Statistica Sinica **26** (2016), 273-293
doi:http://dx.doi.org/10.5705/ss.2014.180

ASYMPTOTICS IN UNDIRECTED RANDOM GRAPH MODELS PARAMETERIZED BY THE STRENGTHS OF VERTICES

Ting Yan[1], Hong Qin[1] and Hansheng Wang[2]

[1]*Central China Normal University and* [2]*Peking University*

Abstract: To capture the heterozygosity of vertex degrees of networks and understand their distributions, a class of random graph models parameterized by the strengths of vertices is proposed. These models have a framework of mutually independent edges, where the number of parameters matches the size of the network. The asymptotic properties of the maximum likelihood estimator have been derived in such models as the β-model, but general results are lacking. In these models, the likelihood equations are identical to the moment equations. Here, we establish a unified asymptotic result that includes the consistency and asymptotic normality of the moment estimator instead of the maximum likelihood estimator, when the number of parameters goes to infinity. We apply it to the generalized β-model, maximum entropy models, and Poisson models.

Key words and phrases: Asymptotical normality, consistency, increasing number of parameters, moment estimators, undirected network models.

1. Introduction

Exploring the generated mechanisms of networks is an important topic in network analysis. The degrees of vertices have received wide attention recently (e.g., Britton, Deijfen, and Martin-Löf (2006); Blitzstein and Diaconis (2010); Bickel, Chen, and Levina (2012); Zhao, Levina, and Zhu (2012); Rinaldo, Petrović, and Fienberg (2013); Hillar and Wibisono (2013)). A larger degree indicates more ability of that vertex to participate in network connections. The Erdös-Rényi model (Erdös and Rényi (1959)) is generally acknowledged as one of the earliest random graph models, in which each edge occurs with the same probability independent of any other edge. Here one has homogenous random graphs in which all vertex degrees have the same binomial distribution. For large number of vertices, they are approximated by the Poisson distributions.

Consider an undirected random graphs with n vertices and binary edges. To capture the heterozygosity of vertex degrees, a class of n-parameter models is proposed, where the parameter of a vertex reflects the propensity of this vertex to form network connections. The β-model (a name coined by Chatterjee, Diaconis, and Sly (2011)), an undirected version of the p_1 directed exponential random model, originally proposed by Holland and Leinhardt (1981), suggests that the probability of an edge only depends on the sum of the strength parameters of the corresponding two vertices and has a logistic representation. In this model, the degree sequence is the exclusively natural sufficient statistic. Lauritzen (2003; 2008) addressed the β-model as the natural model for representing exchangeable binary arrays whose distribution only depends on the row and column totals. When the number of parameters (or vertices) goes to infinity, Chatterjee, Diaconis, and Sly (2011) proved the uniform consistency of the maximum likelihood estimator (MLE); Yan and Xu (2013) further derived its asymptotic normality. Rinaldo, Petrović, and Fienberg (2013) established the necessary and sufficient condition for the existence and uniqueness of the MLE when a sample is given. A similar model is Chung and Lu's (2002) log-linear model, where the edge probability p_{ij} between vertices i and j is $w_i w_j/(\sum_{k=1}^n w_k)$ under the normalization constraint $w_i^2 \leq \sum_{k=1}^n w_k$, with w_i referred to as the weight of vertex i. Janson (2010) obtained conditions under which these two models are asymptotically equivalent as $n \to \infty$, and Perry and Wolfe (2012) demonstrated that they give rise to essentially the same likelihood-based estimates of link probabilities in the sparse finite-sample regimes. Moreover, Olhede and Wolfe (2012) derive the sampling properties of undirected networks parameterized by independent Bernoulli trials.

An edge in undirected networks takes not only binary ("present" or "absent") values but also weighted edges in many scenarios. Examples include collaboration networks, where an edge represents the number of papers coauthored by two scholars (e.g., Newman (2001)) or the number of bills cosponsored by legislators (Fowler (2006)); social categorical recorded networks, where an edge denotes the intensity of friendship such as "very good", "good", "general", etc, or the similarity (e.g., Bernard, Killworth, and Sailer (1980; 1982)); and neural networks, where an edge represents discrete or continuous electrical pulses. In the case of weighted edges, Hillar and Wibisono (2013) have proposed maximum entropy models, in which the distribution of the edge (i, j) only depends on the sum of strength parameters of vertices i and j independent of any other edges, and proved that the MLE is uniformly consistent; Yan, Zhao, and Qin (2015) derived the corresponding asymptotical normality. Ranola et al. (2010) used the Poisson distribution to model the multiple edges of biological networks. However, asymptotic properties of the MLE in this model are still unknown.

At present, all asymptotic results of the MLE in these models are ad-hoc or there are none. The lack of a unified theoretical framework is something we wish to resolve. Two notable characterizations appeared in the above models.

First, studying the distributions of vertex degrees is in the framework of mutually independent edges. Second, the probability-mass or density function of the edge (i,j) only depends on the sum of $\alpha_i+\alpha_j$, where α_i denotes the strength parameter of vertex i. Under these two assumptions, we study asymptotic properties of the moment estimator instead of the MLE. The reason we use the method of moments is that the degrees of vertices are explicitly represented in the moment equations; they are not available for the maximum likelihood equations since we consider a general probability mass function or density function for the edges. Using the moment equations, it is convenient to establish the relationship between the degrees and the parameters that are important for deriving the asymptotic properties of the moment estimator. However, when the degrees of vertices are the sufficient statistics in the exponential family distribution for graphs, the moment equations are identical to the maximum likelihood equations (e.g., the β-model, maximum entropy models, Poisson model).

For the remainder of this paper, we proceed as follows. In Section 2, we establish a unified asymptotic result including the consistency and asymptotic normality for the moment estimator of the parameter vector as the number of parameters goes to infinity. In Section 3, we illustrate several applications on our main results. In the generalized β-model, we relax the assumption of Hillar and Wibisono (2013) that all the parameters are bounded by a constant to guarantee the consistency of the MLE. Further discussion is in Section 4. Proofs are in the Appendix.

2. Main Results

We derive asymptotic results for the moment estimator. As Yan and Xu (2013) were only concerned with the asymptotic normality of the MLE in a special β-model, the new challenge is to derive the consistency and asymptotic normality simultaneously for the moment estimator in general network models based on the degrees of vertices. We study the conditions guaranteeing such results.

2.1. Notation and preliminaries

Let $\mathbb{R}_+ = (0,\infty)$, $\mathbb{R}_0 = [0,\infty)$, $\mathbb{N} = \{1,2,\ldots\}$, $\mathbb{N}_0 = \{0,1,2,\ldots\}$. For a vector $\mathbf{x} = (x_1,\ldots,x_n)^\top \in R^n$, denote by $\|\mathbf{x}\|_\infty = \max_{1\le i\le n}|x_i|$ the ℓ_∞ norm of $\mathbf{x}$. For an $n\times n$ matrix $J=(J_{ij})$, $\|J\|_\infty$ denotes the matrix norm induced by the $\|\cdot\|_\infty$-norm on vectors in $\mathbb{R}^n$:

$$\|J\|_\infty = \max_{\mathbf{x}\neq 0}\frac{\|J\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \max_{1\le i\le n}\sum_{j=1}^n |J_{ij}|.$$

Let D be an open convex subset of $\mathbb{R}^n$. We say an $n \times n$ function matrix $G(\mathbf{x})$ whose elements $G_{ij}(\mathbf{x})$ are functions on vectors $\mathbf{x}$, is Lipschitz continuous on D if there exists a real number λ such that for any $\mathbf{v} \in R^n$ and any $\mathbf{x}, \mathbf{y} \in D$,

$$\|G(\mathbf{x})(\mathbf{v}) - G(\mathbf{y})(\mathbf{v})\|_\infty \le \lambda \|\mathbf{x} - \mathbf{y}\|_\infty \|\mathbf{v}\|_\infty, \tag{2.1}$$

where λ may depend on n but is independent of $\mathbf{x}$ and $\mathbf{y}$. For every fixed n, λ is a constant.

