EMPIRICAL LIKELIHOOD FOR NETWORK DATA

YUKITOSHI MATSUSHITA AND TAISUKE OTSU

Abstract. Analysis on network data is becoming increasingly important in various fields of data science, and the literature on statistical modelling and estimation algorithms for networks is rapidly growing. However, general statistical inference methods for networks are still less developed. This article develops concept of nonparametric likelihood for network data based on the 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 loses its asymptotic pivotalness under the sparse network asymptotics and develop a modified statistic which converges to a chi-squared distribution under both the sparse and dense network asymptotics.

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 modelling and estimation for network data (see, Crane, 2018, for a survey on recent developments). In particular, based on the Aldous-Hoover representation result 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, e.g., 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 modelling 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 concept of nonparametric likelihood for network data based on the network moments, and proposes general inference methods by adapting the theory of jackknife empirical likelihood. The method of jackknife empirical likelihood proposed by Jing, Yuan and Zhou (2009) 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

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

introduce its jackknife pseudo values by using delete-one vertex subgraphs, construct an empirical likelihood function, and study its asymptotic properties under the possibly sparse network model of Bickel and Chen (2009). 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 paper makes two contributions. First, we show that the jackknife empirical likelihood statistic loses its asymptotic pivotalness and converges to a weighted chi-squared distribution under the sparse network asymptotics, where the edge formation probability ρ_n is of order $O(n^{-1})$ for the number of vertices n. We argue that this lack of asymptotic pivotalness is understood as emergence of Efron and Stein's (1981) bias of the jackknife variance estimator in the first order. Second, we develop a modified empirical likelihood statistic to recover asymptotic pivotalness and converges to a chi-squared distribution under both the dense and sparse network asymptotics (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.

Recently several authors proposed inference methods for network data. 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 paper proposes a nonparametric likelihood-based inference method based on jackknife empirical likelihood. Also we emphasize that this paper considers inference under more general conditions on the sparsity level. In particular, the above papers exclude the case of $\rho_n = O(n^{-1})$, which is analyzed in Bickel, Chen and Levina (2011) as a sparse case but inference for this case is an open question.

This paper is organized as follows. In Section 2, we introduce our setup and jackknife empirical likelihood for network moments, derive its asymptotic properties and develop a modified statistic to recover asymptotic pivotalness for the scalar case (Section 2.1) and vector case (Section 2.2). Section 3 presents goodness-of-fit testing based on our empirical likelihood approach, which covers block size testing for stochastic block models. Sections 4 and 5 illustrate our methodology by a simulation study and real data example, respectively. 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 $R = \{(i,j) : 1 \le i < j \le n\}$ 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 are interested in occurrence probability of R, that is

$$P(R) = \mathbb{P}\{\mathcal{E}(G_n(R)) = R\}.$$

Let $\mathrm{Iso}(R)$ be the set of subgraphs that are isomorphic to R in G_n and $|\mathrm{Iso}(R)|$ be its number of elements.[1] Bickel, Chen and Levina (2011) proposed to estimate $P(R)$ by

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

where p is the numbers of vertices and isomorphic subgraphs of R, $\mathcal{G}$ is the set of all subgraphs of G_n, and $\mathbb{I}\{\cdot\}$ is the indicator function. Obviously $\hat{P}(R)$ is unbiased for $P(R)$, and Bickel, Chen and Levina (2011) established the asymptotic distribution of $\hat{P}(R)$.

In this paper, we regard $\hat{P}(R) - P(R)$ as an estimating equation for $P(R)$ and construct the jackknife empirical likelihood function to conduct inference on $P(R)$. More precisely, we introduce the jackknife pseudo value:

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

where $\hat{P}_{-i}(R)$ is the leave-i counterpart of $\hat{P}(R)$ defined as

$$\hat{P}_{-i}(R) = \frac{1}{\binom{n-1}{p} N(R)} \sum_{S \in \mathcal{G}_i} \mathbb{I}\{S \sim R\},$$

and $\mathcal{G}_i$ is the set of all subgraphs that do not contain the i-th vertex. Since $\frac{1}{n}\sum_{i=1}^n (V_i - P(R)) = \hat{P}(R) - P(R)$, the contrast $V_i - P(R)$ can be employed as an estimating function for $P(R)$.

More generally, for fixed sets $\{R_1, \dots, R_k\}$, we can analogously define the estimators $(\hat{P}(R_1), \dots, \hat{P}(R_k))$ and the vector of jackknife pseudo values $V_i = (V_{1i}, \dots, V_{ki})'$ for $\theta = (P(R_1), \dots, P(R_k))'$. Based on this notation, we construct the jackknife empirical likelihood function for θ as

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

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

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

[1] 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 $V(R_1)$ to $V(R_2)$ such that the map $(i,j) \to (\sigma_i, \sigma_j)$ is one-to-one from $E(R_1)$ to $E(R_2)$.

In practice, we use this dual form to implement empirical likelihood inference. In the next subsections, we study asymptotic properties of the empirical likelihood statistic $\ell(\theta)$ for the cases where θ is scalar (Section 2.1) and θ is a vector (Section 2.2).

2.1. **Case of scalar θ.** This subsection considers the case of $d = 1$, where θ, $P(R)$, $\hat{P}(R)$, and V_i are scalar and the empirical likelihood function is written as $\ell(\theta) = 2\sup_\lambda \sum_{i=1}^n \log(1+\lambda(V_i-\theta))$.

We first note that the estimator $\hat{P}(R)$ in (1) can be written as

$$\hat{P}(R) = \frac{1}{\binom{n}{p}} \sum_{1 \le i_1 < \cdots < i_p \le n} Y_{i_1 \ldots i_p}(R),$$

where

$$Y_{i_1 \ldots i_p}(R) = \frac{1}{N(R)} \sum_{S \sim R} \prod_{(i_k, i_l) \in S} A_{i_k i_l} \prod_{(i_k, i_l) \in \bar{S}} (1 - A_{i_k i_l}),$$

is a jointly exchangeable array and $\bar{S} = \{(i,j) \notin S, i \in V(S), j \in V(S)\}$. 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 $Y_{ijl}(R) = \frac{1}{3}\{A_{ij}A_{jl}(1 - A_{il}) + A_{ij}(1 - A_{jk})A_{il} + (1 - A_{ij})A_{jl}A_{il}\}$.

Let $\rho_n = \mathbb{P}\{A_{ij} = 1\}$ be the edge occurrence probability. In our setup, the parameter $P(R)$ typically depends on n, and note that $d_n = (n-1)\rho_n$ is the expected degree. As in Bickel, Chen and Levina (2011), this paper is mainly concerned with the case of $\rho_n \to 0$, and we call networks with $n\rho_n \to \infty$ and $n\rho_n \to C \in (0, \infty)$ as dense and sparse networks, respectively.

To study the asymptotic properties of $\hat{P}(R)$ and $\ell(\theta)$, we employ the nonparametric latent variable model in Bickel, Chen and Levina (2011) and Bhattacharyya and Bickel (2015):

$$A_{ij} = \mathbb{I}\{\xi_{ij} \le \rho_n w(\xi_i, \xi_j) \wedge 1\}, \tag{3}$$

for $i, j \in \{1, \ldots, n\}$, where $(\xi_1, \ldots, \xi_n, \xi_{11}, \ldots, \xi_{nn})$ are iid $U(0,1)$, and $w(\cdot, \cdot)$ is positive, symmetric, and $\int_0^1 \int_0^1 w(s,t)dsdt = 1$. This model is derived from a general representation theorem of the adjacency matrix A (Bickel and Chen, 2009) 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).

