Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Empirical Likelihood for Network Data

Yukitoshi Matsushita & Taisuke Otsu

To cite this article: Yukitoshi Matsushita & Taisuke Otsu (23 Aug 2023): Empirical Likelihood for Network Data, Journal of the American Statistical Association, DOI: 10.1080/01621459.2023.2250091

To link to this article: https://doi.org/10.1080/01621459.2023.2250091

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2023, VOL. 00, NO. 0, 1–12: Theory and Methods
https://doi.org/10.1080/01621459.2023.2250091

Taylor & Francis
Taylor & Francis Group

Empirical Likelihood for Network Data

Yukitoshi Matsushita[a] and Taisuke Otsu[b]

[a]Graduate School of Economics, Hitotsubashi University, Kunitachi, Tokyo, Japan; [b]Department of Economics, London School of Economics, London, UK

ABSTRACT
This article develops a concept of nonparametric likelihood for network data based on network moments, and proposes general inference methods by adapting the theory of jackknife empirical likelihood. Our methodology can be used not only to conduct inference on population network moments and parameters in network formation models, but also to implement goodness-of-fit testing, such as testing block size for stochastic block models. Theoretically we show that the jackknife empirical likelihood statistic for acyclic or cyclic subgraph moments loses its asymptotic pivotalness in severely or moderately sparse cases, respectively, and develop a modified statistic to recover pivotalness in such cases. The main advantage of our modified jackknife empirical likelihood method is its validity under weaker sparsity conditions than existing methods although it is computationally more demanding than the unmodified version. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received November 2020
Accepted August 2023

KEYWORDS
Bootstrap/resampling; Goodness-of-fit methods; Nonparametric methods

1. Introduction

Analysis on network data is becoming increasingly important in various fields of data science, such as social networks, technological networks for communications, transportation, and energy, biological networks for food webs and protein interactions, and information networks for collaborations and semantic relationships (see, e.g., Kolaczyk 2009, for a review). With this surge of various network data as a background, there is a rapidly growing literature on modeling and estimation for network data (see, Crane 2018, for a survey on recent developments). In particular, based on the Aldous-Hoover representation for exchangeable random arrays (see, Kallenberg 2005), various statistical models and their sampling properties are studied for network data viewed as exchangeable random graphs; see, for example, Bickel and Chen (2009), Bickel, Chen, and Levina (2011), Bickel et al. (2013), Chatterjee, Diaconis, and Sly (2011), Diaconis and Janson (2008), and Hoff, Raftery, and Handcock (2002). Given this literature on modeling and estimation for network data, substantial progress has been made in recent years for inference methods, such as uncertainty quantification for network moments or functionals, parameter hypotheses testing, and goodness-of-fit testing; see references below.

In this article, we develop a concept of nonparametric likelihood for network data based on network moments, and propose general inference methods by adapting the methodology of jackknife empirical likelihood (JEL). The method of JEL proposed by Jing, Yuan, and Zhou (2009) and extended by Matsushita and Otsu (2020) is an extension of Owen's (1988) empirical likelihood for U-statistics, and constructs a likelihood function for estimating equations based on jackknife pseudo values for the U-statistics. Based on the method of moments estimator by Bickel, Chen, and Levina (2011), we introduce its jackknife pseudo values by using delete-one vertex subgraphs and construct an empirical likelihood function. Then we study its asymptotic properties under the latent variable model of Bickel and Chen (2009), which allows sparse network models. As in Bickel and Chen (2009), our methodology is general enough to cover various network models (e.g., stochastic block models, preferential attachment models, and random dot product graph models) and can be used not only to conduct inference on population network moments and parameters in network models but also to implement goodness-of-fit testing, such as testing block size for stochastic block models.

Theoretically this article makes two contributions. First, we introduce three types of sparsity (mild, moderate, and severe) and show that the JEL statistic for acyclic or cyclic subgraph moments loses its asymptotic pivotalness and converges to a weighted Chi-squared distribution in the severely or moderately sparse case, respectively. A walk on a graph of length at least three, where the starting and ending vertices are the same but all other vertices are distinct from each other, is called a cycle. A graph containing no cycles is called an acyclic graph.[1] We argue that this lack of asymptotic pivotalness is caused by an analogous reason to Efron and Stein's (1981) bias of the jackknife variance estimator. Under the conventional iid setup, Efron and Stein (1981, Theorem 1) established a general higher-order bias formula of the jackknife variance estimator. In our setup for sparse networks, analogous higher-order bias

[1]Typical examples of cyclic graphs are covered by p-cycles (Example 4 in Bhattacharyya and Bickel 2015) which include triangles and squares. Also popular examples of acyclic graphs are wheels (Definition 1 in Bickel, Chen, and Levina 2011) which include edges and stars.

CONTACT Taisuke Otsu t.otsu@lse.ac.uk Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

terms indeed appear in the first-order terms of the JEL statistic, which cause the lack of asymptotic pivotalness. Second, we develop a modified JEL statistic, which recovers asymptotic pivotalness and converges to a Chi-squared distribution in those cases of sparsity (i.e., emergence of Wilks' phenomenon). The basic idea is to incorporate leave-two-out adjustments as in Hinkley (1978) and Efron and Stein (1981) into the estimating equations by the jackknife pseudo values. We emphasize that the main advantage of our modified jackknife empirical likelihood method is its validity under weaker sparsity conditions than existing methods. Although it is computationally more demanding than the unmodified version, this is not a problem for commonly used subgraphs, such as two stars and triangles.

In the statistics literature, several authors proposed inference methods for network data. Anandkumar et al. (2013) employed a method of moments approach for community detection in network models with overlapping communities and studied its guarantees for support or membership recovery. Bhattacharyya and Bickel (2015) developed subsampling methods for smooth functions of network moments. Green and Shalizi (2017) proposed bootstrap procedures based on the empirical graphon. Levin and Levina (2019) proposed a two-step bootstrap procedure involving estimating the latent positions under the assumption of a random dot product graph. Lin, Lunde, and Sarkar (2020a) showed that the network jackknife procedure leads to conservative estimates of the variance for network functionals. They also showed the consistency of the jackknife variance estimates for count functionals under some sparsity conditions. Lin, Lunde, and Sarkar (2020b) proposed a multiplier bootstrap procedure for count functionals and showed that it exhibits higher-order correctness under appropriate sparsity conditions. In contrast to these papers employing some resampling methods, this article proposes a nonparametric likelihood-based inference method based on JEL. Also we emphasize that this article considers inference under more general conditions on the sparsity level. In particular, the above papers exclude the severely sparse case (for acyclic subgraph moments) and the moderately sparse case (for cyclic subgraph moments). Such cases were discussed in Bickel, Chen, and Levina (2011) but inference for these cases is an open question. To the best of our knowledge, this is the first article that establishes an asymptotically valid inference method under such sparsity conditions. Furthermore, our simulation result illustrates desirable finite sample performance of the (modified) JEL inference even for a very small size of network.

This article is organized as follows. In Section 2, we introduce our setup and JEL for network moments, derive its asymptotic properties, and develop a modified statistic to recover asymptotic pivotalness for the scalar case (Section 2.1), vector case (Section 2.2), smooth functions of network moments (Section 2.3), and alternative network moments (Section 2.4). Section 3 presents applications of the proposed (modified) JEL approach for specification testing (Section 3.1), two-sample testing (Section 3.2), goodness-of-fit testing for stochastic block models (Section 3.3), and other network models (Section 3.4). Sections 4 illustrates our methodology by a simulation study. Section 5 presents real data examples for the karate club data (Section 5.1) and Facebook data (Section 5.2). Section 6 concludes. All proofs and derivations are contained in Appendix.

2. Empirical Likelihood

Consider a random graph G_n on vertices $1, \dots, n$ represented by an $n \times n$ adjacency matrix A, where $A_{ij} = 1$ if there is an edge from node i to j and 0 otherwise. We assume that the graph is undirected (i.e., A is symmetric) and contains no self-loops (i.e., diagonals of A are all zero). Let $\mathbb{P}$ be the probability measure of A and $\mathbb{E}$ be its expectation.

A subset $R \subseteq \{(i,j) : 1 \leq i < j \leq n\}$ is identified by the edge set $\mathcal{E}(R) = R$ and the vertex set $\mathcal{V}(R) = \{i : (i,j) \text{ or } (j,i) \in R \text{ for some } j\}$. Typical examples of R include particular patterns, such as triangles, stars, and wheels. Let $G_n(R)$ be the subgraph induced by $\mathcal{V}(R)$. We consider two types of count functionals. The first one is occurrence probability of R defined as

$$P(R) = \mathbb{P}\{\mathcal{E}(G_n(R)) = R\}, \tag{1}$$

and the second one is probability of an induced subgraph containing the subgraph R defined as

$$Q(R) = \mathbb{P}\{R \subseteq G_n(R)\}. \tag{2}$$

These functionals are also studied by Bickel, Chen, and Levina (2011) and Lin, Lunde, and Sarkar (2020a). See Section 2.4 for some advantages of considering $Q(R)$ with some normalization.

To define the method of moments estimator for $P(R)$ and $Q(R)$, we introduce some notion. Two graphs R_1 and R_2 are called isomorphic (denoted by $R_1 \sim R_2$) if there exists a one-to-one map σ of $\mathcal{V}(R_1)$ to $\mathcal{V}(R_2)$ such that the map $(i,j) \to (\sigma_i, \sigma_j)$ is one-to-one from $\mathcal{E}(R_1)$ to $\mathcal{E}(R_2)$. Let $\text{Iso}(R)$ be the set of subgraphs that are isomorphic to R in G_n and $|\text{Iso}(R)|$ be its number of elements. Bickel, Chen, and Levina (2011) proposed to estimate $P(R)$ by

$$\widehat{P}(R) = \frac{1}{\binom{n}{p}|\text{Iso}(R)|} \sum_{S \in \mathcal{G}} \mathbb{I}\{S \sim R\}, \tag{3}$$

where p is the number of vertices of R, $\mathcal{G}$ is the set of all subgraphs of G_n, and $\mathbb{I}\{\cdot\}$ is the indicator function. Obviously $\widehat{P}(R)$ is an unbiased estimator for $P(R)$, and Bickel, Chen, and Levina 2011) developed the asymptotic theory for $\widehat{P}(R)$ under certain sparsity conditions.