Given $m, M > 0$, we say an $n \times n$ matrix $V = (v_{ij})$ belongs to the matrix class $\mathcal{L}_n(m, M)$ if V is a diagonally balanced matrix with positive elements bounded by m and M,

$$\begin{array}{l} v_{ii} = \sum_{j=1, j\neq i}^n v_{ij}, \quad i = 1, \ldots, n, \\ m \le v_{ij} \le M, \quad i, j = 1, \ldots, n; i \ne j. \end{array} \tag{2.2}$$

We use V to denote the Jacobian matrix induced by the moment equations and show that it belongs to the matrix class $\mathcal{L}_n(m, M)$. We require the inverse of V, which doesn't have a closed form. Yan and Xu (2013) proposed approximating the inverse V^{-1} of V by a matrix $S = (s_{ij})$, where

$$s_{ij} = \frac{\delta_{ij}}{v_{ii}} - \frac{1}{v_{..}}. \tag{2.3}$$

Here, $v_{..} = \sum_{i,j=1}^n (1 - \delta_{ij}) v_{ij}$ with δ_{ij} the Kronecker delta function. We also use S to approximate V^{-1}, whose approximate errors are given in Proposition A.2 of the Appendix.

2.2. Asymptotic results

Consider a probability distribution $\mathbb{P}$ on the adjacency matrix $A = (a_{ij})_{n\times n}$ of an undirected random graph $\mathcal{G}$, with each edge a_{ij} $(i \ne j)$ having the form of the discrete probability distribution

$$P(a_{ij} = a) = f((\alpha_i + \alpha_j)a),$$

or the density function $f((\alpha_i + \alpha_j)a)$, where $f(\cdot)$ is a probability mass or density function and α_i denotes the strength parameter of vertex i. If f is not well behaved, there may exist multiple solutions for the moment estimate. For example, if f takes the continuous distribution with the density $3\alpha x^2 e^{-\alpha x^3}$ (a special case of the Weibull distribution), then its expectation is $\Gamma(4/3)\alpha^{-1/3}$ and the corresponding moment equations may have multiple solutions since this system of equations is involved with the polynomial formula, where $\Gamma(\cdot)$ is the Gamma function. If f behaves well, like the logistic function, then the solution is unique if it exists. If edges only take two states "present" or "absent", then a is the dichotomous value "1" or "0". In communication networks, if edges denote the number of emails between two persons, then a takes values from the set $\mathbb{N}_0$; and if edges denote the calling time, then a takes nonnegative continuous real values. For convenience, let $a_{ii} = 0$, $i = 1, \ldots, n$, so there are no self-loops. We assume that $a_{ij}, 1 \leq i < j \leq n$ are mutually independent.

Here $\mathbb{E}(a_{ij})$ only depends on the sum $\alpha_i + \alpha_j$. Let $\mu(\alpha_i + \alpha_j) = \mathbb{E}(a_{ij})$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$. Let $d_i = \sum_j a_{ij}$ be the degree of vertex i and $\mathbf{d} = (d_1, \ldots, d_n)$ be the degree sequence of graph $\mathcal{G}$. Define a system of functions:

$$F_i(\boldsymbol{\alpha}) = d_i - \mathbb{E}(d_i) = d_i - \sum_{j=1; j \neq i}^{n} \mu(\alpha_i + \alpha_j), \quad i = 1, \ldots, n,$$
$$F(\boldsymbol{\alpha}) = (F_1(\boldsymbol{\alpha}), \ldots, F_n(\boldsymbol{\alpha}))^\top.$$

The solution to $F(\boldsymbol{\alpha}) = 0$ is the moment estimator of $\boldsymbol{\alpha}$ induced by the moment equation $\mathbf{d} = \mathbb{E}(\mathbf{d})$. Henceforth, we use $\widehat{\boldsymbol{\alpha}}$ to denote the moment estimator of $\boldsymbol{\alpha}$ satisfying $F(\widehat{\boldsymbol{\alpha}}) = 0$. Let $F'(\boldsymbol{\alpha})$ denote the Jacobin matrix of $F(\boldsymbol{\alpha})$ on $\boldsymbol{\alpha}$.

A first a proposition has a proof in Appendix A.

Proposition 1. *Assume that*

(C1) $V := Var(\mathbf{d}) \in \mathcal{L}_n(m, M)$;

(C2) $(d_i - E(d_i))/v_{ii}^{1/2}$, $i = 1, \ldots, n$ *and* $\sum_i (d_i - E(d_i))/(2v_{\cdot\cdot})^{1/2}$ *are asymptotically standard normal as* $n \to \infty$.

If $M/m^2 = o(n)$, *then for any fixed* k, *the first* k *elements of* $S(\mathbf{d} - \mathbb{E}(\mathbf{d}))$ *are asymptotically normal distribution with mean zero and covariance matrix given by the upper* $k \times k$ *submatrix of the diagonal matrix* $B = \mathrm{diag}(1/v_{11}, \ldots, 1/v_{nn})$, *where* S *is the approximate inverse of* V *defined at* (2.3).

Remark 1. Since d_i and $(1/2)\sum_i d_i$ are the respective sums of n and $n(n-1)/2$ independent random variables, condition C2 can be easily verified by Lyapunov's (Billingsley (1995), p.362), Lindeberg's (1922), or Loève's (1977, p.289) Central Limit Theorem.

Given $\boldsymbol{\alpha}$ with $q_n \leq \alpha_i + \alpha_j \leq Q_n$, assume $A \sim \mathbb{P}_{\boldsymbol{\alpha}}$. We make the following assumptions.

(C3) $F'(\boldsymbol{\alpha}) \in \mathcal{L}_n(m, M)$ or $-F'(\boldsymbol{\alpha}) \in \mathcal{L}_n(m, M)$, where $m = m(q_n, Q_n)$ and $M = M(q_n, Q_n)$.

(C4) $F'(\boldsymbol{\alpha})$ is Lipschitz continuous with $\lambda = (n-1)\phi_1$, where $\phi_i := \phi_i(q_n, Q_n)$, $i = 1, 2, 3$.

(C5) With probability approaching one,

$$\max_{i=1,\ldots,n} |d_i - \mathbb{E}(d_i)| \leq \phi_2 \sqrt{(n-1)\log(n-1)}. \tag{2.4}$$

(C6) $|\mu''(\hat{\theta}_{ij})| = O_p(\phi_3)$, where $\hat{\theta}_{ij} = t(\alpha_i + \alpha_j) + (1-t)(\hat{\alpha}_i + \hat{\alpha}_j)$, $0 \le t \le 1$.

Theorem 1.

(1) *Assume that* (C3)$-$(C5) *hold and*

$$\frac{M^2\phi_2}{m^3}\sqrt{\frac{\log(n-1)}{(n-1)}} = o(1), \tag{2.5}$$

$$\frac{M^4\phi_1\phi_2}{m^5}\sqrt{\frac{\log(n-1)}{(n-1)}} = o(1), \tag{2.6}$$

then as n goes to infinity, with probability approaching one, the moment estimator $\widehat{\boldsymbol{\alpha}}$ exists and satisfies

$$\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_\infty = O_p(\frac{M^2\phi_2}{m^3}\sqrt{\frac{\log n}{n}}) = o_p(1). \tag{2.7}$$

(2) *If inequality* (2.7) *and conditions* (C1), (C2), *and* (C6), *and if*

$$\frac{M^6\phi_2\phi_3\log n}{m^9 n^{1/2}} = o(1), \tag{2.8}$$

and $M^2/m^3 = o(n)$, then for any fixed $k \ge 1$, as $n \to \infty$, the vector consisting of the first k elements of $(B^{-1})^{1/2}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ is asymptotically standard multivariate normal, where $(B^{-1})^{1/2} = \mathrm{diag}(v_{11}^{1/2}, \ldots, v_{nn}^{1/2})$.

The proof of the theorem is in Appendix B. We use the Newton-Kantovorich Theorem, Proposition A.1 in the Appendix, to prove the consistency of the moment estimator by constructing the Newton's iterative sequence. This technical step is different from Chatterjee, Diaconis, and Sly (2011) and yields a simple proof. It requires that the Jacobin matrix $F'(\boldsymbol{\alpha})$ is Lipschitz continuous and restricts the increasing rate of the Lipschitz λ on the dimension n of $\boldsymbol{\alpha}$; C4 has the Lipschitz λ not faster than n. Generally, it is of the magnitude of $n-1$, since $F_i(\boldsymbol{\alpha})$ is the sum of $n-1$ items. Condition C5 guarantees that the ℓ_∞-norm of $F(\boldsymbol{\alpha}) = \mathbf{d} - E(\mathbf{d})$ is bounded in the magnitude of $(n\log n)^{1/2}$, with probability approaching one. If the random variables are sub-exponential, then this condition can be verified by the concentration inequality (Vershynin (2012)). Condition C6 requires that the second derivative of $\mu(\boldsymbol{\alpha})$ exit and be bounded by a function of q_n and Q_n with probability approaching one. If $\widehat{\boldsymbol{\alpha}}$ is a consistent estimator of $\boldsymbol{\alpha}$, then the upper bound of the second derivation of $\mu(\hat{\theta}_{ij})$ is mainly determined by q_n and Q_n since $q_n \le \alpha_i + \alpha_j \le Q_n$ for all $i \ne j$. If ϕ_1, ϕ_2, and ϕ_3 are at least in the magnitude of $O(1)$, then (2.6) implies (2.5) and (2.8) implies $M^2/m^3 = o(n)$.