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 (3), there exists a measurable function $f : [0,1]^{1+p+(p-1)p/2} \to [0,1]$ such that

$$Y_{i_1 \ldots i_p}(R) = f(\mu, \xi_{i_1}, \ldots, \xi_{i_p}, \xi_{i_1 i_2}, \ldots, \xi_{i_{p-1} i_p}),$$

for each $(i_1, \dots, i_p) \in \bar{\mathbb{N}}$, where $\mu \sim U(0,1)$ corresponds to the mixing distribution in de Finetti's theorem and is not identifiable. To proceed, we introduce some notation. Let

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

Note that $Y_{1\dots p} = \mathbb{E}[Y_{1\dots p}|\xi_1, \dots, \xi_p, \xi_{12}, \dots, \xi_{p-1,p}]$. Let $|R| = |E(R)|$ 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

$$\rho_n^{-|R|}\{\hat{P}(R) - P(R)\} = \frac{1}{n}\sum_{i=1}^{n} \beta_i + \frac{1}{n^2}\sum_{i_1 < i_2} \beta_{i_1 i_2} + \dots + \frac{1}{n^p}\sum_{i_1 < \dots < i_p} \beta_{i_1 \dots i_p}, \tag{4}$$

where

$$\beta_i = \rho_n^{-|R|} p g_1(i), \ \beta_{i_1 i_2} = \rho_n^{-|R|} 2! \binom{p}{2} g_2(i_1 i_2), \dots, \ \beta_{i_1 \dots i_p} = \rho_n^{-|R|} p! g_p(i_1 \dots i_p).$$

Note that all the random variables on the right side of (4) have zero mean and no correlation.

This paper focuses on the cases, where (I) R is a wheel, and (II) R is a cyclic graph.[2] In Section A.1, we show that for Case (I), it holds

$$\frac{1}{n}\sum_{i=1}^{n} \beta_i = O_p\left(\frac{1}{\sqrt{n}}\right), \qquad \frac{1}{n^s}\sum_{i_1 < \dots < i_s}^{n} \beta_{i_1 \dots i_s} = O_p\left(\frac{1}{\sqrt{n^s \rho_n^{s-1}}} \vee \frac{1}{\sqrt{n^s}}\right), \tag{5}$$

for $s = 2, \dots, p$. Thus, for dense networks (i.e., $n\rho_n \to \infty$), the linear term $\frac{1}{n}\sum_{i=1}^{n} \beta_i$ will be a leading term in (4). On the other hand, for Case (II), it holds

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \beta_i = O_p\left(\frac{1}{\sqrt{n}}\right), \\
&\frac{1}{n^s}\sum_{i_1 < \dots < i_s}^{n} \beta_{i_1 \dots i_s} = O_p\left(\frac{1}{\sqrt{n^s \rho_n^{s-1}}} \vee \frac{1}{\sqrt{n^s}}\right), \quad \text{for } s = 2, \dots, p-1, \\
&\frac{1}{n^p}\sum_{i_1 < \dots < i_p}^{n} \beta_{i_1 \dots i_p} = O_p\left(\frac{1}{\sqrt{n^p \rho_n^{p}}} \vee \frac{1}{\sqrt{n^p}}\right).
\end{aligned} \tag{6}$$

The difference of the orders for $s = p$ is due to different orders of the variances in the main term of $\beta_{i_1 \dots i_p}$. For Case (II), we need the condition $n\rho_n \to \infty$ for the consistency, $\rho_n^{-|R|}\{\hat{P}(R) - P(R)\} \xrightarrow{p}$

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

0, due to the last term in (6). Hence $\frac{1}{n^s}\sum_{i_1<\cdots<i_s}^{n}\beta_{i_1\ldots i_s}=o_p\left(\frac{1}{n}\sum_{i=1}^{n}\beta_i\right)$ 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 $\hat{P}(R)$ 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 $\hat{P}(R)$ is determined by the linear and last terms.

The limiting distribution of the jackknife empirical likelihood $\ell(\theta)$ is obtained as follows. 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

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

Theorem 1. *Consider the setup of this section. Suppose $\rho_n\to 0$.*

Case (I): *If R is a wheel, then*

$$\ell(\theta)\xrightarrow{d}\begin{cases}\chi_1^2 & \text{under } n\rho_n\to\infty \text{ and } \mathbb{E}[\beta_1|\xi_1] \text{ is random},\\ \sigma_*^{-2}\chi_1^2 & \text{otherwise.}\end{cases}$$

Case (II): *If R is cyclic and $n\rho_n\to\infty$, then*

$$\ell(\theta)\xrightarrow{d}\begin{cases}\chi_1^2 & \text{under } n^{p-1}\rho_n^p\to\infty \text{ and } \mathbb{E}[\beta_1|\xi_1] \text{ is random},\\ \sigma_*^{-2}\chi_1^2 & \text{otherwise.}\end{cases}$$

This theorem shows that the limiting distribution of the jackknife empirical likelihood statistic $\ell(\theta)$ depends on the behavior of the expected degrees $n\rho_n$. If the network is dense in the sense that $n\rho_n\to\infty$ (for Case (I)) or $n^{p-1}\rho_n^p\to\infty$ (for Case (II)) and the term $\mathbb{E}[\beta_1|\xi_1]$ is random, then the jackknife empirical likelihood statistic is asymptotically pivotal. However, for sparse networks with $n\rho_n\nrightarrow\infty$ and possibly degenerate $\mathbb{E}[\beta_1|\xi_1]$, the jackknife empirical likelihood statistic is no longer asymptotically pivotal and its limiting distribution depends on $\sigma_*^2=\lim_{n\to\infty}(\sigma_n^2/\omega_n)$. It is interesting to note that the discrepancies of the coefficients in σ_n^2 and ω_n can be understood as 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 under the sparse network asymptotics.

The case where $\mathbb{E}[\beta_1|\xi_1]$ becomes random corresponds to non-degeneracy 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 non-degenerate distribution, where the conditional means given ξ_1 happen to be close to constant. We note that this degeneracy yields a non-standard limiting distribution of $\ell(\theta)$ only when $\rho_n=O(1)$ (i.e., the network is very dense), which is excluded in the above theorem. In particular, the terms of order $O_p\left(\sqrt{\frac{1}{n^s}}\right)$ in (5) and (6) will induce non-standard limiting behaviors.

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

$$
\begin{aligned}
M_{i_1\dots i_p} &= n\hat{P}(R) - (n-1)\left(\sum_{i=1}^{p} \hat{P}_{-i}(R)\right) + (n-2)\left(\sum_{i_1<i_2}^{p} \hat{P}_{-i_1,-i_2}(R)\right) + \cdots \\
&\quad +(-1)^p(n-p)\hat{P}_{-i_1,\dots,-i_p}(R).
\end{aligned}
$$

These terms are used in Efron and Stein (1981) to correct the higher-order bias of the jackknife variance estimator. Since $M_{i_1\dots i_p}$ is asymptotically expressed as a function of $\beta_{i_1\dots i_p}$'s but not $\beta_{i_1\dots i_s}$'s with $s<p$ (see, proof of Lemma 1), 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 jackknife empirical likelihood statistic as

$$
\ell^m(\theta) = 2\sup_{\lambda} \sum_{i=1}^{n} \log(1+\lambda V_i^m(\theta)), \tag{7}
$$

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

$$
\begin{aligned}
\hat{\Gamma} &= \sqrt{\sum_{i=1}^{n}(V_i-\hat{\theta})^2}, \qquad (8)\\
\tilde{\Gamma} &= \sqrt{\sum_{i=1}^{n}(V_i-\hat{\theta})^2 - \sum_{i_1<i_2}^{n} M_{i_1 i_2}^2 - \cdots - (-1)^p \sum_{i_1<\cdots<i_p}^{n} M_{i_1\dots i_p}^2}.
\end{aligned}
$$

The asymptotic property of the modified jackknife empirical likelihood statistic is obtained as follows.