We note that the estimator $\widehat{P}(R)$ can be alternatively written as

$$\widehat{P}(R) = \frac{1}{\binom{n}{p}} \sum_{1 \leq i_1 < \cdots < i_p \leq n} Y_{i_1 \dots i_p}(R), \tag{4}$$

where

$$Y_{i_1 \dots i_p}(R) = \frac{1}{|\text{Iso}(R)|} \sum_{S \sim R, \mathcal{V}(S) = \{i_1, \dots, i_p\}} \prod_{(i_k, i_l) \in S} A_{i_k i_l} \prod_{(i_k, i_l) \in \bar{S}} (1 - A_{i_k i_l}),$$

and $\bar{S} = \{(i,j) \notin S, i \in \mathcal{V}(S), j \in \mathcal{V}(S)\}$. For example, (i) if R is an "edge", then $p = 2$ and $Y_{ij}(R) = A_{ij}$; (ii) if R is a "triangle", then $p = 3$ and $Y_{ijl}(R) = A_{ij}A_{jl}A_{il}$; and (iii) if R is a "2-star" (or (1, 2)-wheel), then $p = 3$ and $Y_{ijl}(R) = \frac{1}{3}\{A_{ij}A_{jl}(1 - A_{il}) + A_{ij}(1 - A_{jk})A_{il} + (1 - A_{ij})A_{jl}A_{il}\}$. A (k,l)-wheel is a graph with

$kl+1$ vertices and kl edges isomorphic to the graph with edges $\{(1,2),\ldots,(k,k+1);(1,k+2),\ldots,(2k,2k+1);\ldots,(1,(l-1)k+2),\ldots,(lk,lk+1)\}$. See Bickel, Chen, and Levina (2011, p. 2286).

Similarly, as in Lin, Lunde, and Sarkar (2020a), $Q(R)$ can be estimated by

$$\widehat{Q}(R) = \frac{1}{\binom{n}{p}} \sum_{1\leq i_1<\cdots<i_p\leq n} Y^Q_{i_1\ldots i_p}(R), \tag{5}$$

where

$$Y^Q_{i_1\ldots i_p}(R) = \frac{1}{|\mathrm{Iso}(R)|} \sum_{S\sim R, \mathcal{V}(S)=\{i_1,\ldots,i_p\}} \prod_{(i_k,i_l)\in S} A_{i_k i_l}.$$

Based on the representations in (4) and (5), the only difference of $\widehat{P}(R)$ and $\widehat{Q}(R)$ is presence of the factor "$\prod_{(i_k,i_l)\in\bar{S}}(1-A_{i_k i_l})$". As can be seen from our proof, this additional factor is asymptotically negligible after suitable normalization so that $\rho_n^{-|R|}\widehat{P}(R)$ and $\rho_n^{-|R|}\widehat{Q}(R)$ have the same limiting distribution, where ρ_n is defined in Assumption A. Thus, we hereafter focus on $\widehat{P}(R)$ for the asymptotic analysis.

For statistical inference on $P(R)$, we extend the notion of JEL developed by Jing, Yuan, and Zhou (2009) to network data. To overcome computational and theoretical difficulties of Owen's (1988) original empirical likelihood for the U-statistics, Jing, Yuan, and Zhou (2009) proposed to construct an empirical likelihood function based on jackknife pseudo values for the statistical object of interest, say T. More precisely, for the leave-i-out counterpart T_{-i} of T, its jackknife pseudo value is defined as $V_{T,i} = nT-(n-1)T_{-i}$, and the JEL function for $\tau = \mathrm{plim}_{n\to\infty} T$ is constructed as

$$\ell_T(\tau) = -2 \sup_{\{w_i\}_{i=1}^n} \sum_{i=1}^n \log(nw_i), \quad \text{s.t. } w_i \geq 0,$$
$$\sum_{i=1}^n w_i = 1, \sum_{i=1}^n w_i(V_{T,i}-\tau) = 0. \tag{6}$$

Note that if T is a sample mean (say, $T = n^{-1}\sum_{i=1}^n Z_i$), then it holds $V_{T,i} = Z_i$ and $\ell_T(\tau)$ reduces to Owen's (1988) original empirical likelihood for the population mean of Z_i. However, the JEL approach is applicable to more general statistics, and indeed Jing, Yuan, and Zhou (2009) showed that $\ell_T(\tau)$ converges to Chi-squared distributions for the one- and two-sample U-statistics. Intuitively, they verified Tukey's (1958) conjecture that the jackknife pseudo values can be treated like independent observations. Theoretically even though $V_{T,i}$'s are not independent, they are asymptotically independent for the one- and two-sample U-statistics and relevant limit theorems can still yield Wilks' theorem for the JEL statistic $\ell_T(\tau)$.

In this article we extend the JEL approach to conduct inference on the network moment estimator $\widehat{P}(R)$ in (4) and show that analogous results to Jing, Yuan, and Zhou (2009) can be established even for network data. In particular, we construct the JEL function based on the statistic $T = \widehat{P}(R)$. For the leave-i counterpart T_{-i} in the context of network data analysis, we employ the leave ith vertex out counterpart:

$$\widehat{P}_{-i}(R) = \frac{1}{\binom{n-1}{p}|\mathrm{Iso}(R)|} \sum_{S\in\mathcal{G}_i} \mathbb{I}\{S\sim R\}, \tag{7}$$

where $\mathcal{G}_i$ is the set of all subgraphs that do not contain the ith vertex. Then the jackknife pseudo value for $\widehat{P}(R)$ is defined as

$$V_i = n\widehat{P}(R) - (n-1)\widehat{P}_{-i}(R),$$

for $i = 1,\ldots,n$, and the JEL function for $P(R)$ is obtained as in (6) by replacing $V_{T,i}$ with V_i.

More generally, for subsets $\{R_1,\ldots,R_k\}$, we can analogously define the estimators $(\widehat{P}(R_1),\ldots,\widehat{P}(R_k))$ and the vector of jackknife pseudo values $V_i = (V_{1i},\ldots,V_{ki})'$ for $\theta = (P(R_1),\ldots,P(R_k))'$. Based on this notation, the JEL function for θ is defined as

$$\ell(\theta) = -2 \sup_{\{w_i\}_{i=1}^n} \sum_{i=1}^n \log(nw_i), \quad \text{s.t. } w_i \geq 0,$$
$$\sum_{i=1}^n w_i = 1, \ \sum_{i=1}^n w_i(V_i-\theta) = 0.$$

By applying the Lagrange multiplier method, the dual form of $\ell(\theta)$ is written as

$$\ell(\theta) = 2\sup_{\lambda} \sum_{i=1}^n \log(1+\lambda'(V_i-\theta)). \tag{8}$$

In practice, we use this dual form to implement the JEL inference. In the following sections, we study asymptotic properties of the JEL statistic $\ell(\theta)$ and then develop a modified statistic that exhibits desirable robustness for sparse network data.

2.1. Case of Scalar θ

This section considers the case of $k = 1$, where θ and V_i are scalar and the JEL function is written as $\ell(\theta) = 2\sup_\lambda \sum_{i=1}^n \log(1+\lambda(V_i-\theta))$.

Let $a\wedge b = \min\{a,b\}$ and $a\vee b = \max\{a,b\}$ for $a,b\in\mathbb{R}$. To study the asymptotic properties of $\widehat{P}(R)$ and $\ell(\theta)$, we assume that the network data $\{A_{ij}\}$ are generated from the nonparametric latent variable model in Bickel, Chen, and Levina (2011) and Bhattacharyya and Bickel (2015).

Assumption A. $\{A_{ij}\}$ are generated from

$$A_{ij} = \mathbb{I}\{\xi_{ij} \leq \rho_n w(\xi_i,\xi_j)\wedge 1\}, \tag{9}$$

for $i,j\in\{1,\ldots,n\}$, where $\{\rho_n\}$ is a sequence of positive constants satisfying $\rho_n \to 0$ as $n\to\infty$, $(\xi_1,\ldots,\xi_n,\xi_{11},\ldots,\xi_{nn})$ are iid $U(0,1)$ random variables, and $w(\cdot,\cdot)$ is a positive and symmetric function satisfying $\int_0^1\int_0^1 w(s,t)dsdt = 1$.

This model is derived from a general representation theorem of the adjacency matrix A (Kallenberg 2005, Theorem 7.22) and is flexible to cover popular network formation models, such as stochastic block models, latent variable models, and preferential attachment models (see Kolaczyk 2009, for a review). However, this assumption does not cover the inhomogeneous Erdős-Rényi model and degree-corrected block model by Karrer and Newman (2011) (unless degree parameters are randomized appropriately), for example. Also extending this setup to accommodate covariates would be an important direction of future research.

When $\sup_{a,b\in[0,1]^2} |\rho_n w(a,b)| \leq 1$, the object ρ_n may be interpreted as the edge occurrence probability, and then $d_n = (n-1)\rho_n$ is interpreted as the expected degree. Following the literature (e.g., Borgs, Chayes, and Smith 2015; Klopp, Tsybakov, and Verzelen 2017), we call networks with $\rho_n \to 0$ as *sparse* networks. Otherwise, they are called *dense* networks. Throughout this article, we assume $\rho_n \to 0$ and focus on sparse networks. Indeed many real world networks are considered to be sparse. See Remark 5 for an extension of our result for dense networks.

For sparse networks with $\rho_n \to 0$, we further focus on three cases:

$$\text{Mildly sparse case} : n^{p-1}\rho_n^p \to \infty,$$
$$\text{Moderately sparse case} : n\rho_n \to \infty,\ n^{p-1}\rho_n^p \to C_1 \in [0,\infty),$$
$$\text{Severely sparse case} : n\rho_n \to C_2 \in (0,\infty).$$

For example, if $\rho_n \propto n^{-a}$, the mildly, moderately, and severely sparse cases correspond to $a \in (0,(p-1)/p)$, $a \in [(p-1)/p, 1)$, and $a = 1$, respectively. Intuitively, the mildly and moderately sparse cases yield unbounded average degrees as the number of vertices n diverges, and for the severely sparse case, average degrees remain bounded as n diverges. As we will clarify in Theorem 1, the distinction between the mildly/moderately and severely sparse cases is critical for inference on $P(R)$ with acyclic R. On the other hand, the distinction between the mildly and moderately sparse cases is critical for inference on $P(R)$ with cyclic R. For cyclic R in the severely sparse case, the method of moments estimator $\widehat{P}(R)$ is even inconsistent for $P(R)$.