3. Applications

3.1. Generalized beta model

The beta model has been studied by many authors (e.g., Jackson (2008), Blitzstein and Diaconis (2010), Chatterjee, Diaconis, and Sly (2011)). Hillar and Wibisono (2013) have given its discrete version and proved the consistency of the MLE under the assumption that all the parameters are bounded by a constant, similar to Chatterjee, Diaconis, and Sly (2011). We consider here that the range of parameters varies as n increases. Assume that a_{ij}, $1 \le i \ne j \le n$, take values from the set $\Omega = \{0, 1, \ldots, r-1\}$ with r a fixed constant, and are distributed independently with

$$P(a_{ij} = a) = \frac{e^{a(\alpha_i+\alpha_j)}}{\sum_{k=0}^{r-1} e^{k(\alpha_i+\alpha_j)}}.$$

The moment equations are

$$d_i = \sum_{j=1;j\ne i}^{n} \sum_{a=0}^{r-1} \frac{a e^{a(\widehat{\alpha}_i+\widehat{\alpha}_j)}}{\sum_{k=0}^{r-1} e^{k(\widehat{\alpha}_i+\widehat{\alpha}_j)}}, \quad i = 1, \ldots, n, \tag{3.1}$$

and identical to the maximum likelihood equations. Here, we consider the symmetric parameter space

$$D = \{\boldsymbol{\alpha} \in R^n : -Q_n \le \alpha_i + \alpha_j \le Q_n \ \text{ for } \ 1 \le i < j \le n\}.$$

The Jacobian matrix $F'(\boldsymbol{\alpha})$ of $F(\boldsymbol{\alpha})$ is, for $i, j = 1, \ldots, n$,

$$\frac{\partial F_i}{\partial \alpha_i} = \sum_{j=1;j\ne i}^{n} \frac{\sum_{0\le k<l\le r-1}(k-l)^2 e^{(k+l)(\alpha_i+\alpha_j)}}{(\sum_{a=0}^{r-1} e^{a(\alpha_i+\alpha_j)})^2},$$

$$\frac{\partial F_i}{\partial \alpha_j} = \frac{\sum_{0\le k<l\le r-1}(k-l)^2 e^{(k+l)(\alpha_i+\alpha_j)}}{(\sum_{a=0}^{r-1} e^{a(\alpha_i+\alpha_j)})^2}, \quad j = 1, \ldots, n; j \ne i.$$

Since $e^{2k(\alpha_i+\alpha_j)} \le e^{(k+(k-1))(\alpha_i+\alpha_j)+Q_n}$, we have

$$\sum_{k=0}^{r-1} e^{2k(\alpha_i+\alpha_j)} \le \sum_{0\le k\ne l\le r-1} e^{(k+l)(\alpha_i+\alpha_j)} e^{Q_n}.$$

Therefore,

$$\begin{aligned}\frac{(1/2)\sum_{k\ne l} e^{(k+l)(\alpha_i+\alpha_j)}}{(\sum_{a=0}^{r-1} e^{a(\alpha_i+\alpha_j)})^2} &= \frac{(1/2)\sum_{k\ne l} e^{(k+l)(\alpha_i+\alpha_j)}}{\sum_{k\ne l} e^{(k+l)(\alpha_i+\alpha_j)} + \sum_{k=0}^{m-1} e^{2k(\alpha_i+\alpha_j)}} \\ &\ge \frac{\sum_{k\ne l} e^{(k+l)(\alpha_i+\alpha_j)}}{2(1+e^{Q_n})\sum_{k\ne l} e^{(k+l)(\alpha_i+\alpha_j)}} \ge \frac{1}{2(1+e^{Q_n})}.\end{aligned}$$

On the other hand,

$$\frac{(1/2)\sum_{k\neq l}(k-l)^2 e^{(k+l)(\alpha_i+\alpha_j)}}{(\sum_{a=0}^{r-1} e^{a(\alpha_i+\alpha_j)})^2} \le \frac{1}{2}\max_{k\neq l}(k-l)^2 \le \frac{r^2}{2}.$$

Consequently, when $-Q_n \le \alpha_i + \alpha_j \le Q_n$ for any $i \neq j$,

$$\frac{1}{2(1+e^{Q_n})} \le \Big|\frac{\partial F_i}{\partial \alpha_j}\Big| \le \frac{r^2}{2}.$$

Thus $F'(\boldsymbol{\alpha}) \in \mathcal{L}_n(m, M)$, where $m = (2(1+e^{Q_n}))^{-1}$ and $M = r^2/2$. Let

$$\mathbf{g}_{ij}(\alpha) = \Big(\frac{\partial^2 F_i}{\partial \alpha_1 \partial \alpha_j}, \ldots, \frac{\partial^2 F_i}{\partial \alpha_n \partial \alpha_j}\Big)^T.$$

It is easy to verify that

$$\frac{\partial^2 F_i}{\partial \alpha_i^2} = \sum_{j=1; j\neq i}^{n} \Bigg[\frac{(1/2)\sum_{k\neq l,a}(k-l)^2(k+l-2a)e^{(k+l+a)(\alpha_i+\alpha_j)}}{(\sum_{a=0}^{r-1} e^{a(\alpha_i+\alpha_j)})^3}\Bigg],$$

$$\frac{\partial^2 F_i}{\partial \alpha_j \partial \alpha_i} = \Bigg[\frac{(1/2)\sum_{k\neq l,a}(k-l)^2(k+l-2a)e^{(k+l+a)(\alpha_i+\alpha_j)}}{(\sum_{a=0}^{r-1} e^{a(\alpha_i+\alpha_j)})^3}\Bigg].$$

As $\sum_{k\neq l,a} e^{(k+l+a)(\alpha_i+\alpha_j)} \le (\sum_{a=0}^{r-1} e^{a(\alpha_i+\alpha_j)})^3$,

$$\Big|\frac{\partial^2 F_i}{\partial \alpha_i^2}\Big| \le (r-1)^3(n-1), \quad \Big|\frac{\partial^2 F_i}{\partial \alpha_j \partial \alpha_i}\Big| \le (r-1)^3. \tag{3.2}$$

This leads to $\|\mathbf{g}_{ii}(\alpha)\|_1 \le 2(n-1)(r-1)^3$, where $\|\mathbf{x}\|_1 = \sum_i |x_i|$ for a general vector $\mathbf{x}$. When $i \neq j$ and $k \neq i, j$,

$$\frac{\partial^2 F_i}{\partial \alpha_k \partial \alpha_j} = 0,$$

so we have $\|\mathbf{g}_{ij}(\alpha)\|_1 \le 2(r-1)^3$, for $j \neq i$. Consequently, for a vector $\mathbf{v}$,

$$\begin{aligned}
\max_i\Big\{\sum_j \Big[\frac{\partial F_i}{\partial \alpha_j}(\mathbf{x}) - \frac{\partial F_i}{\partial \alpha_j}(\mathbf{y})\Big] v_j\Big\} &\le \|\mathbf{v}\|_\infty \max_i \sum_{j=1}^n \Big|\frac{\partial F_i}{\partial \alpha_j}(\mathbf{x}) - \frac{\partial F_i}{\partial \alpha_j}(\mathbf{y})\Big| \\
&= \|\mathbf{v}\|_\infty \max_i \sum_{j=1}^n \Big|\int_0^1 g_i(t\mathbf{x} + (1-t)\mathbf{y})(\mathbf{x}-\mathbf{y})dt\Big| \\
&\le 4(r-1)^3(n-1)\|\mathbf{v}\|_\infty \|\mathbf{x}-\mathbf{y}\|_\infty.
\end{aligned}$$

This shows

$$m = \frac{1}{2(1+e^{Q_n})}, \quad M = \frac{r^2}{2}, \quad \phi_1 = 4(r-1)^3.$$

We derive ϕ_2. Since $a_{ij} \in [0, r-1]$ and d_i is a sum of $n-1$ independent random variables $a_{i,j}$, $j = 1, \ldots, n; j \neq i$, by Hoeffding's inequality (Hoeffding (1963)), we have

$$\begin{aligned} P(|d_i - \mathbb{E}(d_i)| \geq (r-1)\sqrt{(n-1)\log(n-1)}) \\ &\leq 2\exp\left(- \frac{2\left((r-1)\sqrt{(n-1)\log(n-1)}\right)^2}{(n-1)(r-1)^2} \right) \\ &\leq \frac{2}{(n-1)^2}. \end{aligned}$$

Since

$$P(\max_i |d_i - \mathbb{E}(d_i)| \geq x) \leq \sum_i P(|d_i - \mathbb{E}(d_i)| \geq x),$$

with probability approaching one, we have

$$\max_i |d_i - \mathbb{E}(d_i)| \leq \phi_2 \sqrt{(n-1)\log(n-1)},$$

where $\phi_2 = r - 1$. Thus, condition (C5) holds. We have

$$\frac{M^4 \phi_1 \phi_2}{m^5} \sqrt{\frac{\log n}{n}} = O\left(e^{5Q_n} \sqrt{\frac{\log n}{n}} \right).$$

If $e^{Q_n} = o\left((n/\log n)^{1/10}\right)$, then (2.6) is satisfied. By Theorem 1, the uniform consistency of $\widehat{\boldsymbol{\alpha}}$ follows.