Theorem 2. *Consider the setup of this section. Suppose $\rho_n \to 0$.*
Case (I): *If R is a wheel, then $\ell_m(\theta) \xrightarrow{d} \chi_1^2$.*
Case (II): *If R is cyclic and $n\rho_n \to \infty$, then $\ell_m(\theta) \xrightarrow{d} \chi_1^2$.*

This theorem says that the modified jackknife empirical likelihood statistic $\ell_m(\theta)$ is asymptotically pivotal and converges to the χ_1^2 distribution regardless of sparsity of the network (i.e., $n\rho_n \to \infty$ or not for Case (I) and $n^{p-1}\rho_n^p \to \infty$ or not for Case (II)). Also, these limiting behaviors are robust to degeneracy of the component $\mathbb{E}[\beta_1|\xi_1]$. Based on this result, the asymptotic $1-\alpha$ confidence set for θ can be obtained as $ELCI_\alpha = \{\theta : \ell_m(\theta) \le \chi_{1,\alpha}^2\}$ for the $1-\alpha$ critical value of the χ_1^2 distribution.

2.2. **Case of vector θ.** For a vector case, we can apply the decomposition in (4) for each element in the vector $(\hat{P}(R_1) - P(R_1), \dots, \hat{P}(R_k) - P(R_k))$ with corresponding components $\{\beta_i^{(j)}, \dots, \beta_{i_1\dots i_{p_j}}^{(j)}\}$ for $j = 1,\dots,k$. Define $\sigma_{s,n}^{(j,h)2} = \mathbb{E}[\beta_{i_1\dots i_s}^{(j)}\beta_{i_1\dots i_s}^{(h)}]$ for $s = 1,\dots,p$, and

$$
\begin{aligned}
\Omega_n &= k\times k \text{ matrix with } (j,h)\text{-th element } \frac{\sigma_{1,n}^{(j,h)2}}{n} + \frac{\sigma_{2,n}^{(j,h)2}}{2n^2} + \frac{\sigma_{3,n}^{(j,h)2}}{6n^3} + \cdots + \frac{\sigma_{p_j\wedge p_h,n}^{(j,h)2}}{p!n^{p_j\wedge p_h}}, \qquad (9)\\
\Sigma_n &= k\times k \text{ diagonal matrix with } (j,h)\text{-th element } \frac{\sigma_{1,n}^{(j,h)2}}{n} + \frac{\sigma_{2,n}^{(j,h)2}}{n^2} + \frac{\sigma_{3,n}^{(j,h)2}}{2n^3} + \cdots + \frac{\sigma_{p_j\wedge p_h,n}^{(j,h)2}}{(p-1)!n^{p_j\wedge p_h}}.
\end{aligned}
$$

Based on the above notation, the limiting distribution of the jackknife empirical likelihood statistic $\ell(\theta)$ in (2) is obtained as follows. To simplify the presentation, we only present the result for Case (I).

Theorem 3. *Consider the setup of this section. Suppose $\rho_n \to 0$, and Ω_n and Σ_n are positive definite for all n large enough. If $(R_1, \ldots, R_k)$ are wheels, 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 jackknife empirical likelihood statistic is not asymptotically pivotal and depends on the unknown component Σ^*. When $n\rho_n \to \infty$ and $\mathbb{E}[\beta_1^{(j)}|\xi_1]$ is random for all $j = 1, \ldots, k$, we can recover asymptotic pivotalness as $\ell(\theta) \xrightarrow{d} \chi^2_k$. The discrepancies of the coefficients in 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.

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 jackknife empirical likelihood statistic is modified as follows

$$\ell^m(\theta) = 2 \sup_\lambda \sum_{l=1}^n \log(1 + \lambda' V_i^m(\theta)), \tag{10}$$

where $V_i^m(\theta) = (V_i - \hat{\theta}) + \hat{\Gamma}\tilde{\Gamma}^{-1}(\hat{\theta} - \theta)$ with $\hat{\theta} = (\hat{P}(R_1), \ldots, \hat{P}(R_k))'$, and $\hat{\Gamma}$ and $\tilde{\Gamma}$ are given by

$$\begin{aligned}
\hat{\Gamma}\hat{\Gamma}' &= \sum_{i=1}^n (V_i - \hat{\theta})(V_i - \hat{\theta})', \\
\tilde{\Gamma}\tilde{\Gamma}' &= k \times k \text{ matrix with } (j,h)\text{-th element} \\
&\quad \sum_{i=1}^n (V_i^{(j)} - \hat{\theta}^{(j)})(V_i^{(h)} - \hat{\theta}^{(h)}) - \sum_{i_1 < i_2} M^{(j)}_{i_1 i_2} M^{(h)}_{i_1 i_2} - \cdots - (-1)^{p_j \wedge p_h} \sum_{i_1 < \cdots < i_{p_j \wedge p_h}} M^{(j)}_{i_1 \ldots i_{p_j \wedge p_h}} M^{(h)}_{i_1 \ldots i_{p_j \wedge p_h}}.
\end{aligned}$$

The asymptotic property of the modified jackknife empirical likelihood statistic is obtained as follows.

Theorem 4. *Consider the setup of this section. Suppose $\rho_n \to 0$, and Ω_n and Σ_n are positive definite for all n large enough. If $(R_1, \ldots, R_k)$ are wheels, 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 for the vector θ case, the modified jackknife empirical likelihood statistic $\ell_m(\theta)$ is asymptotically pivotal and converges to the χ^2_k distribution regardless of sparsity of the network.

Finally we note that this theorem can be modified to deal with the case where the object of interest is a smooth function of θ, say $\vartheta = h(\theta)$. One important example is 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, an analogous argument to Hall and La Scala (1990, Theorem 2.1) yields

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

3. Goodness-of-fit tests

In this section, we develop a goodness-of-fit test using empirical likelihood.

3.1. **Stochastic block model.**

Consider the function $w(\cdot,\cdot)$ in (3) corresponding to a K-block model defined by parameters $\eta \equiv (\pi, \rho_n, S)$, where π is a $K \times 1$ vector of probabilities for block assignment, and a $K \times K$ matrix S satisfies

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

for $a, b = 1, \dots, K$. The number of free parameters in the block model is $K - 1$ for π and $K(K+1)/2$ for F. Note that ρ_n is determined by $\sum_{a=1}^{K} \sum_{b=1}^{K} F_{ab} = \rho_n$. For example, when $K = 3$, the number of free parameter is 8, which can be identified by 8 moments. In this subsection, 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 $Q(R) = \mathbb{P}\{A_{ij} = 1, \text{for all } (i,j) \in R\}$. Note that $Q(R)$ is written by using $P(\cdot)$ as

$$Q(R) = \sum \{P(S) : S \supset R, \mathcal{V}(S) = \mathcal{V}(R)\},$$

where $R \subset S$ refers to $S \subset \{(i,j) : i,j \in \mathcal{V}(R)\}$ (see, Proposition 1 of Bickel, Chen and Levina, 2011).