Existing papers, such as Bickel, Chen, and Levina (2011), Bhattacharyya and Bickel (2015), and Lin, Lunde, and Sarkar (2020a, 2020b), consider the moderately sparse case for acyclic R, and the mildly sparse case for cyclic R. To the best of our knowledge this is the first article which establishes valid inference on network moments in (i) the mildly, moderately, and severely sparse cases for acyclic R, and (ii) the mildly and moderately sparse cases for cyclic R.

Let $\bar{\mathbb{N}}$ consist of all finite sequences $(i_1,\ldots,i_p)$ with distinct entries $i_1,\ldots,i_p \in \mathbb{N}$. By using the latent variables in (9), there exists a measurable function $f : [0,1]^{p+(p-1)p/2} \to [0,1]$ such that

$$Y_{i_1\ldots i_p}(R) = f(\xi_{i_1},\ldots,\xi_{i_p},\xi_{i_1i_2},\ldots,\xi_{i_{p-1}i_p}),$$

for each $(i_1,\ldots,i_p) \in \bar{\mathbb{N}}$. To proceed, we introduce some notation. Let

$$\begin{aligned}
g_1(1) &= \mathbb{E}[Y_{1\ldots p}|\xi_1] - \mathbb{E}[Y_{1\ldots p}],\\
g_2(12) &= \mathbb{E}[Y_{1\ldots p}|\xi_1,\xi_2,\xi_{12}] - g_1(1) - g_1(2) - \mathbb{E}[Y_{1\ldots p}],\\
g_3(123) &= \mathbb{E}[Y_{1\ldots p}|\xi_1,\xi_2,\xi_3,\xi_{12},\xi_{13},\xi_{23}]\\
&\quad - \sum_{i=1}^{3} g_1(i) - \sum_{1\le i_1<i_2\le 3} g_2(i_1i_2) - \mathbb{E}[Y_{1\ldots p}],\\
&\ \ \vdots\\
g_p(12\ldots p) &= \mathbb{E}[Y_{1\ldots p}|\xi_1,\ldots,\xi_p,\xi_{12},\ldots,\xi_{p-1,p}]\\
&\quad - \sum_{i=1}^{p} g_1(i) - \sum_{1\le i_1<i_2\le p} g_p(i_1i_2)\\
&\quad - \cdots - \sum_{1\le i_1<\cdots<i_{p-1}\le p} g_{p-1}(i_1\ldots i_{p-1})\\
&\quad - \mathbb{E}[Y_{1\ldots p}].
\end{aligned}$$

Note that $Y_{1\ldots p} = \mathbb{E}[Y_{1\ldots p}|\xi_1,\ldots,\xi_p,\xi_{12},\ldots,\xi_{p-1,p}]$. Let $|R| = |\mathcal{E}(R)|$ be the number of edges in R. Based on the above notation and repeated add and subtractions, the estimation error admits the following ANOVA-type decomposition.

Proposition 1. Suppose Assumption A holds true. Then it holds

$$\rho_n^{-|R|}\{\widehat{P}(R) - P(R)\} = \frac{1}{n}\sum_{i=1}^{n}\beta_i + \frac{1}{n^2}\sum_{i_1<i_2}\beta_{i_1i_2} + \cdots + \frac{1}{n^p}\sum_{i_1<\cdots<i_p}\beta_{i_1\ldots i_p}, \quad (10)$$

where $\beta_i = \rho_n^{-|R|}pg_1(i)$, $\beta_{i_1i_2} = \rho_n^{-|R|}2!\binom{p}{2}g_2(i_1i_2),\ldots,\beta_{i_1\ldots i_p} = \rho_n^{-|R|}p!g_p(i_1\ldots i_p)$. Furthermore, it holds $\frac{1}{n}\sum_{i=1}^n \beta_i = O_p\left(\frac{1}{\sqrt{n}}\right)$, $\frac{1}{n^s}\sum_{i_1<\cdots<i_s}^n \beta_{i_1\ldots i_s} = O_p\left(\frac{1}{\sqrt{n^s\rho_n^{s-1}}} \vee \frac{1}{\sqrt{n^s}}\right)$ for $s = 2,\ldots,p-1$, and

$$\frac{1}{n^p}\sum_{i_1<\cdots<i_p}^{n}\beta_{i_1\ldots i_p} = \begin{cases} O_p\left(\frac{1}{\sqrt{n^p\rho_n^{p-1}}} \vee \frac{1}{\sqrt{n^p}}\right) & \text{if } R \text{ is acyclic}\\ O_p\left(\frac{1}{\sqrt{n^p\rho_n^{p}}} \vee \frac{1}{\sqrt{n^p}}\right) & \text{if } R \text{ is cyclic}\end{cases}. \quad (11)$$

Note that all the random variables on the right side of (10) have zero mean and no correlation. The above decomposition is also employed by Lin, Lunde, and Sarkar (2020a), but is different from the one by Bickel, Chen, and Levina (2011). The decomposition in (10) is particularly suitable for our asymptotic analysis since the uncorrelatedness of the components in (10) enables to apply Efron and Stein's (1981) argument for the modification on certain discrepancy in the variance components; see a remark on Theorem 1 for further detail. Note that for the Erdős-Rényi model, β_i (or ξ_i) is zero for all i, and the first term in (10) disappears.

Remark 1 (Implication of (11)). It is important to note that the term in (11) exhibits different stochastic orders for the cases of acyclic and cyclic R. This difference is due to different orders of the variances in the main term of $\beta_{i_1\ldots i_p}$. If R is acyclic, for the mildly and moderately sparse cases (i.e., $\rho_n \to 0$ and $n\rho_n \to \infty$), the linear term $\frac{1}{n}\sum_{i=1}^n \beta_i$ will be a leading term in (10). On the other hand, if R is cyclic, we need the condition $n\rho_n \to \infty$ for the consistency, $\rho_n^{-|R|}\{\widehat{P}(R) - P(R)\} \xrightarrow{p} 0$, due to the order in (11). Thus, it holds $\frac{1}{n^s}\sum_{i_1<\cdots<i_s}^n \beta_{i_1\ldots i_s} = o_p\left(\frac{1}{n}\sum_{i=1}^n \beta_i\right)$ for $s = 2,\ldots,p-1$. If $n^{p-1}\rho_n^p \to \infty$ (i.e., $n^p\rho_n^p/n \to \infty$), then the limiting distribution of $\widehat{P}(R)$ is determined by the linear term $\frac{1}{n}\sum_{i=1}^n \beta_i$. If $n^{p-1}\rho_n^p = O(1)$ (i.e., $n^p\rho_n^p/n = O(1)$), then the limiting distribution of $\widehat{P}(R)$ is determined by the two terms $\frac{1}{n}\sum_{i=1}^n \beta_i$ and $\frac{1}{n^p}\sum_{i_1<\cdots<i_p}^n \beta_{i_1\ldots i_p}$. In other words, the distinction between the mildly and moderately sparse cases is critical when R is cyclic.

We now present the limiting distribution of the JEL statistic $\ell(\theta)$. Let $\mathbb{V}(\cdot)$ mean the variance. Define $\sigma_{s,n}^2 = \mathbb{V}(\beta_{1\ldots s})$ for $s = 1, \ldots, p$, and $\sigma_*^2 = \lim_{n\to\infty}(\sigma_n^2/\omega_n)$, where

$$\omega_n = \frac{\sigma_{1,n}^2}{n} + \frac{\sigma_{2,n}^2}{2n^2} + \frac{\sigma_{3,n}^2}{6n^3} + \cdots + \frac{\sigma_{p,n}^2}{p!n^p},$$
$$\sigma_n^2 = \frac{\sigma_{1,n}^2}{n} + \frac{\sigma_{2,n}^2}{n^2} + \frac{\sigma_{3,n}^2}{2n^3} + \cdots + \frac{\sigma_{p,n}^2}{(p-1)!n^p}. \quad (12)$$

Theorem 1. Suppose Assumption A holds true.

(i) If R is acyclic, then

$$\ell(\theta) \xrightarrow{d} \begin{cases} \chi_1^2 & \text{for mildly or moderately sparse case with random } \mathbb{E}[\beta_1|\xi_1], \\ \sigma_*^{-2}\chi_1^2 & \text{for severely sparse case, or nonrandom } \mathbb{E}[\beta_1|\xi_1]. \end{cases}$$

(ii) If R is cyclic, then

$$\ell(\theta) \xrightarrow{d} \begin{cases} \chi_1^2 & \text{for mildly sparse case with random } \mathbb{E}[\beta_1|\xi_1], \\ \sigma_*^{-2}\chi_1^2 & \text{for moderately sparse case, or nonrandom } \mathbb{E}[\beta_1|\xi_1]. \end{cases}$$

This theorem shows that the limiting distribution of the JEL statistic $\ell(\theta)$ depends on the degree of network sparsity. Part (i) of this theorem is on the case of acyclic R. For the mildly and moderately sparse cases (i.e., as far as $n\rho_n \to \infty$) with random $\mathbb{E}[\beta_1|\xi_1]$, the JEL statistic is asymptotically pivotal. However, for the severely sparse case (i.e., $n\rho_n \to C_2 \in (0, \infty)$) or degenerate $\mathbb{E}[\beta_1|\xi_1]$, the JEL statistic is no longer asymptotically pivotal.

For the case of cyclic R, Part (ii) of this theorem shows that lack of asymptotic pivotalness of the JEL statistic occurs in the moderately sparse case. Recall that for cyclic R, the method of moments estimator $\widehat{P}(R)$ is even inconsistent for $P(R)$ in the severely sparse case.