Corollary 1. *If $e^{Q_n} = o\left((n/\log n)^{1/10}\right)$, then as n goes to infinity, with probability approaching one, $\widehat{\boldsymbol{\alpha}}$ exists and satisfies*

$$\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_\infty = O_p\left(e^{3Q_n} \sqrt{\frac{\log n}{n}} \right).$$

Now $d_i = \sum_{k \neq i} a_{ik}$ and $(\sum_i d_i)/2 = \sum_{1 \leq i < j \leq n} a_{ij}$ are sums of $n-1$ and $n(n-1)/2$ independent discrete random variables, respectively. The covariance matrix of $\mathbf{d} - \mathbb{E}(\mathbf{d})$ is $F'(\alpha)$, denoted by V, such that condition C1 holds. By the central limit theorem for the bounded case in (Loève (1977), p.289), we know that $v_{ii}^{-1/2}\{d_i - \mathbb{E}(d_i)\}$ and $(2v_{\cdot\cdot})^{-1/2}[\sum_{i=1}^n \{d_i - \mathbb{E}(d_i)\}]$ are asymptotically standard normal if v_{ii} diverges. Since

$$\frac{(n-1)}{(1+e^{Q_n})} \leq v_{ii} \leq (n-1)(r-1), \quad i = 1, \ldots, n; \quad v_{\cdot\cdot} \geq \frac{n(n-1)}{1+e^{Q_n}}.$$

If $e^{Q_n} = o(n)$, then $v_{ii} \to \infty$, $v_{\cdot\cdot} \to \infty$ and $M/m^2 = o(n)$ such that conditions of Proposition 1 hold. By (3.2), we have $|\mu''(\theta_{ij})| \leq (r-1)^3$. We have $M^6\phi_2\phi_3/m^9 = (r^2/2)^6 \cdot (r-1) \cdot (r-1)^3 \cdot (2(1+e^{Q_n}))^9$.

Corollary 2. *If $e^{Q_n} = o\left(n^{1/18}/(\log n)^{1/9}\right)$, then for any fixed $k \geq 1$, as $n \to \infty$, the vector consisting of the first k elements of $(B^{-1})^{1/2}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ is asymptotically standard multivariate normal, where $(B^{-1})^{1/2} = \operatorname{diag}(v_{11}^{1/2}, \ldots, v_{nn}^{1/2})$.*

3.2. Weighted graphs with continuous weights

When using the maximum entropy distributions to model weighted graphs with continuous weights, Hillar and Wibisono (2013) showed that a_{ij}, $1 \leq i \neq j \leq n$, are mutually independent exponential random variables with density

$$f(a) = \frac{1}{(\alpha_i + \alpha_j)} e^{-(\alpha_i + \beta_j)a}, \quad \alpha_i + \alpha_j > 0.$$

The moment estimating equations are

$$d_i = \sum_{k \neq i} (\hat{\alpha}_i + \hat{\alpha}_k)^{-1}, \quad i = 1, \ldots, n, \tag{3.3}$$

which are identical to the likelihood equations. Correspondingly,

$$\begin{aligned} F_i(\boldsymbol{\alpha}) &= d_i - \sum_{k \neq i} (\alpha_i + \alpha_k)^{-1}, \quad i = 1, \ldots, n, \\ F(\boldsymbol{\alpha}) &= (F_1(\boldsymbol{\alpha}), \ldots, F_n(\boldsymbol{\alpha}))'. \end{aligned}$$

The Jacobian matrix $F'(\boldsymbol{\alpha})$ of $F(\boldsymbol{\alpha})$ can be calculated as, for $i = 1, \ldots, n$,

$$\frac{\partial F_i}{\partial \alpha_j} = \frac{1}{(\alpha_i + \alpha_j)^2}, \quad j = 1, \ldots, n; j \neq i; \quad \frac{\partial F_i}{\partial \alpha_i} = \sum_{k \neq i} \frac{1}{(\alpha_i + \alpha_k)^2}.$$

From $\mathcal{L}_n(m, M)$ in (2.2), we can see that $F'(\boldsymbol{\alpha}) \in \mathcal{L}_n(m, M)$ with $m = Q_n^{-2}$ and $M = q_n^{-2}$ if $q_n \leq \alpha_i + \alpha_k \leq Q_n$, $i \neq k$. Theorem 1 can be applied. The consistency of the MLE has been derived by Hillar and Wibisono (2013) and the corresponding asymptotic normality has been done in Yan, Zhao, and Qin (2015).

3.3. Weighted graphs with discrete weights

When considering weighted graphs with discrete weights, Hillar and Wibisono (2013) showed that the a_{ij}, $1 \leq i \neq j \leq n$, are mutually independent geometric random variables with the probability a_{ij} at $a \in \mathbb{N}_0$

$$P(a_{ij} = a) = (1 - e^{-(\alpha_i + \alpha_j)}) e^{-(\alpha_i + \alpha_j)a}, \quad \alpha_i + \alpha_j > 0.$$

The moment estimating equations are

$$d_i = \sum_{k \neq i} \frac{e^{-(\hat{\alpha}_i + \hat{\alpha}_j)}}{1 - e^{-(\hat{\alpha}_i + \hat{\alpha}_j)}} = \frac{1}{e^{(\hat{\alpha}_i + \hat{\alpha}_j)} - 1}, \quad i = 1, \ldots, n. \tag{3.4}$$

Correspondingly,

$$F_i(\boldsymbol{\alpha}) = d_i - \sum_{k\neq i} \frac{1}{e^{(\alpha_i+\alpha_k)}-1}, \quad i=1,\ldots,n,$$
$$F(\boldsymbol{\alpha}) = (F_1(\boldsymbol{\alpha}),\ldots,F_n(\boldsymbol{\alpha}))'.$$

The Jacobin matrix $F'(\boldsymbol{\alpha})$ of $F(\boldsymbol{\alpha})$ can be calculated as follows. For $i=1,\ldots,n$,

$$\frac{\partial F_i}{\partial \alpha_j} = \frac{e^{(\alpha_i+\alpha_j)}}{(e^{(\alpha_i+\alpha_j)}-1)^2}, \quad j=1,\ldots,n; j\neq i; \quad \frac{\partial F_i}{\partial \alpha_i} = \sum_{k\neq i} \frac{e^{(\alpha_i+\alpha_k)}}{(e^{(\alpha_i+\alpha_k)}-1)^2}.$$

We see that $-F'(\boldsymbol{\alpha}) \in \mathcal{L}_n(M,m)$ with $m = e^{Q_n}(e^{Q_n}-1)^{-2}$ and $M = e^{q_n}(e^{q_n}-1)^{-2}$ if $q_n \le \alpha_i+\alpha_k \le Q_n$, $i\neq k$. Theorem 1 can be applied. The consistency of the MLE has been derived by Hillar and Wibisono (2013) and the corresponding asymptotic normality has been done in Yan, Zhao, and Qin (2015).