From Bickel, Chen and Levina (2011, Theorem 2), stochastic block models are generally identified by some set of wheels. Therefore, this subsection focuses on the case where R's are wheels. If the graph R is a (k,l)-wheel, it can be written as

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

Based on this formula, we can compute the moment, say $Q(R;\eta)$, implied from a given $\eta = (\pi, \rho_n, S)$.[3]

For wheels $(R_1,\dots,R_k)$, we consider the estimator $\hat{Q}(R_j) = \sum\{\hat{P}(S) : S \supset R_j, \mathcal{V}(S) = \mathcal{V}(R_j)\}$ of the moment $Q(R_j,\eta)$ for $j = 1,\dots,k$. Then the jackknife pseudo value can be defined as

$$V_{ji} = n\hat{Q}(R_j) - (n-1)\hat{Q}_{-i}(R_j),$$

for $j = 1,\dots,k$. Then the modified jackknife empirical likelihood function $\ell_m(\eta)$ can be defined as in (10) by setting

$$V_i^m(\eta) = (V_i - \hat{\theta}) + \hat{\Gamma}\tilde{\Gamma}^{-1}(\hat{\theta} - \theta(\eta)), \tag{11}$$

where $V_i = (V_{1i},\dots,V_{ki})'$, $\hat{\theta} = (\hat{Q}(R_1),\dots,\hat{Q}(R_k))'$, $\theta(\eta) = (Q(R_1;\eta),\dots,Q(R_k;\eta))'$, and $\hat{\Gamma}$ and $\tilde{\Gamma}$ are defined as in (8) by replacing $\{\hat{P}(R_1),\dots,\hat{P}(R_k)\}$ with $\{\hat{Q}(R_1),\dots,\hat{Q}(R_k)\}$. Then the goodness-of-fit statistic based $\ell_m(\theta)$ on is defined as

$$T_n = \min_{\eta\in\Upsilon} \ell_m(\eta), \tag{12}$$

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

Theorem 5. *Consider the setup of this subsection. Assume (i) there exists a unique $\eta_0 \in int(\Upsilon)$ such that $Q(R_j) = Q(R_j;\eta_0)$ is satisfied for all $j = 1,\dots k$, 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) Ω_n and Σ_n defined in (9) 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 $1-\alpha$ quantile of the $\chi^2_{k-\dim(\eta_0),\alpha}$ distribution.

The proof is similar to that of Theorem 4 with modifications for the overidentified case as in Qin and Lawless (1994, Corollary 4). Assumptions are standard for overidentified models. Identification of η_0 needs to be verified for each application (see, Theorem 2 of Bickel, Chen and Levina, 2011, for stochastic block models).

3.2. **Other network models.** The goodness-of-fit testing approach in the previous subsection can be applied to other network models. Once we specify the function $w(\cdot,\cdot;\eta)$ in (3) with parameters η, we can take a set of subgraphs $(R_1,\dots,R_k)$ and characterize the moments

[3]For example, when $K = 2$, it can be written as $T(1)(\xi) = v_1$ for $\xi \in [0,\pi_1]$ and v_2 for $\xi \in (\pi_1, 1]$, where $v_j = \pi_1 F_{1j} + (1-\pi_1)F_{2j}$ with $v_1 < v_2$. Let W_{kl} be a (k,l)-wheel. Thus the first three moments of $T(1)(\xi)$ are

$$\mathbb{E}[\{T(1)(\xi)\}^l] = \mathbb{E}[Q(W_{1l})] = \pi_1 v_1^l + (1-\pi_1)v_2^l,$$

for $l = 1,2,3$. Similarly, we have $T(1)^2(\xi) = \pi_1 v_1 F_{11} + (1-\pi_1)v_1 F_{21}$ for $\xi \in [0,\pi_1]$ and $\pi_1 v_1 F_{12} + (1-\pi_1)v_2 F_{22}$ for $\xi \in (\pi_1, 1]$. Thus, the first three moments of $T(1)^2(\xi)$ are

$$\mathbb{E}[\{T(1)^2(\xi)\}^l] = \mathbb{E}[Q(W_{2l})] = \pi_1\{\pi_1 v_1 F_{11} + (1-\pi_1)v_2 F_{21}\}^l + (1-\pi_1)\{\pi_1 v_1 F_{12} + (1-\pi_1)v_2 F_{22}\}^l.$$

$\theta(\eta) = (Q(R_1;\eta), \ldots, Q(R_k;\eta))'$ implied from the model $w(\cdot,\cdot;\eta)$. Then the jackknife empirical likelihood goodness-of-fit statistic is obtained as in (11) and (12).

For example, the preferential attachment model, where the $m+1$-th vertex attaches to one of the preceding m vertices with probability proportional to degree, may be tested by setting (see, Section 5.3 of Bickel, Chen and Levina, 2011)

$$w(u,v) = (1-u)^{-1/2}(1-v)^{-1/2}.$$

Other examples include the β-model with $w(u,v) = \exp(u+v)/\{1+\exp(u+v)\}$ (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, Holmes and Janson, 2008).

4. Simulation

This section conducts a simulation study to evaluate the finite sample properties of the jackknife empirical likelihood inference methods. In particular, we focus on the jackknife empirical likelihood inference under the sparse network asymptotics in Section 2, and consider a stochastic block model with $K = 2$ equal-sized communities and the following edge probabilities

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

We set $S = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.4 \end{pmatrix}$ and vary s_n such that $\theta_n = \pi' F \pi \in (0.5, 0.1, 0.05)$ with $\pi = (0.5, 0.5)'$. The network size is $n = 100$.

We compare four methods to construct confidence intervals for (i) $(1,2)$-wheels and (ii) 3-cycles (or triads or triangles): (i) Wald-type confidence interval (Wald), which is defined as $[\hat{\theta} \pm 1.96\hat{\sigma}]$ with $\hat{\sigma}^2 = \frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}^{(i)} - \hat{\theta})^2$, (ii) bootstrap confidence interval (Boot), which is defined as $[\hat{\theta} - c^*_{97.5}\hat{\sigma},\ \hat{\theta} - c^*_{2.5}\hat{\sigma}]$ with the α-th percentile of the bootstrap approximation c^*_{α} based on the node resampling network bootstrap by Green and Shalizi (2017) with 999 bootstrap replications, (iii) jackknife empirical likelihood confidence interval (JEL), and (iv) modified jackknife empirical likelihood confidence interval (mJEL).

Tables 1 and 2 gives the empirical coverage rates and average lengths of the confidence intervals above across 1,000 Monte Carlo replications for (i) $(1,2)$-wheels and (ii) 3-cycles, respectively. The nominal rate is 0.95. The main findings from the simulation study are in line with our theoretical results. The Wald and jackknife empirical likelihood confidence intervals tend to over-cover especially when the network is sparse, which verifies our theoretical results. The bootstrap-based intervals are more accurate than the Wald and jackknife empirical likelihood, but still tend to over-cover for sparse network. The modified jackknife empirical likelihood confidence intervals are most robust to the sparsity of 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 modified jackknife empirical likelihood outperforms other methods for all cases.