Remark 2 (Nonpivotal distribution). It is interesting to note that the nonpivotal limiting distribution depends on $\sigma_*^2 = \lim_{n\to\infty}(\sigma_n^2/\omega_n)$, which is the limit of the ratio of a normalized sample variance $\frac{\rho_n^{-2|R|}}{n^2}\sum_{i=1}^n (V_i - \theta)^2$ to the population variance $\omega_n = \mathbb{V}\left(\frac{\rho_n^{-|R|}}{n}\sum_{i=1}^n (V_i - \theta)\right)$. Note that $\sigma_*^2 \le 1$ by the definitions in (12). Thus, if we use the χ^2 critical value $\chi_{1,\alpha}^2$ for the JEL statistic $\ell(\theta)$ in the severely sparse case (for acyclic R) or moderately sparse case (for cyclic R), the resulting JEL confidence interval $\{\theta : \ell(\theta) \le \chi_{1,\alpha}^2\}$ will exhibit over-coverage. The discrepancy of σ_n^2 and ω_n in (12) is analogous to Efron and Stein's (1981) bias in this context. In other words, the Efron-Stein bias for the jackknife variance estimator emerges in the first-order asymptotics in the moderately and severely sparse cases. The discrepancy of σ_n^2 and ω_n can be large when the components $(\sigma_{2,n}^2, \ldots, \sigma_{p,n}^2)$ are relatively large compared to $\sigma_{1,n}^2$. For example, the Erdős-Rényi model satisfies $\sigma_{1,n}^2 = 0$ so that the discrepancy tends to be large.

Remark 3 (Degenerate case). The case where $\mathbb{E}[\beta_1|\xi_1]$ becomes random corresponds to nondegeneracy of the U-statistic in the current context (see also Menzel 2018). This only excludes the possibility that $\mathbb{E}[Y_{1\cdots p}|\xi_1]$ has a degenerate distribution, where the conditional means given ξ_1 happen to be constant. We note that this degeneracy yields a nonstandard limiting distribution of $\ell(\theta)$ only when ρ_n converges to a nonzero constant (i.e., the network is dense), which is excluded in Theorem 1. In particular, the terms of order $O_p(1/\sqrt{n^s})$ in Proposition 1 will induce nonstandard limiting behaviors.

Remark 4 (Inference on ρ_n). An application of Theorem 1 (i) is inference on ρ_n, which can be interpreted as the link formation probability $\mathbb{P}(A_{ij} = 1)$. Note that it can be written as $\mathbb{P}(A_{ij} = 1) = P(R_0)$ for the $(1, 1)$-wheel R_0, which is an acyclic graph. Thus, Theorem 1 directly applies by using the estimator $\widehat{P}(R_0) = \binom{n}{2}^{-1}\sum_{k=1}^n \sum_{l=k+1}^n A_{kl}$ for ρ_n. This special case is studied by Matsushita and Otsu (2020).

Our next step is to modify the JEL statistic to recover asymptotic pivotalness. To this end, we employ the bias correction method suggested by Efron and Stein (1981). Let $\widehat{P}_{-i_1,\ldots,-i_s}(R)$ be the leave-$(i_1, \ldots, i_s)$-out version of $\widehat{P}(R)$, and define

$$M_{i_1\ldots i_s} = n\widehat{P}(R) - (n-1)\left(\sum_{i=1}^s \widehat{P}_{-i}(R)\right) + (n-2)\left(\sum_{i_1<i_2}^s \widehat{P}_{-i_1,-i_2}(R)\right) + \cdots + (-1)^s(n-p)\widehat{P}_{-i_1,\ldots,-i_s}(R),$$

for $s = 2, \ldots, p$. These terms are used in Efron and Stein (1981) to correct the higher-order bias of the jackknife variance estimator when $p = 2$. Since the second sample moment of $M_{i_1\ldots i_s}$ is related to the components $(\sigma_{s,n}^2, \ldots, \sigma_{p,n}^2)$ as shown in Lemma 1 in Appendix, it can be used to adjust mismatch in the variance components of σ_*^2 due to Efron and Stein's bias.

By using these terms, we modify the JEL statistic as

$$\ell^m(\theta) = 2\sup_\lambda \sum_{i=1}^n \log(1 + \lambda V_i^m(\theta)), \quad (13)$$

where $V_i^m(\theta) = (V_i - \widehat{\theta}) + \widehat{\Gamma}\widetilde{\Gamma}^{-1}(\widehat{\theta} - \theta)$ with $\widehat{\theta} = \widehat{P}(R)$, and $\widehat{\Gamma}$ and $\widetilde{\Gamma}$ are given by

$$\widehat{\Gamma} = \sqrt{\sum_{i=1}^n (V_i - \widehat{\theta})^2}, \quad (14)$$
$$\widetilde{\Gamma} = \sqrt{\sum_{i=1}^n (V_i - \widehat{\theta})^2 - \sum_{i_1<i_2}^n M_{i_1 i_2}^2 - \cdots - (-1)^p \sum_{i_1<\cdots<i_p}^n M_{i_1\ldots i_p}^2}.$$

The asymptotic property of the modified JEL statistic is obtained as follows.

Theorem 2. Suppose Assumption A holds true.

(i) If R is acyclic, then $\ell_m(\theta) \xrightarrow{d} \chi_1^2$ in the mildly, moderately, and severely sparse cases.

(ii) If R is cyclic, then $\ell_m(\theta) \xrightarrow{d} \chi_1^2$ in the mildly and moderately sparse cases.

The modified JEL statistic $\ell_m(\theta)$ is our main proposal and this theorem clarifies its major advantage, robust size properties even for the severely sparse case, which is a challenging case in the literature and is not covered by the existing methods. The cost for this advantage is computational one, particularly leave-out operations to obtain $\widetilde{\Gamma}$ in (14).

In practice, it is common to employ smaller subgraphs, such as two stars and triangles, as R. In this case, the computational cost for obtaining $\widetilde{\Gamma}$ is relatively small. Moreover, for certain subgraphs, it is relatively easy to compute the leave-out objects $M_{i_1\ldots i_p}$. For example, in the case of triangles, one may use A, A^2, and A^3 to compute $M_{i_1\ldots i_p}$ since the ith diagonal element of A^3 contains all triangles for the ith vertex.

Part (i) of this theorem says that for acyclic R, the modified JEL statistic $\ell_m(\theta)$ is asymptotically pivotal and converges to the χ^2_1 distribution regardless of sparsity of the network. We emphasize that this theorem covers the severely sparse case (i.e., $n\rho_n \to C \in (0,\infty)$). To the best of our knowledge, this is the first result to provide an asymptotically valid inference in the severely sparse case. Part (ii) of this theorem says that for cyclic R, asymptotic pivotalness of the modified JEL statistic $\ell_m(\theta)$ is maintained in the mildly and moderately cases.

Moreover, these limiting behaviors are robust to degeneracy of the component $\mathbb{E}[\beta_1|\xi_1]$. Based on this theorem, the asymptotic $1-\alpha$ confidence set for θ can be obtained as $ELCI_\alpha = \{\theta : \ell_m(\theta) \le \chi^2_{1,\alpha}\}$, where $\chi^2_{1,\alpha}$ is the $(1-\alpha)$th quantile of the χ^2_1 distribution.

Remark 5 (Dense network). Although this article focuses on sparse networks satisfying $\rho_n \to 0$, our proof of Theorem 1 can also be adapted to cover dense networks, where ρ_n converges to a nonzero constant. For example, as in Theorem 1, if $\mathbb{E}[\beta_1|\xi_1]$ is random, then $\ell(\theta) \xrightarrow{d} \chi^2_1$ even for dense networks. Furthermore, a similar argument to the proof of Theorem 2 yields $\ell_m(\theta) \xrightarrow{d} \chi^2_1$ if $\mathbb{E}[\beta_1|\xi_1]$ is random.

Remark 6 (Computationally cheaper modification). To reduce the computational burden of calculating $\widetilde{\Gamma}$ in (14) which involves the leave-$(i_1,\ldots,i_p)$-out versions of $\widehat{P}(R)$, we could replace $\widetilde{\Gamma}$ by a size $b = cn$ subsample counterpart for some $c \in (0,1)$:

$$\widetilde{\Gamma}_{sub} = \sqrt{\sum_{i=1}^{n}(V_i-\widehat{\theta})^2 - \sum_{i_1<i_2}^{b} M^{*2}_{i_1i_2} - \cdots - (-1)^p \sum_{i_1<\cdots<i_p}^{b} M^{*2}_{i_1\ldots i_p}},$$

where $M^{*2}_{i_1\ldots i_p} = \frac{n(n-1)\cdots(n-p+1)}{b(b-1)\cdots(b-p+1)} M^2_{i_1\ldots i_p}$. We note that the statements in Theorem 2 still hold under the same assumptions even if we use $\widetilde{\Gamma}_{sub}$ instead of $\widetilde{\Gamma}$. For the simulation study in Section 4 and empirical example in Section 5.2, we use $\widetilde{\Gamma}_{sub}$ with $b = 50$.

2.2. Case of Vector θ

For a vector case, we can apply the decomposition in (10) for each element in the vector $(\widehat{P}(R_1) - P(R_1), \ldots, \widehat{P}(R_k) - P(R_k))$ with corresponding components $\{\beta^{(j)}_i, \ldots, \beta^{(j)}_{i_1\ldots i_{p_j}}\}$ for $j = 1,\ldots,k$. Define $\sigma^{(j,h)2}_{s,n} = \mathbb{E}[\beta^{(j)}_{i_1\ldots i_s}\beta^{(h)}_{i_1\ldots i_s}]$ for $s = 1,\ldots,p$, and

$$\Omega_n = k \times k \text{ matrix with } (j,h)\text{th element } \frac{\sigma^{(j,h)2}_{1,n}}{n} + \frac{\sigma^{(j,h)2}_{2,n}}{2n^2} + \frac{\sigma^{(j,h)2}_{3,n}}{6n^3} + \cdots + \frac{\sigma^{(j,h)2}_{p_j\wedge p_h,n}}{p!n^{p_j\wedge p_h}}, \tag{15}$$

$$\Sigma_n = k \times k \text{ diagonal matrix with } (j,h)\text{th element } \frac{\sigma^{(j,h)2}_{1,n}}{n} + \frac{\sigma^{(j,h)2}_{2,n}}{n^2} + \frac{\sigma^{(j,h)2}_{3,n}}{2n^3} + \cdots + \frac{\sigma^{(j,h)2}_{p_j\wedge p_h,n}}{(p-1)!n^{p_j\wedge p_h}}.$$

Based on the above notation, the limiting distribution of the JEL statistic $\ell(\theta)$ in (8) is obtained as follows. To simplify the presentation, we only present the result corresponding to Part (i) of Theorem 1.