3.4. Poisson models

Assume that each a_{ij} is Poisson distributed with parameter $\alpha_i+\alpha_j>0$. This model has been considered by Ranola et al. (2010) and Sadinle (2012) in its directed version. For convenience, we transform α_i to e^{α_i} $(i=1,\ldots,n)$ with

$$P(a_{ij}=a) = \frac{e^{a(\alpha_i+\alpha_j)}}{a!}\exp(e^{\alpha_i+\alpha_j}).$$

The moment equations are

$$d_i = \sum_{j\neq i;j=1}^{n} e^{\hat{\alpha}_i+\hat{\alpha}_j}, \quad i=1,\ldots,n. \tag{3.5}$$

With $F_i(\boldsymbol{\alpha}) = d_i - \mathbb{E}(d_i) = d_i - \sum_{j\neq i} e^{\alpha_i+\alpha_j}$, $i=1,\ldots,n$, we have

$$\frac{\partial F_i(\boldsymbol{\alpha})}{\partial \alpha_i} = -\sum_{j\neq i} e^{\alpha_i+\alpha_j}, \quad \frac{\partial F_i(\boldsymbol{\alpha})}{\partial \alpha_j} = -e^{\alpha_i+\alpha_j}, j\neq i.$$

Consider the parameter space $D=\{\boldsymbol{\alpha} : -Q_n \le \alpha_i+\alpha_j \le Q_n\}$. Then $-F'(\boldsymbol{\alpha}) \in \mathcal{L}_n(m,M)$, where $m = 1/e^{Q_n}$ and $M = e^{Q_n}$. Let $g_{kij} = \partial^2 F_i/(\partial\alpha_k\partial\alpha_j)$ and $\mathbf{g}_{ij} = (g_{1ij},\ldots,g_{nij})^\top$. We have

$$g_{kij} = \begin{cases} -\sum_{l\neq i} e^{\alpha_i+\alpha_l}, & k=i=j, \\ -e^{\alpha_i+\alpha_k}, & k\neq i, i=j, \\ -e^{\alpha_i+\alpha_k}, & k=j, i\neq j, \\ -e^{\alpha_k+\alpha_j}, & k=i, i\neq j, \\ 0, & \text{otherwise}. \end{cases}$$

By the mean-value theorem for vector-valued functions (Lang (1993, p.341)), we have

$$\begin{aligned}\|(F'(\mathbf{x})-F'(\mathbf{y}))\mathbf{v}\|_\infty &= \|\mathbf{v}\|_\infty \max_i \sum_j \Big|\frac{\partial F_i(\mathbf{x})}{\partial \alpha_j} - \frac{\partial F_i(\mathbf{y})}{\partial \alpha_j}\Big| \\ &= \|\mathbf{v}\|_\infty \max_i \sum_j \Big| \int_0^1 \mathbf{g}_{ij}^\top (t\mathbf{x}+(1-t)\mathbf{y})dt(\mathbf{x}-\mathbf{y})\Big| \\ &\le \|\mathbf{v}\|_\infty \|(\mathbf{x}-\mathbf{y})\|_\infty \max_i \sum_{j,k} \Big| \int_0^1 g_{kij}(t\mathbf{x}+(1-t)\mathbf{y})dt\Big| \\ &\le 4(n-1)e^{Q_n}\|\mathbf{v}\|_\infty \|(\mathbf{x}-\mathbf{y})\|_\infty.\end{aligned}$$

Therefore, we can choose $\phi_1 = 4e^{Q_n}$. Lemma C.2 in Appendix C shows

$$\phi_2 = 2c_3\sqrt{\frac{2e^{4Q_n}}{\gamma}}, \tag{3.6}$$

where c_3 is an absolute constant. We have

$$\begin{aligned}\frac{M^4\phi_1\phi_2}{m^5}\sqrt{\frac{\log(n-1)}{(n-1)}} &= e^{9Q_n}\times 4e^{Q_n}\times 2c_3\sqrt{\frac{2e^{4Q_n}}{\gamma}} \times \sqrt{\frac{\log(n-1)}{(n-1)}} \\ &= O(e^{12Q_n}\sqrt{\frac{\log n}{n}}).\end{aligned}$$

Corollary 3. *If* $e^{Q_n} = o((n/\log n)^{1/24})$, *then*

$$\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|_\infty = O_p\left(e^{7Q_n}\sqrt{\frac{\log n}{n}}\right) = o_p(1).$$

The covariance matrix of $\mathbf{d}-\mathbb{E}(\mathbf{d})$ is $V=-F'(\alpha)\in\mathcal{L}_n(m,M)$, so C1 holds. Note that $d_i=\sum_{k\neq i} a_{ik}$ and $(\sum_i d_i)/2 = \sum_{1\le i<j\le n} a_{ij}$ are sums of $n-1$ and $n(n-1)/2$ Poisson random variables. As the third moment of the Poisson with parameter λ is $\lambda+3\lambda^2+\lambda^3$, we have

$$\frac{\sum_{j\neq i}\mathbb{E}(a_{ij}^3)}{v_{ii}^{3/2}} \le \frac{(n-1)(e^{Q_n}+3e^{2Q_n}+e^{3Q_n})}{(n-1)^{3/2}e^{-3Q_n/2}} \le \frac{5e^{7Q_n/2}}{(n-1)^{1/2}}.$$

If $e^{Q_n}=o(n^{1/7})$, then the above expression goes to zero. This shows that the condition for the Lyapunov Central Limit Theorem, holds. Therefore, $v_{ii}^{-1/2}\{d_i-\mathbb{E}(d_i)\}$ and $(2v_{\cdot\cdot})^{-1/2}\sum_i\{d_i-\mathbb{E}(d_i)\}$ are asymptotically standard normal if $e^{Q_n}=o(n^{1/7})$. Thus, C2 holds. If $e^{Q_n}=o((n/\log n)^{1/24})$, then by Corollary 3,

$$\begin{aligned}|\mu''(\hat\theta_{ij})| &= \exp(t_{ij}(\hat\alpha_i+\hat\alpha_j)+(1-t_{ij})(\alpha_i+\alpha_j)) \\ &= O_p(e^{Q_n}\times e^{e^{7Q_n}(\log n)^{1/2}n^{-1/2}}).\end{aligned}$$

Thus, C6 holds with ϕ_3 as above. If $e^{Q_n} = n^{1/36}/(\log n)^{1/18}$, then

$$\frac{M^6\phi_2\phi_3\log n}{m^9 n^{1/2}} = \frac{\log n}{n^{1/2}} \times e^{15Q_n} \times 2c_3\sqrt{\frac{2e^{4Q_n}}{\gamma}} \times e^{Q_n} \times e^{e^{7Q_n}(\log n)^{1/2}n^{-1/2}} = o(1).$$

Corollary 4. *If $e^{Q_n} = n^{1/36}/(\log n)^{1/18}$, then for any fixed $k \geq 1$, as $n \to \infty$, the vector consisting of the first k elements of $(B^{-1})^{1/2}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ is asymptotically standard multivariate normal, where $(B^{-1})^{1/2} = \mathrm{diag}(v_{11}^{1/2}, \ldots, v_{nn}^{1/2})$.*

3.5. Rayleigh distribution

We consider a Rayleigh distribution for edges in which the moment equations are different from the maximum likelihood equations. The density for the Rayleigh distribution with the parameter $\sigma > 0$ is $f(x) = x\sigma^{-2}\exp(-x^2/(2\sigma^2))$ (Papoulis (1991), p.78). Let a_{ij} have the Rayleigh distribution with the parameter $e^{(\alpha_i+\alpha_j)/2}$, so that the density of a_{ij} is

$$f(a) = \frac{a}{e^{\alpha_i+\alpha_j}} e^{-a^2/(2e^{\alpha_i+\alpha_j})}.$$

As $\mathbb{E}(a_{ij}) = \sqrt{\pi/2}e^{(\alpha_i+\alpha_j)/2}$, the moment equations are

$$d_i = \sum_{j\neq i; j=1}^{n} \sqrt{\frac{\pi}{2}} e^{(\hat{\alpha}_i+\hat{\alpha}_j)/2}, \quad i = 1, \ldots, n, \tag{3.7}$$

while the maximum likelihood equations are

$$\frac{1}{2}\sum_{j\neq i} a_{ij}^2 e^{-(\bar{\alpha}_i+\bar{\alpha}_j)} = n-1, \quad i = 1, \ldots, n,$$

where $\bar{\alpha}_i$ is the MLE of α_i. The moment equations are simpler than the maximum likelihood equations. Since (3.7) is similar to (3.5) for the Poisson model, deriving the asymptotic results for the moment estimator is also similar, and is omitted.

4. Discussion

We have established a unified asymptotic theory for the moment estimator in a class of undirected random graph models parameterized by the strengths of vertices, and have illustrated applications to the generalized β-model, maximum entropy models on graphs, and the Poisson model. The moment estimator is induced by the moment equations based on the degree sequence; in particular, the MLE is exactly the moment estimator in exponential family distributions on graphs with the degree sequence as the sufficient statistic. The numerical evaluations on asymptotic properties of the moment estimator have been provided in Yan and Xu (2013), Yan, Zhao, and Qin (2015).