	Coverage rates				Average interval lengths			
θ_n	Wald	Boot	JEL	mJEL	Wald	Boot	JEL	mJEL
0.5	0.978	0.965	0.983	0.946	0.0134	0.0124	0.0134	0.0113
0.1	0.982	0.960	0.987	0.944	0.0044	0.0037	0.0044	0.0033
0.05	0.983	0.938	0.989	0.947	0.0017	0.0014	0.0018	0.0013

Table 1. Coverage rates and average lengths of 95% confidence intervals for $R = (1, 2)$-wheel with $n = 100$

	Coverage rates			
θ_n	Wald	Boot	JEL	mJEL
0.5	0.959	0.937	0.962	0.944
0.1	0.990	0.972	0.991	0.941
0.05	0.986	0.978	0.995	0.944

Table 2. Coverage rates and average lengths of 95% confidence intervals for $R = 3$-cycle with $n = 200$

We also analyze the power properties of the tests for the null $H_0 : \theta_n = \theta_0$ against the alternative hypotheses $H_1 : \theta_n = \theta_0 + \Delta$ for $\Delta \in (-0.02, -0.01, 0.01, 0.02)$. Table 2 gives the calibrated powers of all the tests across 1,000 Monte Carlo replications, i.e., the rejection frequencies of these tests, where the critical values are given by the Monte Carlo 95th percentiles of these test statistics under H_0. The results suggest that the proposed modified jackknife empirical likelihood test exhibits good calibrated power.

Appendix A. Mathematical Appendix

A.1. **Proof of (5) and (6).** For sets of edges S_1, S_2, define

$$W(S_1, S_2) = \prod_{(i_k,i_l)\in S_1} w(\xi_{i_k}, \xi_{i_l}) \prod_{(i_k,i_l)\in S_2} (1-\rho_n w(\xi_{i_k}, \xi_{i_l})).$$

First, we note that

$$\rho_n^{-|R|}\mathbb{E}[Y_{i_1\dots i_p}|\xi_{i_1}] = \frac{1}{N(R)} \sum_{S\sim R} \mathbb{E}[W(S,\bar{S})|\xi_{i_1}], \tag{13}$$

for all n large enough. Similarly, we have

$$\rho_n^{-|R|}\mathbb{E}[Y_{i_1\dots i_p}|\xi_{i_1},\dots,\xi_{i_s}] = \frac{1}{N(R)} \sum_{S\sim R} \mathbb{E}[W(S,\bar{S})|\xi_{i_1},\dots,\xi_{i_s}], \tag{14}$$

$$\begin{aligned}&\rho_n^{-|R|}\mathbb{E}[Y_{i_1\dots i_p}|\xi_{i_1},\xi_{i_2},\xi_{i_1 i_2}]\\ =\ &\frac{1}{N(R)}\sum_{S\sim R}\{\rho_n^{-1}A_{i_1i_2}\mathbb{E}[W(S\setminus(i_1,i_2),\bar{S})|\xi_{i_1},\xi_{i_2}] + (1-A_{i_1i_2})\mathbb{E}[W(S,\bar{S}\setminus(i_1,i_2))|\xi_{i_1},\xi_{i_2}]\},\end{aligned} \tag{15}$$

and

$$\begin{aligned}&\rho_n^{-|R|}\mathbb{E}[Y_{i_1\dots i_p}|\xi_{i_1},\dots,\xi_{i_s},\xi_{i_1i_2},\dots,\xi_{i_{s-1}i_s}]\\ =\ &\frac{1}{N(R)}\sum_{S\sim R}\Bigg\{\rho_n^{-(s-1)}\prod_{(i_k,i_l)\in S\cap\{(i_1,i_2),\dots,(i_{s-1},i_s)\}} A_{i_ki_l}\prod_{(i_k,i_l)\in\bar{S}\cap\{(i_1,i_2),\dots,(i_{s-1},i_s)\}}(1-A_{i_ki_l})\\ &\times\mathbb{E}[W(S\setminus\{(i_1,i_2),\dots,(i_{s-1},i_s)\},\bar{S}\setminus\{(i_1,i_2),\dots,(i_{s-1},i_s)\})|\xi_{i_1},\dots,\xi_{i_s}]\\ &\cdots+(1-A_{i_1i_2})\cdots(1-A_{i_{s-1}i_s})\times\mathbb{E}[W(S,\bar{S}\setminus\{(i_1,i_2),\dots,(i_{s-1},i_s)\})|\xi_{i_1},\dots,\xi_{i_s}]\Big\},\end{aligned} \tag{16}$$

for $s = 2,\dots,p-1$, and

$$\rho_n^{-|R|}\mathbb{E}[Y_{i_1\dots i_p}|\xi_{i_1},\dots,\xi_{i_s},\xi_{i_1i_2},\dots,\xi_{i_{p-1}i_p}] = \rho_n^{-|R|}\frac{1}{N(R)}\sum_{S\sim R}\prod_{(i_k,i_l)\in S}A_{i_ki_l}\prod_{(i_k,i_l)\in\bar{S}}(1-A_{i_ki_l}). \tag{17}$$

Based on these conditional moments, we now characterize stochastic orders of the right hand side of (4). Note that all β's in (4) have zero mean and no correlation. By (13), we have

$$\mathbb{V}\left(\frac{1}{n}\sum_{i=1}^n\beta_i\right) = \frac{1}{n}\mathbb{V}(\beta_1) = \frac{p^2}{n}\mathbb{V}\left(\rho_n^{-|R|}\mathbb{E}[Y_{1\dots p}|\xi_1]\right) = O\left(\frac{1}{n}\right),$$

which yields the first statements in (5) and (6).

Similarly, by (14) and (15), we have

$$\begin{aligned}\mathbb{V}\left(\frac{1}{n^2}\sum_{i_1<i_2}^n\beta_{i_1i_2}\right) &= \frac{1}{n^4}\binom{n}{2}\mathbb{V}(\beta_{12}) = \frac{1}{n^4}\binom{n}{2}\mathbb{E}[Var(\beta_{12}|\xi_1,\xi_2)] + \frac{1}{n^4}\binom{n}{2}\mathbb{V}(\mathbb{E}[\beta_{12}|\xi_1,\xi_2])\\ &= \frac{1}{n^4}\binom{n}{2}(C_1\rho_n^{-1}+C_2\rho_n) + \frac{1}{n^4}\binom{n}{2}C_3 = O\left(\frac{1}{n^2\rho_n}\right) + O\left(\frac{1}{n^2}\right),\end{aligned}$$

for some positive constants C_1, C_2, and C_3, where the third equality follows from the fact that $\mathbb{E}[\mathbb{V}(A_{12}|\xi_1,\xi_2)] = \mathbb{E}[\rho_n w(\xi_1,\xi_2)(1-\rho_n w(\xi_1,\xi_2))] = O(\rho_n)$.

Generally, we decompose

$$\begin{aligned}
\mathbb{V}\left(\frac{1}{n^s}\sum_{i_1<\cdots<i_s}^{n}\beta_{i_1\ldots i_s}\right) &= \frac{1}{n^{2s}}\binom{n}{s}\mathbb{V}(\beta_{1\ldots s}) \\
&= \frac{1}{n^{2s}}\binom{n}{s}\mathbb{E}[\mathbb{V}(\beta_{1\ldots s}|\xi_1,\ldots,\xi_s)] + \frac{1}{n^{2s}}\binom{n}{s}\mathbb{V}(\mathbb{E}[\beta_{1\ldots s}|\xi_1,\ldots,\xi_s]) \\
&\equiv M_1 + M_2.
\end{aligned}$$

For M_2, by (14), we have

$$M_2 = O\left(\frac{1}{n^s}\right). \tag{18}$$

For M_1, we first consider the case with $s = 2,\ldots,p-1$. By (16), the leading term of $\beta_{i_1\ldots i_s}$ is of the form

$$\begin{aligned}
\tilde{\beta}_{i_1\ldots i_s} &= \frac{1}{N(R)}\sum_{S\sim R}\rho_n^{-(s-1)}\prod_{(i_k,i_l)\in S\cap\{(i_1,i_2),\ldots,(i_{s-1}i_s)\}}A_{i_k i_l}\prod_{(i_k,i_l)\in \bar{S}\cap\{(i_1,i_2),\ldots,(i_{s-1},i_s)\}}(1-A_{i_k i_l}) \\
&\quad\times\mathbb{E}[W(S\setminus\{(i_1,i_2),\ldots,(i_{s-1},i_s)\},\bar{S}\setminus\{(i_1,i_2),\ldots,(i_{s-1},i_s)\})|\xi_{i_1},\ldots,\xi_{i_s}].
\end{aligned}$$