Theorem 3. Suppose Assumption A holds true, and Ω_n and Σ_n are positive definite for all n large enough. If R_j is acyclic for each $j = 1,\ldots,k$, then

$$\ell(\theta) \xrightarrow{d} \zeta'\Sigma^{*-1}\zeta,$$

where $\Sigma^* = \lim_{n\to\infty}\Omega_n^{-1/2}\Sigma_n\Omega_n^{-1/2}$ and $\zeta \sim N(0, I_k)$.

Since the proof is similar to that of Theorem 1, it is omitted. Similar to Theorem 1 for the case of scalar θ, the JEL statistic is not asymptotically pivotal and depends on the unknown component Σ^*. When $n\rho_n \to \infty$ (the mildly and moderately sparse cases) and $\mathbb{E}[\beta^{(j)}_1|\xi_1]$ is random for all $j = 1,\ldots,k$, we can recover asymptotic pivotalness as $\ell(\theta) \xrightarrow{d} \chi^2_k$. The discrepancy of Σ_n and Ω_n can be understood as Efron and Stein's (1981) bias in this context. Note that the variance components Σ_n and Ω_n only contain the covariance terms up to the order $p_j \wedge p_h$. This is due to uncorrelatedness of $\beta^{(j)}_{i_1\ldots i_s}$'s. Finally, analogous results can be derived for the case where some or all of $(R_1,\ldots,R_k)$ are cyclic. In this case, we need to impose the additional condition $n\rho_n \to \infty$.

To recover asymptotic pivotalness for the case of vector θ, the JEL statistic is modified as follows

$$\ell^m(\theta) = 2\sup_\lambda \sum_{l=1}^{n}\log(1+\lambda' V^m_i(\theta)), \tag{16}$$

where $V^m_i(\theta) = (V_i - \widehat{\theta}) + \widehat{\Gamma}\widetilde{\Gamma}^{-1}(\widehat{\theta}-\theta)$ with $\widehat{\theta} = (\widehat{P}(R_1),\ldots,\widehat{P}(R_k))'$, and $\widehat{\Gamma}$ and $\widetilde{\Gamma}$ are given by

$$\widehat{\Gamma}\widehat{\Gamma}' = \sum_{i=1}^{n}(V_i-\widehat{\theta})(V_i-\widehat{\theta})',$$

$$\widetilde{\Gamma}\widetilde{\Gamma}' = k \times k \text{ matrix with } (j,h)\text{th element } \left\{\begin{array}{l} \sum_{i=1}^{n}(V^{(j)}_i - \widehat{\theta}^{(j)})(V^{(h)}_i - \widehat{\theta}^{(h)}) - \sum_{i_1<i_2} M^{(j)}_{i_1i_2}M^{(h)}_{i_1i_2} \\ - \cdots - (-1)^{p_j\wedge p_h}\sum_{i_1<\cdots<i_{p_j\wedge p_h}} M^{(j)}_{i_1\ldots i_{p_j\wedge p_h}} M^{(h)}_{i_1\ldots i_{p_j\wedge p_h}} \end{array}\right\}. \tag{17}$$

The asymptotic property of the modified JEL statistic is obtained as follows.

Theorem 4. Suppose Assumption A holds true, and Ω_n and Σ_n are positive definite for all n large enough. If R_j is acyclic for each $j = 1, \ldots, k$, then

$$\ell_m(\theta) \xrightarrow{d} \chi^2_k.$$

Also the same result can be obtained even if some or all of $(R_1, \ldots, R_k)$ are cyclic under the additional condition $n\rho_n \to \infty$.

Since the proof is similar to that of Theorem 2, it is omitted. Similar comments to Theorem 2 apply. Even if θ is a vector, the modified JEL statistic $\ell_m(\theta)$ is asymptotically pivotal and converges to the χ^2_k distribution under mild conditions on network sparsity.

2.3. Inference on Smooth Function of θ

The asymptotic theory for the modified JEL statistic in the last section can be extended to deal with the case where the object of interest is a smooth function of θ, say $\vartheta = h(\theta)$. Examples include inference on the normalized object $\vartheta = P(R_0)^{-|R|}P(R)$ with the (1,1)-wheel R_0, and the transitivity index $\vartheta = P(R_1)/\{P(R_1) + P(R_2)\}$, where R_1 is a 3-cycle and R_2 is a (1,2)-wheel.

In this case, we can adapt the argument in Hall and La Scala (1990, Theorem 2.1) to establish the asymptotic property of the modified JEL statistic.

Theorem 5. Suppose Assumption A holds true, and Ω_n and Σ_n defined in (15) are positive definite for all n large enough. If R_j is acyclic for each $j = 1, \ldots, k$, $h(\cdot)$ is continuously differentiable in a neighborhood of θ, and $\partial h(\theta)/\partial\theta'$ has the full column rank, then

$$\ell_m(\vartheta) = \min_{\theta \in \{\theta : h(\theta) = \vartheta\}} \ell_m(\theta) \xrightarrow{d} \chi^2_{\dim(\vartheta)}.$$

Also an analogous result can be obtained even if some or all of $(R_1, \ldots, R_k)$ are cyclic under the additional condition $n\rho_n \to \infty$.

The asymptotic $1-\alpha$ confidence set for ϑ can be obtained as $\{\vartheta : \ell_m(\vartheta) \le \chi^2_{\dim(\vartheta),\alpha}\}$. It should be noted that our modified JEL statistic $\ell_m(\vartheta)$ and its confidence set do not suffer from the linearization errors as in the delta method whose effect may be nontrivial in finite samples particularly for highly nonlinear objects, such as $\vartheta = P(R_0)^{-|R|}P(R)$ with large $|R|$.

2.4. Inference on $Q(R)$

Based on the representations in (4) and (5), our theoretical developments for $\widehat{P}(R)$ so far can be adapted to the estimator $\widehat{Q}(R)$ for $Q(R)$ in (2). By using the leave-i counterpart $\widehat{Q}_{-i}(R)$ of $\widehat{Q}(R)$, the jackknife pseudo value for $\widehat{Q}(R)$ can be defined as $V_{Q,i} = n\widehat{Q}(R) - (n-1)\widehat{Q}_{-i}(R)$, and the modified JEL function $\ell_m(\theta_Q)$ for $\theta_Q = Q(R)$ can be defined as in (16) with

$$V^m_{Q,i}(\theta_Q) = (V_{Q,i} - \widehat{\theta}_Q) + \widehat{\Gamma}_Q \widetilde{\Gamma}_Q^{-1}(\widehat{\theta}_Q - \theta_Q), \tag{18}$$

where $\widehat{\theta}_Q = \widehat{Q}(R)$, and $\widehat{\Gamma}_Q$ and $\widetilde{\Gamma}_Q$ are defined as in (14) by replacing $\widehat{P}(R)$ with $\widehat{Q}(R)$.

By a similar argument for Theorem 4, we obtain

$$\ell_m(\theta_Q) \xrightarrow{d} \chi^2_{\dim(\theta_Q)},$$

under the same assumption of Theorem 4.

We note that there are some advantages to consider $\rho_n^{-|R|}Q(R)$ instead of $\rho_n^{-|R|}P(R)$. The first is that it is often more computationally tractable to find "edge" matches for $\widehat{Q}(R)$ rather than "exact" matches for $\widehat{P}(R)$. The second is that $\rho_n^{-|R|}Q(R) = E\left[\prod_{\{i,j\}\in R} w(\xi_i, \xi_j)\right]$ is independent of n, whereas $\rho_n^{-|R|}P(R) = E\left[\prod_{\{i,j\}\in R} w(\xi_i, \xi_j) \prod_{\{i,j\}\in \bar{R}}(1 - \rho_n w(\xi_i, \xi_j))\right]$ does depend on n. For example, when we consider two-sample testing of network homogeneity as in Section 3.2, it is more natural to assess homogeneity of $\rho_n^{-|R|}Q(R)$. As discussed in Section 2.3, Theorem 5 can be applied to conduct inference on $\rho_n^{-|R|}P(R) = P(R_0)^{-|R|}P(R)$ with the (1,1)-wheel R_0. Similarly, Theorem 5 can be adapted to the normalized object $\rho_n^{-|R|}Q(R) = P(R_0)^{-|R|}Q(R)$, which is a smooth function of $(P(R_0), Q(R))$.

3. Applications

In this section, we apply the methodology of (modified) JEL in the last section to several statistical models and problems for network data.

3.1. Specification Test

Consider the network model

$$A_{ij} = \mathbb{I}\{\xi_{ij} \le \rho_n^0 w^0(\xi_i, \xi_j) \wedge 1\}, \tag{19}$$

where $w^0(\cdot,\cdot)$ and ρ_n^0 are specified by the researcher and contain no nuisance parameter. Validity of this specified model may be assessed by testing the hypothesis $H_0 : P(R_j) = P^0(R_j)$ for $j = 1, \ldots, k$, where R_j's are chosen by researcher's interest. For example, Bhattacharyya and Bickel (2015, sec. 6.1) specified a stochastic block model and preferential attachment model to conduct inference on high school network data. Also Zhang and Xia (2022) studied this testing in detail. Our (modified) JEL approach can provide an alternative method for this testing problem. We emphasize that in contrast to these existing papers, our method allows the severely sparse case (for acyclic R_j) and moderately sparse case (for cyclic R_j).

Let $\{P^0(R_j) : j = 1, \ldots, k\}$ be the network moments implied by the null model in (19). Then the modified JEL goodness-of-fit statistic is given by $\ell_m^{\text{gof}} = \ell_m(P^0(R_1), \ldots, P^0(R_k))$ using the definition in (16), which converges to χ^2_k by Theorem 4. Similarly we can conduct goodness-of-fit testing based on $Q^0(R)$ specified by the researcher.