If the maximum likelihood equations are identical to the moment equations, then condition C3 implies C1, the Fisher information matrix of the parameter vector and the covariance matrix of the degree sequence are the same. Conditions C3–C5 crucially depend on values q_n and Q_n that measure the minimum and maximum values of the set $\{\alpha_i + \alpha_j : i \neq j\}$. If q_n is too small or Q_n is too large, then the moment estimate doesn't exist. In contrast with the conditions guaranteeing the consistency of the moment estimator, those guaranteeing its asymptotic normality seems more severe, as illustrated by the examples in Section 3. It is of interest to see whether these conditions can be relaxed.

In this paper, we assume that the network edges are mutually independent. This assumption holds when we only consider the distribution of the vertex degrees with them as the exclusively sufficient statistics. If edges are dependent, as long as the moment condition is correct, we should be able to obtain a consistent estimator since our method is driven by moment condition. However, without the mutual independence assumption, the resulting estimator's asymptotic distribution is not clear. A more complex dependent case is that other network statistics, such as triangle measuring transitivity effect, are involved. In this case, we still have the moment equations based on these network statistics as well as the degree sequence. In such situation, it is intricate to investigate the asymptotic properties of the moment estimator since the asymptotic background is involved with not only an increasing dimension of the parameter space but also dependent edges (Fienberg (2012)). It would be of interest to see whether the current method can be extended after some modifications. On the other hand, if the exponential family for network statistics such as k-stars and triangles is used, there is the problem of model degeneracy. See (Chatterjee and Diaconis (2013), Schweinberger (2011), Strauss (1986)). A probability distribution for random graphs needs to be properly chosen in order to avoid this problem.

Acknowledgements

We are very grateful to two anonymous referees and the Editor Hsin-Cheng Huang for their valuable comments that have greatly improved the manuscript. This research was supported by National Natural Science Foundation of China (11341001, 11401239, 11271147, 11131002, 11271032). Yan's research was supported by Postdoctoral Science Foundation of China (2014M552064).

Appendix A.

For a subset $C \subset \mathbb{R}^n$, let C^0 and $\overline{C}$ denote the interior and closure of C in R^n, respectively. Let $\Omega(\mathbf{x}, r)$ denote the open ball $\{\mathbf{y} : \|\mathbf{x} - \mathbf{y}\| < r\}$, and $\overline{\Omega(\mathbf{x}, r)}$ be its closure. We use Newton's iterative sequence to prove the existence and consistency of the moment estimates relying on results of Gragg and Tapia (1974).

Proposition A.1. *Let $F(\mathbf{x}) = (F_1(\mathbf{x}), \ldots, F_n(\mathbf{x}))^\top$ be a function vector on $\mathbf{x} \in \mathbb{R}^n$. Assume that the Jacobian matrix $F'(\mathbf{x})$ is Lipschitz continuous on an open convex set D with the Lipschitz constant λ. Given $\mathbf{x}_0 \in D$, assume that $[F'(\mathbf{x}_0)]^{-1}$ exists,*

$$\|[F'(\mathbf{x}_0)]^{-1}\|_\infty \le \aleph, \quad \|[F'(\mathbf{x}_0)]^{-1}F(\mathbf{x}_0)\|_\infty \le \delta, \quad h = 2\aleph\lambda\delta \le 1,$$

$$\Omega(\mathbf{x}_0, t^*) \subset D^0, \quad t^* := \frac{2}{h}(1-\sqrt{1-h})\delta = \frac{2}{1+\sqrt{1-h}}\delta \le 2\delta,$$

where $\aleph$ and δ are positive constants that may depend on $\mathbf{x}_0$ and the dimension n of $\mathbf{x}_0$. Then the Newton iterates $\mathbf{x}_{k+1} = \mathbf{x}_k - [F'(\mathbf{x}_k)]^{-1}F(\mathbf{x}_k)$ exist and $\mathbf{x}_k \in \Omega(\mathbf{x}_0, t^*) \subset D^0$ for all $k \ge 0$; $\widehat{\mathbf{x}} = \lim \mathbf{x}_k$ exists, $\widehat{\mathbf{x}} \in \overline{\Omega(\mathbf{x}_0, t^*)} \subset D$ and $F(\widehat{\mathbf{x}}) = 0$.

Thus if $t^* \to 0$, then $\|\widehat{\boldsymbol{x}} - \boldsymbol{x}_0\| = o(1)$.

For a matrix $A = (a_{ij})$, take $\|A\| := \max_{i,j} |a_{ij}|$. Yan and Xu (2013) derived the following.

Proposition A.2. *If $V \in \mathcal{L}_n(m, M)$ at (2.2), and n is large enough,*

$$\|V^{-1} - S\| \le \frac{c_1 M^2}{m^3(n-1)^2},$$

where S is defined at (2.3) and c_1 is a constant that does not depend on M, m, and n.

Lemma A.1. If $V \in \mathcal{L}_n(m, M)$, for large enough n,

$$\|V^{-1}\|_\infty \le \|V^{-1} - S\|_\infty + \|S\|_\infty \le \frac{c_1 n M^2}{m^3(n-1)^2} + \frac{1}{m}\left(\frac{1}{n(n-1)} + \frac{1}{n-1}\right) \le \frac{c_2 M^2}{nm^3},$$

where c_2 is a constant that does not depend on M, m, and n.

Appendix A: Proof of Proposition 1

We have $[S(\mathbf{d} - \mathbb{E}(\mathbf{d}))]_i = (d_i - \mathbb{E}(d_i))/v_{ii} - \sum_i (d_i - \mathbb{E}(d_i))/(v_{\cdot\cdot})$. Since $V \in \mathcal{L}_n(m, M)$,

$$\frac{\max_i v_{ii}}{v_{\cdot\cdot}} \le \frac{(n-1)M}{n(n-1)m^2} = \frac{M/m^2}{n}.$$

By (C2),

$$\sqrt{v_{ii}}[S(\mathbf{d} - \mathbb{E}(\mathbf{d}))]_i = \frac{d_i - \mathbb{E}(d_i)}{\sqrt{v_{ii}}} + O_p\left(\sqrt{\frac{v_{ii}}{v_{\cdot\cdot}}}\right) = \frac{d_i - \mathbb{E}(d_i)}{\sqrt{v_{ii}}} + O_p\left(\sqrt{\frac{M/m^2}{n}}\right).$$

For any fixed k, if v_{ii} diverges, then

$$\begin{aligned}\frac{d_i - \mathbb{E}(d_i)}{\sqrt{v_{ii}}} &= \frac{\sum_{l=1}^{k}(d_{il} - \mathbb{E}(d_{il}))}{\sqrt{v_{ii}}} + \frac{\sum_{l=k+1}^{n}(d_{il} - \mathbb{E}(d_{il}))}{\sqrt{v_{ii}}} \\ &= \frac{\sum_{l=k+1}^{n}(d_{il} - \mathbb{E}(d_{il}))}{\sqrt{v_{ii}}} + o(1).\end{aligned}$$

For any fixed k, $\sum_{l=k+1}^{n}(d_{il} - \mathbb{E}(d_{il}))$, $i = 1, \ldots, k$, are independent. The condition $M/m^2 = o(n)$ implies $mn \to \infty$. The latter further implies $v_{ii} \to \infty$ as $n \to \infty$. Therefore, for any fixed k, $(d_i - \mathbb{E}(d_i))/v_{ii}^{1/2}$, $i = 1, \ldots, k$, are asymptotically independent standard normal.