Since $\mathbb{V}\left(\prod_{(i_k,i_l)\in S\cap\{(i_1,i_2),\ldots,(i_{s-1},i_s)\}}A_{i_k i_l}|\xi_{i_1},\ldots,\xi_{i_s}\right) = O_p(\rho_n^{s-1})$, it holds

$$\mathbb{V}(\tilde{\beta}_{1\ldots s}|\xi_1,\ldots,\xi_s) = O_p(\rho_n^{-(s-1)}),$$

which implies

$$M_1 = O\left(\frac{1}{n^s\rho_n^{s-1}}\right), \quad \text{for } s = 2,\ldots,p-1. \tag{19}$$

We next consider the case with $s = p$. Note that $|R| = p-1$ for Case (I) (i.e., R is a wheel), and $|R| = p$ for Case (II) (i.e., R is a cyclic). By the same argument to derive (19) combined with these orders for $|R|$, we obtain for $s = p$,

$$M_1 = \begin{cases} O\left(\frac{1}{n^p\rho_n^{p-1}}\right) & \text{for Case (I)}, \\ O\left(\frac{1}{n^p\rho_n^{p}}\right) & \text{for Case (II)}. \end{cases} \tag{20}$$

Combining (18)-(20), the remaining statements in (5) and (6) follow by Chebyshev's inequality.

A.2. **Proof of Theorem 1.** Since proofs are similar, we only present the proof for Case (I). First, we prove the convergence to χ_1^2 under the assumptions $n\rho_n \to \infty$ and $E[Y_{1,\ldots,p}|\xi_1]$ is random. By the conventional argument (e.g., Owen, 1991), we can prove the asymptotic equivalence

$$\ell(\theta) = \left[\frac{1}{n}\sum_{i=1}^{n}V_i(\theta)^2\right]^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i(\theta)\right)^2 + o_p(1).$$

Thus, it is enough to show that

$$\frac{1}{\sqrt{\omega_n}n}\sum_{i=1}^{n}\rho_n^{-|R|}V_i(\theta) \xrightarrow{d} N(0,1), \tag{21}$$

$$\frac{1}{\omega_n n^2}\sum_{i=1}^{n}\rho_n^{-2|R|}V_i(\theta)^2 \xrightarrow{p} 1, \tag{22}$$

where $\omega_n = Var(\rho_n^{-|R|}\hat{\theta})$. Note that $\frac{1}{n}\sum_{i=1}^{n}V_i(\theta) = \hat{\theta}-\theta$. By (5), we get

$$Var\left(\frac{1}{n}\sum_{i=1}^{n}\rho_n^{-|R|}V_i(\theta)\right) = \omega_n = \frac{\sigma_{1,n}^2}{n}\{1+o(1)\},$$

Thus, (21) follows from the central limit theorem for U-statistics under our assumptions.

For (22), we first note that

$$
\begin{aligned}
\sum_{i=1}^{n} V_i(\theta)^2 &= \sum_{i=1}^{n}\left[\hat\theta-\theta+(n-1)(\hat\theta-\hat\theta^{(i)})\right]^2 = n(\hat\theta-\theta)^2+(n-1)^2\sum_{i=1}^{n}(\hat\theta-\hat\theta^{(i)})^2 \\
&= n(\hat\theta-\theta)^2+(n-1)^2\frac{1}{n}\sum_{i<i'}(\hat\theta^{(i)}-\hat\theta^{(i')})^2, \qquad (23)
\end{aligned}
$$

where the second equality follows from $\sum_{i=1}^{n}(\hat\theta-\hat\theta^{(i)})=0$, and the third equality follows from a direct calculation. Thus, we have

$$
\begin{aligned}
\frac{1}{\omega_n n^2}\sum_{i=1}^{n}\rho_n^{-2|R|}V_i(\theta)^2 &= \frac{1}{\omega_n\rho_n^{2|R|}}\left[\frac{1}{n}(\hat\theta-\theta)^2+\frac{(n-1)^2}{n^3}\sum_{i<i'}(\hat\theta^{(i)}-\hat\theta^{(i')})^2\right] \\
&= \frac{(n-1)^2}{\omega_n n^3}\rho_n^{-2|R|}\sum_{i<i'}(\hat\theta^{(i)}-\hat\theta^{(i')})^2+o_p(1) \\
&= \frac{(n-1)^2}{\omega_n n^3}\sum_{i<i'}\left\{\frac{1}{n-1}(\beta_{i'}-\beta_i)\right\}^2+o_p(1) = \frac{n-1}{\omega_n n^2}Var(\beta_1)+o_p(1) \\
&\xrightarrow{p} 1,
\end{aligned}
$$

where the second equality follows from $\frac{1}{\omega_n n}\rho_n^{-2|R|}(\hat\theta-\theta)^2\xrightarrow{p}0$ (by the consistency $\rho_n^{-|R|}(\hat\theta-\theta)\xrightarrow{p}0$), and the fourth equality follows from the law of large numbers.

Second, we consider the case where $n\rho_n=O(1)$ or $E[Y_{1\ldots p}|\xi_1]$ degenerates to a constant. For this case, it is enough to show (21) and

$$
\frac{1}{\omega_n n^2}\sum_{i=1}^{n}\rho_n^{-2|R|}V_i(\theta)^2\xrightarrow{p}\sigma_*^2. \qquad (24)
$$

Using the fact that the terms in (4) are uncorrelated, we get

$$
Var\left(\frac{1}{n}\sum_{i=1}^{n}\rho_n^{-|R|}V_i(\theta)\right)=\omega_n=\left(\frac{\sigma_{1,n}^2}{n}+\frac{\sigma_{2,n}^2}{2n^2}+\frac{\sigma_{3,n}^2}{6n^3}+\cdots+\frac{\sigma_{p,n}^2}{p!n^p}\right)\{1+o(1)\}.
$$

Thus, (21) follows from the central limit theorem for U-statistics under our assumptions.

For (24), we have

$$
\begin{aligned}
&\frac{1}{\omega_n n^2}\sum_{i=1}^{n}\rho_n^{-2|R|}V_i(\theta_n)^2 \\
=\ & \frac{1}{\omega_n\rho_n^{2|R|}}\left[\frac{1}{n}(\hat\theta-\theta_n)^2+\frac{(n-1)^2}{n^3}\sum_{i<i'}(\hat\theta^{(i)}-\hat\theta^{(i')})^2\right]=\frac{(n-1)^2}{\omega_n n^3}\rho_n^{-2|R|}\sum_{i<i'}(\hat\theta^{(i)}-\hat\theta^{(i')})^2+o_p(1) \\
=\ & \frac{(n-1)^2}{\omega_n n^3}\sum_{i<i'}\left\{\frac{1}{n-1}(\beta_{i'}-\beta_i)+\frac{1}{(n-1)^2}\sum_{l_1=1}^{(i,i')}(\beta_{i'l_1}-\beta_{il_1})+\cdots+\frac{1}{(n-1)^p}\sum_{l_1<\cdots<l_{p-1}}^{(i,i')}(\beta_{i'l_1\ldots l_{p-1}}-\beta_{il_1\ldots l_{p-1}})\right\}^2 \\
&+o_p(1) \\
=\ & \frac{(n-1)^2}{\omega_n n^2}\left[\frac{\sigma_{1,n}^2}{n-1}+\binom{n-2}{1}\frac{\sigma_{2,n}^2}{(n-1)^3}+\binom{n-2}{2}\frac{\sigma_{3,n}^2}{(n-1)^5}+\cdots+\binom{n-2}{p-1}\frac{\sigma_{p,n}^2}{(n-1)^{2p-1}}\right]+o_p(1) \\
=\ & \frac{(n-1)^2}{\omega_n n^2}\left[\frac{\sigma_{1,n}^2}{n}+\frac{\sigma_{2,n}^2}{n^2}+\frac{\sigma_{3,n}^2}{2n^3}+\cdots+\frac{\sigma_{p,n}^2}{(p-1)!n^p}\right]+o_p(1) \\
\xrightarrow{p}\ & \sigma_*^2,
\end{aligned}
$$