Although it is beyond the scope of this article, we conjecture that consistent goodness-of-fit testing of the specified model in (19) would be possible by letting $k \to \infty$ as $n \to \infty$.

3.2. Two-Sample Test

Our JEL approach can be extended to two-sample testing problems. Suppose we wish to test whether two network data $\{A^1_{ij}\}$

and $\{A_{ij}^2\}$ share the same features in terms of moments for $(R_1, \ldots, R_k)$. For example, Bhattacharyya and Bickel (2015) studied homogeneity of two subnetworks drawn from Facebook network. Let $\{Q^1(R_j), Q^2(R_j) : j = 1, \ldots, k\}$ be the network moments for $\{A_{ij}^1\}$ and $\{A_{ij}^2\}$, respectively. Then the null of the two-sample testing problem is formulated as $H_0 : Q^1(R_j) = Q^2(R_j)$ for $j = 1, \ldots, k$.

By extending the two-sample empirical likelihood (Wu and Yan 2012) to network data, the modified JEL statistic for H_0 can be constructed as $\min_\theta \ell_m^{\text{hom}}(\theta, \theta)$, where

$$\ell_m^{\text{hom}}(\theta_1, \theta_2) = 2 \sup_{\lambda_1} \sum_{l_1=1}^{n} \log(1 + \lambda_1' V_{Q,1,l_1}^m(\theta_1)) + 2 \sup_{\lambda_2} \sum_{l_2=1}^{m} \log(1 + \lambda_2' V_{Q,2,l_2}^m(\theta_2)),$$

and $V_{Q,1,l_1}^m(\theta)$ and $V_{Q,2,l_2}^m(\theta)$ are defined as in (18) using $\{A_{ij}^1\}$ and $\{A_{ij}^2\}$, respectively. If $\{A_{ij}^1\}$ and $\{A_{ij}^2\}$ are independent, we can apply the analogous argument by Wu and Yan (2012) to our context, which implies $\min_\theta \ell_m^{\text{hom}}(\theta, \theta) \xrightarrow{d} \chi_k^2$ under H_0 and analogous conditions to Theorem 4.

3.3. Stochastic Block Model

Consider the function $w(\cdot, \cdot)$ in (9) corresponding to a K-block model so that the implied link formation probabilities are written as

$$\mathbb{P}\{A_{ij} = 1 | i \in a, j \in b\} = \rho_n S_{ab},$$

for $a, b = 1, \ldots, K$ by some $K \times K$ matrix $S = [S_{ab}]$, a $K \times 1$ vector π for block assignment probabilities for the blocks $\{1, \ldots, K\}$, and ρ_n as defined in (9). Let $\eta = (\pi, \rho_n, S)$ and $F = \rho_n S$. The number of free parameters in this block model is $K - 1$ for π and $K(K+1)/2$ for F. Note that ρ_n satisfies $\sum_{a=1}^K \sum_{b=1}^K F_{ab} = \rho_n$. For example, when $K = 1$, this model becomes the Erdős-Rényi model, which contains only one free parameter. Also, when $K = 3$, the number of free parameters is eight, which can be identified by eight moments. In this section, we consider goodness-of-fit (or block size) testing: $H_0 : K = K_0$ for some specified value K_0 versus $H_1 : K > K_0$.

Let $\mathcal{L}_2(0, 1)$ be the L_2 space for functions defined on the interval $(0, 1)$, and $T : \mathcal{L}_2(0, 1) \to \mathcal{L}_2(0, 1)$ be an operator defined by

$$[Tf](u) = \int_0^1 h(u, v) f(v) dv,$$

where $h(u, v) = \rho_n w(u, v)$. For stochastic block models, it is convenient to consider the moment $Q(R) = \mathbb{P}\{A_{ij} = 1, \text{for all } (i, j) \in R\}$. From Bickel, Chen, and Levina (2011, Theorem 1), stochastic block models are generally identified by some set of wheels. Therefore, this section focuses on the case where R's are wheels. If the graph R is a (k, l)-wheel, it can be written as

$$Q(R) = \mathbb{E}\left[\prod_{(i,j) \in \mathcal{E}(R)} h(\xi_i, \xi_j)\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[\prod_{(i,j)} h(\xi_i, \xi_j) : (i, j) \in \mathcal{E}(R) \middle| \xi_1\right]\right]$$
$$= \mathbb{E}\left[\left(\int_0^1 \cdots \int_0^1 h(\xi_1, \xi_2) \cdots h(\xi_k, \xi_{k+1}) d\xi_2 \cdots \xi_{k+1}\right)^l\right]$$
$$= \mathbb{E}[\{T^k(1)(\xi_1)\}^l].$$

Based on this formula, we can compute the moment, say $Q(R; \eta)$, implied from given parameter values $\eta = (\pi, \rho_n, S)$.

For example, when $K = 2$, it can be written as $T(1)(\xi) = \nu_1$ for $\xi \in [0, \pi_1]$ and ν_2 for $\xi \in (\pi_1, 1]$, where $\nu_j = \pi_1 F_{1j} + (1 - \pi_1) F_{2j}$ with $\nu_1 < \nu_2$. Let W_{kl} be a (k, l)-wheel. Thus, the first three moments of $T(1)(\xi)$ are

$$\mathbb{E}[\{T(1)(\xi)\}^l] = \mathbb{E}[Q(W_{1l})] = \pi_1 \nu_1^l + (1 - \pi_1) \nu_2^l,$$

for $l = 1, 2, 3$. Similarly, we have $T(1)^2(\xi) = \pi_1 \nu_1 F_{11} + (1 - \pi_1) \nu_1 F_{21}$ for $\xi \in [0, \pi_1]$ and $\pi_1 \nu_1 F_{12} + (1 - \pi_1) \nu_2 F_{22}$ for $\xi \in (\pi_1, 1]$. Thus, the first three moments of $T(1)^2(\xi)$ are

$$\begin{aligned}\mathbb{E}[\{T(1)^2(\xi)\}^l] &= \mathbb{E}[Q(W_{2l})] \\ &= \pi_1 \{\pi_1 \nu_1 F_{11} + (1 - \pi_1) \nu_2 F_{21}\}^l \\ &\quad + (1 - \pi_1) \{\pi_1 \nu_1 F_{12} + (1 - \pi_1) \nu_2 F_{22}\}^l.\end{aligned}$$

For wheels $(R_1, \ldots, R_k)$, we consider the estimator $\widehat{Q}(R_j)$ in (5) of the moment $Q(R_j, \eta)$ for $j = 1, \ldots, k$. The jackknife pseudo value for $\widehat{Q}(R_j)$ can be defined as $V_{Q,i}^{(j)} = n\widehat{Q}(R_j) - (n-1)\widehat{Q}_{-i}(R_j)$ for $j = 1, \ldots, k$. Then the modified JEL function $\ell_m(\eta)$ can be defined as in (16) by setting

$$V_{Q,i}^m(\eta) = (V_{Q,i} - \widehat{\theta}_Q) + \widehat{\Gamma}\widetilde{\Gamma}^{-1}(\widehat{\theta}_Q - \theta_Q(\eta)), \tag{20}$$

where $V_{Q,i} = (V_{Q,i}^{(1)}, \ldots, V_{Q,i}^{(k)})'$, $\widehat{\theta}_Q = (\widehat{Q}(R_1), \ldots, \widehat{Q}(R_k))'$, $\theta_Q(\eta) = (Q(R_1; \eta), \ldots, Q(R_k; \eta))'$, and $\widehat{\Gamma}$ and $\widetilde{\Gamma}$ are defined as in (17) by replacing $\{\widehat{P}(R_1), \ldots, \widehat{P}(R_k)\}$ with $\{\widehat{Q}(R_1), \ldots, \widehat{Q}(R_k)\}$. The goodness-of-fit statistic based on the modified JEL is defined as

$$T_n = \min_{\eta \in \Upsilon} \ell_m(\eta), \tag{21}$$

and the asymptotic property of this statistic is presented as follows.

Theorem 6. Assume (i) there exists a unique $\eta_0 \in \text{int}(\Upsilon)$ such that $Q(R_j) = Q(R_j; \eta_0)$ is satisfied with all wheels R_j for $j = 1, \ldots k$, and Υ is compact, (ii) $\theta(\eta)$ is continuously differentiable in a neighborhood of η_0 and $\partial\theta(\eta_0)/\partial\eta'$ has the full column rank, and (iii) Assumption A holds true, and Ω_n and Σ_n defined in (15) are positive definite for all n large enough. Then under $H_0 : K = K_0$, it holds

$$T_n \xrightarrow{d} \chi^2_{k - \dim(\eta_0)},$$

Also under $H_1 : K > K_0$, it holds

$$\mathbb{P}\{T_n > \chi^2_{k - \dim(\eta_0), \alpha}\} \to 1,$$

for the $(1 - \alpha)$th quantile $\chi^2_{k - \dim(\eta_0), \alpha}$ of the $\chi^2_{k - \dim(\eta_0)}$ distribution.

The assumptions (i)–(ii) for this theorem are standard for overidentified models. Identification of η_0 needs to be verified for each application (see, Theorem 1 of Bickel, Chen, and Levina 2011, for stochastic block models).

It should be noted that Theorem 6 applies to the case of $K_0 = 1$, where the null model becomes the Erdős-Rényi model containing only one free parameter. In this case, our goodness-of-fit test can be considered as a community detection test, which complements eigenvalue-based testing by Bickel and Sarkar (2016).

Although the goodness-of-fit test in Theorem 6 applies to any finite K in principle, it is computationally expensive to implement our test for large values of K, where we typically need to consider (k, l)-wheels with large k and/or l, and computation of those moments is demanding (see, e.g., Section 5 of Bickel, Chen, and Levina 2011). This practical issue applies to any statistical methods based on subgraph counting, and it is an important direction of future research to build an efficient algorithm to implement our JEL method; see Ribeiro et al. (2021) for recent advances in subgraph counting methods.

Furthermore, although it is also beyond the scope of this article, based on Andrews (1999) for consistent moment selection procedures, we conjecture that sequential testing based on the test statistic T_n for different values of K_0 with suitably chosen critical values (typically letting significance values converge to zero) may yield an asymptotically valid model selection procedure for the stochastic block models. This approach can be a complement to the likelihood-based model selection method developed in Wang and Bickel (2017).