Appendix B: Proof of Theorem 1

To prove the first part of this theorem, it is sufficient to show that the Newton-Kantovorich conditions hold. We only give the proof in case $F'(\boldsymbol{\alpha}) \in \mathcal{L}_n(m, M)$. The proof when $-F'(\boldsymbol{\alpha}) \in \mathcal{L}_n(m, M)$ is similar, and we omit it. In the Newton's iterative step, we take the true parameter vector $\boldsymbol{\alpha}$ as the starting point $\boldsymbol{\alpha}^{(0)} := \boldsymbol{\alpha}$. Let $V = F'(\boldsymbol{\alpha}) \in \mathcal{L}_n(m, M)$ and $W = V^{-1} - S$. By Lemma A.1, we have $\aleph = c_2 M^2/(nm^3)$. Note that $F(\boldsymbol{\alpha}) = \mathbf{d} - \mathbb{E}(\mathbf{d})$. Assuming (2.4), by Proposition A.2 we have

$$\begin{aligned}\|[F'(\boldsymbol{\alpha})]^{-1}F(\boldsymbol{\alpha})\|_\infty &\le n\|W\|\|F(\boldsymbol{\alpha})\|_\infty + \max_i \frac{|F_i(\boldsymbol{\alpha})|}{v_{ii}} + \frac{1}{v_{\cdot\cdot}}\sum_{i=1}^{n}|F_i(\boldsymbol{\alpha})| \\ &\le \left(\frac{c_1 n M^2}{(n-1)^2 m^3} + \frac{2}{m(n-1)}\right) \times \phi_2\sqrt{(n-1)\log(n-1)} \\ &\le \frac{c_3 M^2 \phi_2}{m^3}\sqrt{\frac{\log(n-1)}{(n-1)}},\end{aligned}$$

where c_3 is a constant. Therefore, we can choose

$$\delta = \frac{c_3 M^2 \phi_2}{m^3}\sqrt{\frac{\log(n-1)}{(n-1)}}.$$

If (2.6) holds, by C4,

$$\begin{aligned}h = 2\aleph\lambda\delta &= \frac{c_2 M^2}{(n-1)m^2} \times (n-1)\phi_1 \times \frac{c_3 M^2 \phi_2}{m^3}\sqrt{\frac{\log(n-1)}{(n-1)}} \\ &= \frac{M^4\phi_1\phi_2}{m^5}\sqrt{\frac{\log(n-1)}{(n-1)}} = o(1).\end{aligned}$$

By Proposition A.1, $\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\|_\infty = O((M^2\phi_2/m^3)\sqrt{\log n/n})$. By C5, (2.4) holds with probability approaching one such that (2.7) holds if (2.5) is satisfied. This shows the first part.

For the second part, let $\hat{\gamma}_{ij} = \hat{\alpha}_i + \hat{\alpha}_j - \alpha_i - \alpha_j$ and assume

$$\max_{i\neq j} |\hat{\gamma}_{ij}| = O(\frac{M^2\phi_2}{m^3}\sqrt{\frac{\log n}{n}}). \tag{B.1}$$

Condition C4 holds, $V = \mathrm{Cov}(\mathbf{d}-\mathbb{E}(\mathbf{d})) \in \mathcal{L}_n(m, M)$. Let $W = V^{-1} - S$ and $U = \mathrm{Cov}[W\{\mathbf{d}-\mathbb{E}(\mathbf{d})\}]$. Then

$$\begin{aligned} U &= WVW^T = (V^{-1}-S) - S(I-VS),\\ \{S(I-VS)\}_{i,j} &= \frac{(\delta_{i,j}-1)v_{i,j}}{v_{i,i}v_{j,j}} + \frac{1}{v_{\cdot\cdot}}. \end{aligned}$$

According the definition of $\mathcal{L}_n(m, M)$, since

$$\begin{aligned} |\{S(I-VS)\}_{i,j}| &\le \max\left\{\frac{M}{(n-1)^2m^2}, \frac{1}{n(n-1)m}\right\} \le \frac{2M}{(n-1)^2m^2},\\ \|U\| \le \|V^{-1}-S\| + \|S(I-VS)\| &\le \frac{c_1M^2}{m^3(n-1)^2} + \frac{2M}{m^2(n-1)^2}. \end{aligned}$$

If $M^2/m^3 = o(n)$, then

$$\|U\| = o(n^{-1}). \tag{B.2}$$

For $i = 1, \ldots, n$, by a Taylor's expansion, we have

$$\begin{aligned} d_i - E(d_i) &= \sum_{j\neq i}(\mu(\hat{\alpha}_i+\hat{\alpha}_j) - \mu(\alpha_i+\alpha_j))\\ &= \sum_{j\neq i}[\mu'(\alpha_i+\alpha_j)(\mu(\hat{\alpha}_i+\hat{\alpha}_j) - \mu(\alpha_i+\alpha_j))] + h_i, \end{aligned}$$

where $h_i = (1/2)\sum_{j\neq i}\mu''(\hat{\theta}_{ij})[((\hat{\alpha}_i+\hat{\alpha}_j)-(\alpha_i+\alpha_j))]^2$, and $\hat{\theta}_{ij} = t_{ij}(\alpha_i+\alpha_j) + (1-t_{ij})(\hat{\alpha}_i+\hat{\alpha}_j)$, $0 < t_{ij} < 1$. Writing the above expressions in matrices, $\mathbf{d} - \mathbb{E}\mathbf{d} = V(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}) + \mathbf{h}$, or, equivalently,

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha} &= V^{-1}(\mathbf{d}-\mathbb{E}\mathbf{d}) + V^{-1}\mathbf{h}\\ &= S(\mathbf{d}-\mathbb{E}\mathbf{d}) + W(\mathbf{d}-\mathbb{E}\mathbf{d}) + V^{-1}\mathbf{h}, \end{aligned}$$

where $\mathbf{h} = (h_1, \ldots, h_n)^T$. Assume that $\mu''(\hat{\theta}_{ij}) = O(\phi_3)$. Then

$$|h_i| \le \frac{1}{2}(n-1)\phi_3\hat{\gamma}_{ij}^2.$$

Therefore,

$$\begin{aligned}
|(V^{-1}\mathbf{h})_i| &= |(S\mathbf{h})_i| + |(W\mathbf{h})_i| \\
&\le \max_i \frac{|h_i|}{v_{ii}} + \frac{\sum_i |\mathbf{h}_i|}{v_{\cdot\cdot}} + \|W\| \sum_i |h_i| \\
&\le O\Big(\frac{3\phi_3 \hat{\gamma}_{ij}^2}{2m} + \frac{c_1 M^2}{m^3 (n-1)^2} \times \frac{1}{2} n(n-1)\phi_3 \hat{\gamma}_{ij}^2 \Big) \\
&\le O\bigg(\frac{M^2 \phi_3}{m^3} \times \Big(\frac{M^2 \phi_2}{m^3} \sqrt{\frac{\log n}{n}} \Big)^2 \bigg) = O\Big(\frac{M^6 \phi_2 \phi_3 \log n}{m^9 n} \Big).
\end{aligned}$$

If $M^6 \phi_2 \phi_3 \log n / m^9 = o(n^{1/2})$, then, $(V^{-1}\mathbf{h})_i = o(n^{-1/2})$. By the first part of this theorem, (B.1) holds with probability approaching 1. And by C6, $\mu''(\hat{\theta}_{ij}) = O_p(\phi_3)$. Consequently, by (B.2), we have

$$(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})_i = [S(\mathbf{d} - \mathbb{E}(\mathbf{d}))]_i + o_p(n^{-1/2}).$$

Therefore, the second part of the theorem follows from Proposition 1.

Appendix C: Proof of (3.6)

A real-valued random variable X is *sub-exponential* with parameter $\kappa > 0$ if

$$\mathbb{E}[|X|^p]^{1/p} \le \kappa p \quad \text{for all } p \ge 1.$$

If X is a κ-sub-exponential random variable with finite first moment, then the centered random variable $X - \mathbb{E}[X]$ is also sub-exponential with parameter 2κ since

$$\big[\mathbb{E}\big(\big|X - \mathbb{E}[X]\big|^p\big)\big]^{1/p} \le [\mathrm{E}\left(|X|^p\right)]^{1/p} + \big|\mathbb{E}[X]\big| \le 2[\mathrm{E}\left(|X|^p\right)]^{1/p}.$$

Theorem C.1 (Vershynin (2012), Corollary 5.17)**.** *Let $X_1, \ldots, X_n$ be independent centered random variables, and suppose each X_i is sub-exponential with parameter κ_i. Let $\kappa = \max_{1\le i\le n} \kappa_i$. Then for every $\epsilon \ge 0$,*

$$\mathbb{P}\bigg(\Big| \frac{1}{n} \sum_{i=1}^n (X_i - E(X_i)) \Big| \ge \epsilon \bigg) \le 2 \exp\Big[- n\gamma \cdot \min\Big(\frac{\epsilon^2}{\kappa^2}, \frac{\epsilon}{\kappa} \Big) \Big],$$

where $\gamma > 0$ is an absolute constant.