where the notation $\sum^{(i,i')}$ indicates summations avoiding the values i and i'. The first equality follows from (23), the second equality follows from $\frac{1}{\omega_n n}\rho_n^{-2|R|}(\hat{\theta}-\theta)^2 \xrightarrow{p} 0$ (by the consistency $\rho_n^{-|R|}(\hat{\theta}-\theta) \xrightarrow{p} 0$), and the fourth equality follows from the law of large numbers.

A.3. **Proof of Theorem 2.** Here we present the proof for the case of $p = 3$. As in the proof of Theorem 1, we can prove the asymptotic equivalence

$$\ell^m(\theta) = \left[\frac{1}{\omega_n n^2}\sum_{i=1}^n V_i^m(\theta)^2\right]^{-1}\left(\frac{1}{\sqrt{\omega_n} n}\sum_{l=1}^n V_i^m(\theta)\right)^2 + o_p(1).$$

Thus, it is enough to show

$$\frac{1}{\sqrt{\omega_n} n}\sum_{l=1}^n \rho_n^{-|R|}V_i^m(\theta) \quad \xrightarrow{d} \quad N(0,\sigma_*^2), \tag{25}$$

$$\frac{1}{\omega_n n^2}\sum_{l=1}^n \rho_n^{-2|R|}V_i^m(\theta)^2 \quad \xrightarrow{p} \quad \sigma_*^2. \tag{26}$$

A similar argument to (24) yields

$$\frac{1}{\omega_n n^2}\sum_{l=1}^n \rho_n^{-2|R|}V_i(\hat{\theta})^2 \xrightarrow{p} \sigma_*^2. \tag{27}$$

Thus, the consistency $\rho_n^{-|R|}(\hat{\theta}-\theta) \xrightarrow{p} 0$ implies (26).

It remains to show (25). Since $\sum_{i=1}^n \rho_n^{-|R|}V_i(\hat{\theta}) = 0$, we have

$$\frac{1}{\sqrt{\omega_n} n}\sum_{i=1}^n \rho_n^{-|R|}V_i^m(\theta) = \hat{\Gamma}\tilde{\Gamma}^{-1}\frac{1}{\sqrt{\omega_n} n}\sum_{i=1}^n \rho_n^{-|R|}V_i(\theta)$$

$$= \sqrt{\frac{\frac{1}{\omega_n n^2}\sum_{i=1}^n V_i(\hat{\theta})^2}{\frac{1}{\omega_n n^2}\sum_{i=1}^n V_i(\hat{\theta})^2 - \frac{1}{\omega_n n^2}\sum_{i_1<i_2}^n M_{i_1 i_2}^2 + \frac{1}{\omega_n n^2}\sum_{i_1<i_2<i_3}^n M_{i_1 i_2 i_3}^2}}\frac{1}{\sqrt{\omega_n} n}\sum_{i=1}^n \rho_n^{-|R|}V_i(\theta).$$

By (21), it holds $\frac{1}{\sqrt{\omega_n} n}\sum_{i=1}^n \rho_n^{-|R|}V_i(\theta) \xrightarrow{d} N(0,1)$. Also a similar argument to (24) yields $\frac{1}{\omega_n n^2}\sum_{i=1}^n \rho_n^{-2|R|}V_i(\hat{\theta})^2 \xrightarrow{p} \sigma_*^2$. Thus, for (25), it remains to show that

$$\frac{1}{\omega_n n^2}\sum_{i=1}^n \rho_n^{-2|R|}V_i(\hat{\theta})^2 - \frac{1}{\omega_n n^2}\sum_{i_1<i_2}^n \rho_n^{-2|R|}M_{i_1 i_2}^2 + \frac{1}{\omega_n n^2}\sum_{i_1<i_2<i_3}^n \rho_n^{-2|R|}M_{i_1 i_2 i_3}^2 \xrightarrow{p} 1, \tag{28}$$

which follows from Lemma 1.

Lemma 1. *When $p = 3$, we have*

$$\frac{1}{\omega_n n^2}\sum_{i<j}\rho_n^{-2|R|}M_{ij}^2 \quad = \quad \frac{1}{\omega_n}\left[\frac{\sigma_{2,n}^2}{2n^2} + \frac{\sigma_{3,n}^2}{2n^3}\right] + o_p(1),$$

$$\frac{1}{\omega_n n^2}\sum_{i<j<k}\rho_n^{-2|R|}M_{ijk}^2 \quad = \quad \frac{1}{\omega_n}\left[\frac{\sigma_{3,n}^2}{6n^3}\right] + o_p(1).$$

Proof: By the expansion (4), we have

$$\begin{aligned}
n\rho_n^{-|R|}(\hat{P}(R)-P(R)) &= \sum_{i=1}^{n}\beta_i + \frac{1}{n}\sum_{i_1<i_2}\beta_{i_1i_2} + \frac{1}{n^2}\sum_{i_1<i_2<i_3}\beta_{i_1i_2i_3},\\
(n-1)\rho_n^{-|R|}(\hat{P}_{-l}(R)-P(R)) &= \sum_{i}^{(l)}\beta_i + \frac{1}{n-1}\sum_{i_1<i_2}^{(l)}\beta_{i_1i_2} + \frac{1}{(n-1)^2}\sum_{i_1<i_2<i_3}^{(l)}\beta_{i_1i_2i_3},\\
(n-2)\rho_n^{-|R|}(\hat{P}_{-l,-l'}(R)-P(R)) &= \sum_{i}^{(l,l')}\beta_i + \frac{1}{n-2}\sum_{i_1<i_2}^{(l,l')}\beta_{i_1i_2} + \frac{1}{(n-2)^2}\sum_{i_1<i_2<i_3}^{(l,l')}\beta_{i_1i_2i_3},\\
(n-3)\rho_n^{-|R|}(\hat{P}_{-l,-l',-l''}(R)-P(R)) &= \sum_{i}^{(l,l',l'')}\beta_i + \frac{1}{n-3}\sum_{i_1<i_2}^{(l,l',l'')}\beta_{i_1i_2} + \frac{1}{(n-3)^2}\sum_{i_1<i_2<i_3}^{(l,l',l'')}\beta_{i_1i_2i_3}.
\end{aligned}$$

where the notations $\sum^{(l)}$, $\sum^{(l,l')}$, and $\sum^{(l,l',l'')}$ indicate summations avoiding the value l, the values l and l' and the values l, l', and l'', respectively. Then we have