Remark 7 (Plug-in statistic). When the dimension of the parameters η is high, the minimization in (21) to compute T_n may be computationally expensive. In such a scenario, if some consistent and computationally cheaper estimator $\widetilde{\eta}$ is available (e.g., Bickel et al. 2013), we can adapt the plug-in approach in Zheng, Zhao, and Yu (2012) to construct a goodness-of-fit statistic based on $\widetilde{\eta}$. More precisely, let $V^m_{Q,i}(\eta) = (V^{m(1)}_{Q,i}(\eta)', V^{m(2)}_{Q,i}(\eta)')'$, where $V^{m(1)}_{Q,i}(\eta)$ and $V^{m(2)}_{Q,i}(\eta)$ are $(k - \dim(\eta_0)) \times 1$ and $\dim(\eta_0) \times 1$, respectively. Then define

$$V^m_{*,i}(\widetilde{\eta}) = V^{m(1)}_{Q,i}(\widetilde{\eta}) - \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial V^{m(1)}_{Q,i}(\widetilde{\eta})}{\partial \eta'}\right) \left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial V^{m(2)}_{Q,i}(\widetilde{\eta})}{\partial \eta'}\right]^{-1} V^{m(2)}_{Q,i}(\widetilde{\eta}).$$

The goodness-of-fit statistic is obtained as

$$\widetilde{T}_n = 2\sup_{\lambda}\sum_{i=1}^{n}\log(1 + \lambda' V^m_{*,i}(\widetilde{\eta})).$$

Note that this statistic does not involve minimization for η as in (21). By applying the same argument in Zheng, Zhao, and Yu (2012), we can show that $\widetilde{T}_n$ is asymptotically equivalent to T_n.

3.4. Other Network Models

The goodness-of-fit testing approach in the previous section can be applied to other network models. Once we specify the function $w(\cdot, \cdot; \eta)$ in (9) with parameters η, we can take a set of subgraphs $(R_1, \dots, R_k)$ and characterize the moments $\theta(\eta) = (Q(R_1; \eta), \dots, Q(R_k; \eta))'$ implied from the model $w(\cdot, \cdot; \eta)$. Then the JEL goodness-of-fit statistic is obtained as in (20) and (21).

For example, Bhattacharyya and Bickel (2015, sec. 5.3) considered the function $w(u, v) = (1-u)^{-1/2}(1-v)^{-1/2}$ motivated by the preferential attachment model, where the $(m+1)$th vertex attaches to one of the preceding m vertices with probability proportional to degree. Other examples include $w(u, v) = \exp(u + v)/\{1 + \exp(u + v)\}$ based on the β-model (see, e.g., Chatterjee, Diaconis, and Sly 2011), and the random threshold graphs with $w(u, v) = \mathbb{I}\{F(u) + F(v) \geq \alpha\}$ for some cumulative distribution function F and $\alpha > 0$ (see, e.g., Diaconis and Janson 2008).

4. Simulation

This section conducts a simulation study to evaluate the finite sample properties of the JEL inference methods. In particular, we consider a stochastic block model with $K = 2$ equal-sized communities and the following edge probabilities

$$F_{ab} = P(A_{ij} = 1 | i \in a, j \in b) = s_n S_{ab}, \quad \text{for } a, b \in \{1, 2\}.$$

We set $S = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.4 \end{pmatrix}$ and vary s_n such that $\rho_n = 0.45 s_n = \pi' F \pi \in (0.1, 0.05, 0.02)$ with $\pi = (0.5, 0.5)'$. The cases of $\rho_n = 0.05$ or 0.02 may be considered as sparse networks. For example, Facebook networks considered in Section 5.2 exhibit $\rho_n \approx 0.02$.

We compare four methods to construct confidence intervals for (a) $P(R)$ where R is (1,2)-wheels, (b) $P(R)(= Q(R))$ where R is 3-cycles (or triangles), (c) $Q(R)$ where R is (1,2)-wheels, and (d) $Q(R)$ where R is (1,3)-wheels: (i) Wald-type confidence interval (Wald), which is defined as $[\widehat{\theta} \pm 1.96\widehat{\sigma}]$ with $\widehat{\sigma}^2 = \frac{n-1}{n}\sum_{i=1}^{n}(\widehat{\theta}^{(i)} - \widehat{\theta})^2$, (ii) bootstrap confidence interval (Boot), which is defined as $[\widehat{\theta} - c^*_{97.5}\widehat{\sigma}, \widehat{\theta} - c^*_{2.5}\widehat{\sigma}]$ with the αth percentile of the bootstrap approximation c^*_α based on the node resampling network bootstrap by Green and Shalizi (2017) with 199 bootstrap replications, (iii) jackknife empirical likelihood confidence interval (JEL) using the χ^2 critical value, and (iv) modified jackknife empirical likelihood confidence interval (mJEL).

Tables 1–4 give the empirical coverage rates and average lengths of the confidence intervals above for (a)–(d), respectively. The number of Monte Carlo replications is 1000 for (a)–(c), and 500 for (d). To highlight desirable finite sample performance of our JEL methods, we set the network size as $n = 400$, which is quite small in the context of network data analysis. Our empirical analysis in the next section considers larger networks such as Facebook data. The nominal rate is 0.95. Main findings from the simulation study are in line with our theoretical results. The Wald and JEL confidence intervals (using the normal and χ^2 critical values, respectively) tend to overcover especially when the network is sparse, which verifies our theoretical results mentioned in Remark 2.

The bootstrap-based intervals are more accurate than the Wald and JEL except for the case of $\rho_n = 0.02$. When $\rho_n = 0.02$, the bootstrap-based intervals severely under-cover for wheels, and over-cover for 3-cycles. On the other hand, the mJEL confidence intervals are most robust to the sparsity of

Table 1. Coverage rates and average lengths of 95% confidence intervals for $P(R)$ with $R = (1,2)$-wheel and $n = 400$.

	Coverage rates				Average interval lengths			
ρ_n	Wald	Boot	JEL	mJEL	Wald	Boot	JEL	mJEL
0.1	0.984	0.933	0.981	0.946	0.0013	0.0011	0.0012	0.0011
0.05	0.984	0.917	0.984	0.952	0.00046	0.00039	0.00045	0.00035
0.02	0.992	0.833	0.991	0.944	0.00012	0.00009	0.00012	0.00009

Table 2. Coverage rates and average lengths of 95% confidence intervals for $P(R)$ with $R =$ 3-cycle and $n = 400$.

	Coverage rates				Average interval lengths			
ρ_n	Wald	Boot	JEL	mJEL	Wald	Boot	JEL	mJEL
0.1	0.987	0.966	0.987	0.949	0.00025	0.00023	0.00025	0.00020
0.05	0.991	0.969	0.990	0.940	0.000045	0.000039	0.000046	0.000032
0.02	0.998	0.983	0.998	0.947	0.0000071	0.0000065	0.0000072	0.0000054

Table 3. Coverage rates and average lengths of 95% confidence intervals for $Q(R)$ with $R = (1,2)$-wheel and $n = 400$.

	Coverage rates				Average interval lengths			
ρ_n	Wald	Boot	JEL	mJEL	Wald	Boot	JEL	mJEL
0.1	0.980	0.942	0.980	0.946	0.0015	0.0014	0.0015	0.0013
0.05	0.982	0.924	0.985	0.954	0.00049	0.00042	0.00049	0.00038
0.02	0.992	0.848	0.990	0.942	0.00012	0.00010	0.00012	0.000087

Table 4. Coverage rates and average lengths of 95% confidence intervals for $Q(R)$ with $R = (1,3)$-wheel and $n = 400$.

	Coverage rates				Average interval lengths			
ρ_n	Wald	Boot	JEL	mJEL	Wald	Boot	JEL	mJEL
0.1	0.986	0.888	0.988	0.944	0.00095	0.00082	0.00095	0.00079
0.05	0.988	0.750	0.988	0.954	0.00016	0.00012	0.00016	0.00012
0.02	0.994	0.464	0.994	0.934	0.000016	0.000010	0.000016	0.000011

the network compared to the other intervals, and offer close-to-correct empirical coverages in all cases. Furthermore, in terms of the average lengths of the confidence intervals, the mJEL outperforms other methods for all cases. Overall, the modified JEL method exhibit excellent finite sample performances even for a very small sample size.

5. Real Data Example

5.1. Karate Club Network

We consider the well-known karate club data of Zachary (1977) which describe social network friendships between 34 members of a karate club at a U.S. university in the 1970s. It is known that the members split into two groups after a disagreement on class fees later (Zachary 1977). We assess whether this ground truth fact is formally supported by the observed network data using our method.

Let $W_{k,l}$ be a (k,l)-wheel. Under the null hypothesis that the data is generated from a stochastic block model with $K = 1$ (i.e., Erdős-Rényi model), the modified JEL test based on $Q(W_{11})$ and $Q(W_{12})$ gives the p-value of 0.0043 (the value of the statistic is 8.13) indicating strong evidence to reject the null hypothesis. On the other hand, the JEL and Wald statistics take rather smaller values (4.56 and 3.03, respectively). Under the χ^2 critical values which are valid in the mildly and moderately sparse cases, their p-values are 0.033 for the JEL and 0.082 for the Wald. Thus, the Wald test cannot reject the null of $K = 1$ at 5% significance level.

5.2. Facebook Network

Using Facebook network data by Rossi and Ahmed (2015), we test whether two network data share the same features in terms of the transitivity (i.e., $\vartheta = Q(R_1)/Q(R_2)$ or $P(R_1)/\{P(R_1) + P(R_2)\}$, where R_1 is a 3-cycle and R_2 is a (1,2)-wheel).