Lemma C.1. *Let X be a Poisson random variable with parameter $\lambda > 0$. Then X is sub-exponential with parameter $c_3 e^{-\lambda}$, and the centered random variable $X - \lambda$ is sub-exponential with parameter $2c_3 e^{-\lambda}$, where $1 < c_3 < 3$ is an absolute constant.*

Proof. Let $\lceil p \rceil$ denote the minimum integer among all integers larger than p. Then $p \le \lceil p \rceil + 1$. Direct calculation gives

$$\begin{aligned}
E(X^{\lceil p \rceil}) &= \sum_{k=1}^{\lceil p \rceil} \left\{ \begin{matrix} \lceil p \rceil \\ k \end{matrix} \right\} \lambda^k = \sum_{k=1}^{\lceil p \rceil} \Big[\lambda^k \times \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} j^{\lceil p \rceil} \Big] \\
&\le \sum_{k=1}^{\lceil p \rceil} \sum_{j=0}^{k} \frac{\lambda^k k!}{j!(k-j)!k!} j^{\lceil p \rceil} \le \lambda^{\lceil p \rceil} \sum_{k=1}^{\lceil p \rceil} \sum_{j=0}^{k} j^{\lceil p \rceil} \\
&\le \lambda^{\lceil p \rceil} \sum_{k=1}^{\lceil p \rceil} \int_1^{k+1} x^{\lceil p \rceil} dx \le \lambda^{\lceil p \rceil} \sum_{k=1}^{\lceil p \rceil} \frac{(k+1)^{\lceil p \rceil + 1}}{\lceil p \rceil + 1} \\
&\le \lambda^{\lceil p \rceil} \frac{1}{\lceil p \rceil + 1} \int_2^{\lceil p \rceil + 1} x^{\lceil p \rceil + 1} dx = \lambda^{\lceil p \rceil} \frac{(\lceil p \rceil + 1)^{\lceil p \rceil + 2}}{(\lceil p \rceil + 1)(\lceil p \rceil + 2)}.
\end{aligned}$$

Therefore,

$$[E(X^{\lceil p \rceil + 1})]^{1/p} \le (\lambda^{\lceil p \rceil + 1})^{1/p} [\frac{(\lceil p \rceil + 2)^{\lceil p \rceil + 3}}{(\lceil p \rceil + 2)(\lceil p \rceil + 3)}]^{1/p} \le [\lambda(\lceil p \rceil + 2)]^{1+1/p} \le c_3 \lambda^2 p,$$

where $1 < c_3 < 3$ is an absolute constant.

Lemma C.2. With probability at least $1 - 2n/(n-1)^2$, we have

$$\max_{1 \le i \le n} |d_i - E(d_i)| \le 2c_3 \sqrt{\frac{2e^{4Q_n}}{\gamma}(n-1)\log(n-1)}.$$

Proof. With a_{ij} a Poisson random variable with the parameter $e^{-Q_n} \le \lambda = e^{\alpha_i + \alpha_j} \le e^{Q_n}$, by Lemma C.1, $a_{ij} - e^{(\alpha_i + \alpha_j)}$ is sub-exponential with parameter $2c_3 e^{2(\alpha_i + \alpha_j)} \le 2c_3 e^{2Q_n}$. For each $i = 1, \ldots, n$, the random variables $(a_{ij} - e^{\alpha_i + \alpha_j}, j \ne i)$ are independent sub-exponential random variables, so we can apply Theorem C.1 with $\kappa = 2c_3 e^{2Q_n}$ and

$$\epsilon = \kappa \left(\frac{2\log(n-1)}{\gamma(n-1)} \right)^{1/2}.$$

Assume n is sufficiently large such that $\epsilon/\kappa = \sqrt{2\log(n-1)/\gamma(n-1)} \le 1$. Then by Theorem C.1, for each $i = 1, \ldots, n$ we have

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{n-1} |d_i - Ed_i^*| \ge \kappa \left(\frac{2\log(n-1)}{\gamma(n-1)} \right)^{1/2} \right) &\le 2\exp\left(-(n-1)\gamma \cdot \frac{2\log n}{\gamma(n-1)} \right) \\
&= \frac{2}{(n-1)^2}.
\end{aligned}$$

By the union bound,

$$\begin{aligned}
\mathbb{P}&\left(\|\mathbf{d}-\mathbf{d}^*\|_\infty \geq 2c_3\sqrt{\frac{2e^{Q_n}}{\gamma}(n-1)\log(n-1)} \right) \\
&\leq \sum_{i=1}^{n} \mathbb{P}\left(|d_i - d_i^*| \geq 2c_3\sqrt{\frac{2e^{Q_n}}{\gamma}(n-1)\log(n-1)}\right) \\
&\leq \frac{2n}{(n-1)^2}.
\end{aligned}$$

References

Bernard H. R., Killworth P. and Sailer L. (1980). Informant accuracy in social network data IV. *Social Networks* **2**, 191-218.

Bernard H. R., Killworth P. and Sailer L. (1982). Informant accuracy in social network data V. *Social Sci. Res.* **11**, 30-66.

Bickel, P. J., Chen, A. and Levina, E. (2012). The method of moments and degree distributions for network models. *Ann. Statist.* **39**, 2280–2301.

Billingsley, P. (1995). *P*robability and measure. 3rd edition. Wiley, New York.

Blitzstein, J. and Diaconis, P. (2010). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Math.* **6**, 489-522.

Britton, T., Deijfen, M. and Martin-Löf A. (2006). Generating simple random graphs with prescribed degree distribution. *J. Statist. Phys.* **124**, 1377-1397.

Chatterjee S. and Diaconis P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.* **41**, 2428-2461.

Chatterjee S., Diaconis P. and Sly A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21**, 1400-1435.

Chung, F. and Lu, L. (2002). The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA* **99**, 15879-15882.

Erdös, P. and Rényi, A. (1959). On random graphs. I. *Publicationes Math.* **6**, 290–297.

Fienberg S. E. (2012). A brief history of statistical models for network analysis and open challenges. *J. Comput. Graph. Statist.* **21**, 825-839.

Fowler, J. H. (2006). Connecting the congress: A study of cosponsorship networks. *Polit. Anal.* **14**, 456-487.

Gragg, W. B. and Tapia, R. A. (1974). Optimal error bounds for the Newton-Kantorovich theorem. *SIAM J. Numer. Anal.* **11**, 10-13.

Hillar C. J. and Wibisono A. (2013). Maximum entropy distributions on graphs. Available at `http://arxiv.org/abs/1301.3321`.

Hoeffding W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13-30.

Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *J. Amer. Statist. Assoc.* **76**, 33-65.

Jackson, M. O.(2008). *Social and Economic Networks.* Princeton University Press, Princeton.

Janson, S. (2010). Asymptotic equivalence and contiguity of some random graphs. *Random Struct. Algor.* **36**, 26-45.

Lang S. (1993). *Real and Functional Analysis.* Springer.

Lauritzen, S. L. (2003). Rasch models with exchangeable rows and columns. In *Bayesian Statistics* **7** (Edited by J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West), 215-232, Oxford University Press.

Lauritzen, S. L. (2008). Exchangeable Rasch matrices. *Rendiconti di Math.* **28**, 83-95.

Lindeberg J. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15**, 211-225.

Loève, M. (1977). *Probability Theory.* 4th edition. Springer-Verlag, New York.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proc. Nat. Acad. Sci. USA* **98**, 404-409.

Olhede S. C. and Wolfe P. J. (2012). Degree-based network models. Available at `http://arxiv.org/abs/1211.6537`.

Papoulis A. (1991). *Probability, Random Variables, and Stochastic Processes.* 3rd edition. McGraw-Hill, New York.

Perry, P. O. and Wolfe, P. J. (2012). Null models for network data. Available at `http://arxiv.org/abs/1201.5871`.

Ranola, J. M. O., Ahn, S., Sehl, M., Smith, D. J. and Lange, K. (2010). A Poisson model for random multigraphs. *Bioinformatics* **26**, 2004-2011.

Rinaldo, A., Petrović, S. and Fienberg, S. E. (2013). Maximum likelihood estimation in the Beta model. *Ann. Statist.* **41**, 1085-1110.

Sadinle M. (2012). The strength of arcs and edges in interaction networks: elements of a model-based approach. Available at `http://arxiv.org/abs/1211.0312`.

Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *J. Amer. Statist. Assoc.* **106**, 1361-1370.

Strauss, D. (1986). On a general class of models for interaction. *SIAM Rev.* **28**, 513-527.

Vershynin R. (2012). *Compressed Sensing, Theory and Applications.* Cambridge University Press.

Yan, T. and Xu, J. (2013). A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika* **100**, 519-524.

Yan T., Zhao Y. and Qin H. (2015). Asymptotic normality in the maximum entropy models on graphs with an increasing number of parameters. *J. Multivariate Anal.* **133**, 61-76.

Zhao, Y., Levina, E. and Zhu, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.* **40**, 2266-2292.

Department of Statistics, Central China Normal University, Wuhan 430079, China.

E-mail: tingyanty@mail.ccnu.edu.cn

Department of Statistics, Central China Normal University, Wuhan 430079, China.

E-mail: qinhong@mail.ccnu.edu.cn

Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing 100871, China.

E-mail: hansheng@gsm.pku.edu.cn

(Received May 2014; accepted January 2015)