$$\begin{aligned}
\rho_n^{-|R|}(V_l-\theta) &= \rho_n^{-|R|}\left\{n(\hat{P}(R)-P(R))-(n-1)(\hat{P}_{-l}(R)-P(R))\right\} = \beta_l + (\beta_{l\cdot}-\beta_{\cdot\cdot}) + (\beta_{l\cdot\cdot}-\beta_{\cdots}^{(1)}),\\
\rho_n^{-|R|}M_{ll'} &= \rho_n^{-|R|}\left\{n\hat{P}(R)-(n-1)(\hat{P}_{-l}(R)+\hat{P}_{-l'}(R))+(n-2)\hat{P}_{-l,-l'}(R)\right\}\\
&= \frac{1}{n-2}(\beta_{ll'}-\beta_{l\cdot}-\beta_{l'\cdot}+\beta_{\cdot\cdot}) + (\beta_{ll'\cdot}-\beta_{\cdots}^{(2)}),\\
\rho_n^{-|R|}M_{ll'l''} &= \rho_n^{-|R|}\left\{n\hat{P}(R)-(n-1)\left(\hat{P}_{-l}(R)+\hat{P}_{-l'}(R)+\hat{P}_{-l''}(R)\right)\right.\\
&\quad \left. +(n-2)\left(\hat{P}_{-l,-l'}(R)+\hat{P}_{-l',-l''}(R)+\hat{P}_{-l,-l''}(R)\right)-(n-3)\hat{P}_{-l,-l',-l''}(R)\right\}\\
&= \frac{1}{(n-3)^2}(\beta_{ll'l''}-\beta_{l\cdot\cdot}-\beta_{l'\cdot\cdot}-\beta_{l''\cdot\cdot}+\beta_{ll'\cdot}+\beta_{l'l''\cdot}+\beta_{ll''\cdot}+\beta_{\cdots}^{(3)}),
\end{aligned}$$

where $\beta_{l\cdot} = \frac{1}{n-1}\sum_i^{(l)}\beta_{li}$, $\beta_{\cdot\cdot} = \frac{1}{n(n-1)}\sum_{l<l'}\beta_{ll'}$, $\beta_{l\cdot\cdot} = \frac{1}{(n-1)^2}\sum_{i<i'}^{(l)}\beta_{lii'}$, $\beta_{ll'\cdot} = \frac{1}{(n-2)^2}\sum_i^{(l,l')}\beta_{ll'i}$, $\beta_{\cdots}^{(1)} = \left(\frac{1}{(n-1)^2}-\frac{1}{n^2}\right)\sum_{l<l'<l''}\beta_{ll'l''}$, $\beta_{\cdots}^{(2)} = \left(\left(\frac{1}{(n-2)^2}-\frac{1}{n^2}\right)-2\left(\frac{1}{(n-2)^2}-\frac{1}{(n-1)^2}\right)\right)\sum_{l<l'<l''}\beta_{ll'l''}$, and $\beta_{\cdots}^{(3)} = \left(\left(\frac{1}{(n-3)^2}-\frac{1}{n^2}\right)-3\left(\frac{1}{(n-3)^2}-\frac{1}{(n-1)^2}\right)+3\left(\frac{1}{(n-3)^2}-\frac{1}{(n-2)^2}\right)\right)\sum_{l<l'<l''}\beta_{ll'l''}$.

Thus, the conclusion follows by the law of large numbers using the following moments

$$\begin{aligned}
E\left[\sum_{l=1}^{n}\beta_{l\cdot}^2\right] &= E\left[\frac{1}{(n-1)^2}\sum_{l=1}^{n}\sum_{i}^{(l)}\beta_{li}^2\right] = \frac{n}{n-1}\sigma_{2,n}^2,\\
E\left[\sum_{l=1}^{n}\beta_{l\cdot\cdot}^2\right] &= E\left[\frac{1}{n(n-1)^4}\sum_{l=1}^{n}\sum_{i<i'}^{(l)}\beta_{lii'}^2\right] = \frac{n^2}{2(n-1)^3}\sigma_{3,n}^2,\\
E\left[\sum_{l<l'}\left(\frac{1}{n-2}\beta_{ll'}\right)^2\right] &= E\left[\frac{1}{(n-2)^2}\sum_{l<l'}\beta_{ll'}^2\right] = \frac{n(n-1)}{2(n-2)^2}\sigma_{2,n}^2,\\
E\left[\sum_{l<l'}\beta_{ll'\cdot}^2\right] &= E\left[\frac{1}{(n-2)^4}\sum_{l<l'}\sum_{i}^{(l,l')}\beta_{ll'i}^2\right] = \frac{n(n-1)}{2(n-2)^3}\sigma_{3,n}^2,\\
E\left[\sum_{l<l'<l''}\left(\frac{1}{(n-3)^2}\beta_{ll'l''}\right)^2\right] &= E\left[\frac{1}{(n-3)^4}\sum_{l<l'<l''}\beta_{ll'l''}^2\right] = \frac{n(n-1)(n-2)}{6(n-3)^4}\sigma_{3,n}^2,
\end{aligned}$$

References

[1] Bhattacharyya, S. and P. J. Bickel (2015) Subsampling bootstrap of count features of networks, *Annals of Statistics*, 43, 2384-2411.

[2] Bickel, P. J. and A. Chen (2009) A nonparametric view of network models and Newman-Girvan and other modularities, *Proceedings of the National Academy of Sciences*, 106, 21068-21073.

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

[4] Bickel, P., Choi, D., Chang, X., and H. Zhang (2013) Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels, *Annals of Statistics*, 41, 1922-1943.

[5] Chatterjee, S., Diaconis, P. and A. Sly (2011) Random graphs with a given degree sequence, *Annals of Applied Probability*, 21, 1400-1435.

[6] Crane, H. (2018) *Probabilistic Foundations of Statistical Network Analysis*, CRC Press.

[7] Diaconis, P., Holmes, S. and S. Janson (2008) Threshold graph limits and random threshold graphs, *Internet Mathematics*, 5, 267-320.

[8] Diaconis, P. and S. Janson (2008) Graph limits and exchangeable random graphs, *Rendiconti di Matematica*, 28, 33-61.

[9] Efron, B. and C. Stein (1981) The jackknife estimate of variance, *Annals of Statistics*, 9, 586-596.

[10] Green, A. and C. R. Shalizi (2017) Bootstrapping exchangeable random graphs, Working paper.

[11] Hall, P. and B. La Scala (1990) Methodology and algorithms of empirical likelihood, *International Statistical Review*, 58, 109-127.

[12] Hinkley, D. V. (1978) Improving the jackknife with special reference to correlation estimation, *Biometrika*, 65, 13-21.

[13] Hoff, P., Raftery, A. and M. Handcock (2002) Latent space approaches to social network analysis, *Journal of the American Statistical Association*, 97, 1090–1098.

[14] Jing, B. Y., Yuan, J. and W. Zhou (2009) Jackknife empirical likelihood, *Journal of the American Statistical Association*, 104, 1224-1232.

[15] Kallenberg, O. (2005) *Probabilistic Symmetries and Invariance Principles*, Springer.

[16] Kolaczyk, E. D. (2009) *Statistical Analysis of Network Data*, Springer.

[17] Levin, K. and E. Levina (2019) Bootstrapping networks with latent space structure, Working paper.

[18] Lin, Q., Lunde, R. and P. Sarkar (2020a) On the theoretical properties of the network jackknife, Working paper.

[19] Lin, Q., Lunde, R. and P. Sarkar (2020b) Higher-order correct multiplier bootstraps for count functionals of networks, Working paper.

[20] Owen, A. B. (1988) Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 75, 237-249.

Graduate School of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan.

Email address: `matsushita.y@r.hit-u.ac.jp`

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

Email address: `t.otsu@lse.ac.uk`