The modified JEL statistic is defined as $\min_\vartheta \ell_m^{\text{hom}}(\vartheta, \vartheta)$, where

$$\ell_m^{\text{hom}}(\vartheta_1, \vartheta_2) = 2 \min_{\theta_1 \in \{\theta_1 : h(\theta_1) = \vartheta_1\}} \sup_{\lambda_1} \sum_{l_1=1}^{n} \log(1 + \lambda_1' V_{1l_1}^m(\theta_1)) + 2 \min_{\theta_2 \in \{\theta_2 : h(\theta_2) = \vartheta_2\}} \sup_{\lambda_2} \sum_{l_2=1}^{m} \log(1 + \lambda_2' V_{2l_2}^m(\theta_2)),$$

$\theta_1 = (Q^1(R_1), Q^1(R_2))$, $\theta_2 = (Q^2(R_1), Q^2(R_2))$, $h(\theta) = Q(R_1)/Q(R_2)$, and $V_{1l_1}^m(\theta)$ and $V_{2l_2}^m(\theta)$ are defined as in (16) using $\{A_{ij}^1\}$ and $\{A_{ij}^2\}$, respectively. If $\{A_{ij}^1\}$ and $\{A_{ij}^2\}$ are independent, we have $\min_\theta \ell_m^{\text{hom}}(\vartheta, \vartheta) \xrightarrow{d} \chi_1^2$ under $H_0 : \vartheta_1 = \vartheta_2$ and analogous conditions to Theorem 4.

Transitivity in this example refers to the case where the friend of your friend is also a friend of yours, and is of substantial interest in the social network literature (see, e.g., Section 7 of Graham 2020). We consider two college pairs: (Williams, Wellesley) and (Rice, Johns Hopkins). Table 3 gives a summary of the networks we used. Williams and Wellesley are chosen as examples of strong liberal arts colleges with relatively small number of students. We choose Rice and Johns Hopkins since

Table 5. Summary of college networks we used.

	Williams	Wellesley	Rice	Johns Hopkins
Nodes	2790	2970	4087	5180
Edges	112986	94899	184828	186586
Densities	0.0290	0.0215	0.0221	0.0139
Transitivities	0.2075	0.1727	0.2032	0.1932

they have similar numbers of nodes and edges in our data as shown in Table 3.

For Williams and Wellesley, under the null hypothesis that the two network data share the same features in terms of the transitivity, the p-values are 7.8770e-13 and 1.1957e-13 for the JEL and modified JEL tests, respectively (the values of the statistics are 51.31 and 55.01, respectively). Thus, both tests indicate strong evidence to reject the null hypothesis (Table 5).

On the other hand, for Rice and Johns Hopkins, the JEL test gives the p-value of 0.053 (the value of the test statistic is 3.74) and hence cannot reject the null at 5% level, while the modified JEL gives the p-value of 0.048 (the value of the test statistic is 3.90) and delivers marginal significance at the 5% level.

6. Conclusion

In this article, we extend the methodology of jackknife empirical likelihood for network moments, and study its asymptotic properties for different types of sparsity. We find that the jackknife empirical likelihood statistic for acyclic or cyclic subgraph moments loses its asymptotic pivotalness in the severely or moderately sparse case, respectively, and a modification is possible to recover pivotalness under those sparsity conditions. Several applications for specification, two-sample, or goodness-of-fit testing and numerical examples illustrate usefulness of the proposed approach, particularly the modified jackknife empirical likelihood statistic.

There are several directions of future research. First, it is interesting to extend our method to cover more general network models, such as degree-corrected block models (Karrer and Newman 2011) and inhomogeneous Erdős-Rényi models (Bollobás, Janson, and Riordan 2007), which are not covered by the nonparametric latent variable model considered in this article. Second, it is important to generalize our framework to accommodate covariates. Third, as mentioned in a remark of Theorem 6, it is worthwhile to study use of the (modified) jackknife empirical likelihood statistics for different hypotheses to construct a valid model selection procedure.

Supplementary Materials

The supplementary material contains the proofs for the proposition and theorems in this article.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

Our research is supported by the JSPS KAKENHI (26780133, 18K01541) (Matsushita) and ERC Consolidator Grant (SNP 615882) (Otsu).

References

Anandkumar, A., Ge, R., Hsu, D., and Kakade, S. (2013), "A Tensor Spectral Approach to Learning Mixed Membership Community Models," in *Proceedings of the 26th Annual Conference on Learning Theory* (Vol. 30), pp. 867–881. [2]

Andrews, D. W. K. (1999), "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation, *Econometrica*, 67, 543–564. [9]

Bhattacharyya, S., and Bickel, P. J. (2015), "Subsampling Bootstrap of Count Features of Networks," *Annals of Statistics*, 43, 2384–2411. [1,2,3,4,7,8,9]

Bickel, P. J., and Chen, A. (2009), "A Nonparametric View of Network Models and Newman-Girvan and Other Modularities," *Proceedings of the National Academy of Sciences*, 106, 21068–21073. [1]

Bickel, P. J., Chen, A., and Levina, E. (2011), "The Method of Moments and Degree Distributions for Network Models," *Annals of Statistics*, 39, 2280–2301. [1,2,3,4,8,9]

Bickel, P., Choi, D., Chang, X., and Zhang, H. (2013), "Asymptotic Normality of Maximum Likelihood and its Variational Approximation for Stochastic Blockmodels," *Annals of Statistics*, 41, 1922–1943. [1,9]

Bickel, P. and P. Sarkar (2016), "Hypothesis testing for automated community detection in networks," *Journal of the Royal Statistical Society, Series B*, 78, 253–273. [9]

Bollobás, B., Janson, S., and Riordan, O. (2007), "The Phase Transition in Inhomogeneous Random Graphs," *Random Structures and Algorithms*, 31, 3–122. [11]

Borgs, C., Chayes, J. and A. Smith (2015), "Private graphon estimation for sparse graphs," in Advances in Neural Information Processing Systems, 1369–1377. Nips Foundation. Available at http://books.nips.cc. [4]

Chatterjee, S., Diaconis, P., and Sly, A. (2011), "Random Graphs with a Given Degree Sequence," *Annals of Applied Probability*, 21, 1400–1435. [1,9]

Crane, H. (2018), *Probabilistic Foundations of Statistical Network Analysis*, Boca Raton, FL: CRC Press. [1]

Diaconis, P., Holmes, S., and Janson, S. (2008), "Threshold Graph Limits and Random Threshold Graphs," *Internet Mathematics*, 5, 267–320. [1,9]

Diaconis, P., and Janson, S. (2008), "Graph Limits and Exchangeable Random Graphs," *Rendiconti di Matematica*, 28, 33–61. [1,9]

Efron, B., and Stein, C. (1981), "The Jackknife Estimate of Variance," *Annals of Statistics*, 9, 586–596. [1,2,4,5,6]

Graham, B. S. (2020), "Network Data," in *Handbook of Econometrics* (Vol. 7A), eds. S. N. Durlauf, L. Peter Hansen, J. J. Heckman, and R. L. Matzkin, Amsterdam: Elsevier, pp. 111–218. [10]

Green, A., and Shalizi, C. R. (2017), "Bootstrapping Exchangeable Random Graphs," working paper. [2,9]

Hall, P., and La Scala, B. (1990), "Methodology and Algorithms of Empirical Likelihood," *International Statistical Review*, 58, 109–127. [7]

Hinkley, D. V. (1978), "Improving the jackknife with Special Reference to Correlation Estimation," *Biometrika*, 65, 13–21. [2]

Hoff, P., Raftery, A., and Handcock, M. (2002), "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 97, 1090–1098. [1]

Jing, B. Y., Yuan, J., and Zhou, W. (2009), "Jackknife Empirical Likelihood," *Journal of the American Statistical Association*, 104, 1224–1232. [1,3]

Kallenberg, O. (2005), *Probabilistic Symmetries and Invariance Principles*, New York: Springer. [1,3]

Karrer, B., and Newman, M. E. J. (2011), "Stochastic Blockmodels and Community Structure in Networks," *Physical Review,* 83, 016107. [3,11]

Klopp, O., Tsybakov, A. B. and Verzelen, N. (2017), "Oracle inequalities for network models and sparse graphon estimation," *The Annals of Statistics*, 45, 316–354. [4]

Kolaczyk, E. D. (2009), *Statistical Analysis of Network Data*, New York: Springer. [1,3]

Levin, K., and Levina, E. (2019), "Bootstrapping Networks with Latent Space Structure," working paper. [2]

Lin, Q., Lunde, R., and Sarkar, P. (2020a), "On the Theoretical Properties of the Network Jackknife," working paper. [2,3,4]

——— (2020b), "Higher-Order Correct Multiplier Bootstraps for Count Functionals of Networks," working paper. [2,4]

Matsushita, Y., and Otsu, T. (2020), "Jackknife Empirical Likelihood: Small Bandwidth, Sparse Network and High-Dimension Asymptotics," *Biometrika*, forthcoming. [1,5]

Menzel, K. (2018), "Bootstrap with Cluster-Dependence in Two or More Dimensions," working paper. [5]

Owen, A. B. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249. [1,3]

Ribeiro, P., Paredes, P., Silva, M. E. P., Aparicio, D., and Silva, F. (2021), "A Survey on Subgraph Counting: Concepts, Algorithms, and Applications to Network Motifs and Graphlets," *ACM Computing Surveys*, 54, 1–36. [9]

Rossi, R. A., and Ahmed, N. K. (2015), "The Network Data Repository with Interactive Graph Analytics and Visualization," *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, available at *http://networkrepository.com*. [10]

Tukey, J. W. (1958), "Bias and Confidence in Not-Quite Large Samples," *The Annals of Mathematical Statistics*, 29, 614. [3]

Wang, Y. X. R., and Bickel, P. J. (2017), "Likelihood-based Model Selection for Stochastic Block Models," *Annals of Statistics*, 45, 500–528. [9]

Wu, C. and Y. Yan (2012), "Empirical likelihood inference for two-sample problems," *Statistics and Its Interface*, 5, 345–354. [8]

Zachary, W. (1977), "An Information Flow Model for Conflict and Fission in Small Groups," *Journal of Anthropological Research*, 33, 452–473. [10]

Zhang, Y. and D. Xia (2022), "Edgeworth expansions for network moments," *The Annals of Statistics*, 50, 726–753. [7]

Zheng, M., Zhao, Z. and W. Yu (2012), "Empirical likelihood methods based on influence functions," *Statistics and Its Interface*, 5, 355–366 [9]