ASYMPTOTICS IN DIRECTED EXPONENTIAL RANDOM GRAPH MODELS WITH AN INCREASING BI-DEGREE SEQUENCE

BY TING YAN[1], CHENLEI LENG AND JI ZHU[2]

Central China Normal University, University of Warwick and University of Michigan

Although asymptotic analyses of undirected network models based on degree sequences have started to appear in recent literature, it remains an open problem to study statistical properties of directed network models. In this paper, we provide for the first time a rigorous analysis of directed exponential random graph models using the in-degrees and out-degrees as sufficient statistics with binary as well as continuous weighted edges. We establish the uniform consistency and the asymptotic normality for the maximum likelihood estimate, when the number of parameters grows and only one realized observation of the graph is available. One key technique in the proofs is to approximate the inverse of the Fisher information matrix using a simple matrix with high accuracy. Numerical studies confirm our theoretical findings.

1. Introduction. Recent advances in computing and measurement technologies have led to an explosion in the amount of data with network structures in a variety of fields including social networks [20, 30], communication networks [1, 2, 12], biological networks [3, 32, 48], disease transmission networks [33, 43] and so on. This creates an urgent need to understand the generative mechanism of these networks and to explore various characteristics of the network structures in a principled way. Statistical models are useful tools to this end, since they can capture the regularities of network processes and variability of network configurations of interests, and help to understand the uncertainty associated with observed outcomes [40, 42]. At the same time, data with network structures pose new challenges for statistical inference, in particular asymptotic analysis when only one realized network is observed and one is often interested in the asymptotic phenomena with the growing size of the network [14].

The in- and out-degrees of vertices (or degrees for undirected networks) preliminarily summarize the information contained in a network, and their distributions provide important insights for understanding the generative mechanism of networks. In the undirected case, the degree sequence has been extensively studied

[6, 10, 25, 34, 39, 55]. In particular, its distributions have been explored under the framework of the exponential family parameterized by the so-called "potentials" of vertices recently, for example, the "β-model" by [10] for binary edges or "maximum entropy models" by [25] for weighted edges in which the degree sequence is the exclusively sufficient statistic. It is also worth to note that the asymptotic theory of the maximum likelihood estimates (MLEs) for these models have not been derived until very recently [10, 25, 53, 54]. In the directed case, how to construct and sample directed graphs with given in- and out-degree (sometimes referred as "bi-degree") sequences have been studied [11, 13, 29]. However, statistical inference is not available, especially for asymptotic analysis. The distributions of the bi-degrees were studied in [41] through empirical examples for social networks, but the work lacked theoretical analysis.

In this paper, we study the distribution of the bi-degree sequence when it is the sufficient statistic in a directed graph. Recall the Koopman–Pitman–Darmois theorem or the principle of maximum entropy [49, 50], which states that the probability mass function of the bi-degree sequence must admit the form of the exponential family. We will characterize the exponential family distributions for the bi-degree sequence with three types of weighted edges (binary, discrete and continuous) and conduct the maximum likelihood inference.

In the model we study, one out-degree parameter and one in-degree parameter are needed for each vertex. As a result, the total number of parameters is twice of the number of the vertices. As the size of the network increases, the number of parameters goes to infinity. This makes asymptotic inference very challenging. Establishing the uniform consistency and asymptotic normality of the MLE are the aims of this paper. To the best of our knowledge, it is the first time that such results are derived in directed exponential random graph models with weighted edges. We remark further that our proofs are highly nontrivial. One key feature of our proofs lies in approximating the inverse of the Fisher information matrix by a simple matrix with small approximation errors. This approximation is utilized to derive a Newton iterative algorithm with geometrically fast rate of convergence, which leads to the proof of uniform consistency, and it is also utilized to derive approximately explicit expressions of the estimators, which leads to the proof of asymptotic normality. Furthermore, the approximate inverse makes the asymptotic variances of estimators explicit and concise. We note that [21, 22] have studied problems related to the present paper but the methods therein cannot be applied to the model we study. This is explained in detail at the end of the next section after we state the main theorems.

Next, we formally describe the models considered in this paper. Consider a directed graph $\mathcal{G}$ on $n \geq 2$ vertices labeled by $1, \ldots, n$. Let $a_{i,j} \in \Omega$ be the weight of the directed edge from i to j, where $\Omega \subseteq \mathbb{R}$ is the set of all possible weight values, and $A = (a_{i,j})$ be the adjacency matrix of $\mathcal{G}$. We consider three cases: $\Omega = \{0, 1\}$, $\Omega = [0, \infty)$ and $\Omega = \{0, 1, 2, \ldots\}$, where the first case is the usual binary edge. We assume that there are no self-loops, that is, $a_{i,i} = 0$. Let $d_i = \sum_{j \neq i} a_{i,j}$

be the out-degree of vertex i and $\mathbf{d} = (d_1, \ldots, d_n)^\top$ be the out-degree sequence of the graph $\mathcal{G}$. Similarly, define $b_j = \sum_{i \neq j} a_{i,j}$ as the in-degree of vertex j and $\mathbf{b} = (b_1, \ldots, b_n)^\top$ as the in-degree sequence. The pair $\{\mathbf{b}, \mathbf{d}\}$ or $\{(b_1, d_1), \ldots, (b_n, d_n)\}$ are the bi-degree sequence. Then the density or probability mass function on $\mathcal{G}$ parameterized by exponential family distributions with respect to some canonical measure ν is

$$(1.1) \qquad p(\mathcal{G}) = \exp(\boldsymbol{\alpha}^\top \mathbf{d} + \boldsymbol{\beta}^\top \mathbf{b} - Z(\boldsymbol{\alpha}, \boldsymbol{\beta})),$$

where $Z(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is the log-partition function, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\top$ is a parameter vector tied to the out-degree sequence, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^\top$ is a parameter vector tied to the in-degree sequence. This model can be viewed as a directed version of the β-model [10]. It can also be represented as the log-linear model [16–18] and the algorithm developed for the log-linear model can be used to compute the MLE. As explained by [26], α_i quantifies the effect of an outgoing edge from vertex i and β_j quantifies the effect of an incoming edge connecting to vertex j. If α_i is large and positive, vertex i will tend to have a relatively large out-degree. Similarly, if β_j is large and positive, vertex j tends to have a relatively large in-degree. Note that

$$(1.2)$$
$$\exp(\boldsymbol{\alpha}^\top \mathbf{d} + \boldsymbol{\beta}^\top \mathbf{b}) = \exp\left(\sum_{i,j=1; i \neq j}^{n} (\alpha_i + \beta_j) a_{i,j} \right)$$
$$= \prod_{i,j=1; i \neq j}^{n} \exp((\alpha_i + \beta_j) a_{i,j}),$$

which implies that the $n(n-1)$ random variables $a_{i,j}$, $i \neq j$ are mutually independent and $Z(\boldsymbol{\alpha}, \boldsymbol{\beta})$ can be expressed as

$$(1.3) \quad Z(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{i \neq j} Z_1(\alpha_i + \beta_j) := \sum_{i \neq j} \log \int_\Omega \exp((\alpha_i + \beta_j) a_{i,j}) \nu(da_{i,j}).$$

Since an out-edge from vertex i pointing to j is the in-edge of j coming from i, it is immediate that

$$\sum_{i=1}^{n} d_i = \sum_{j=1}^{n} b_j.$$

Moreover, since the sample is just one realization of the random graph, the density or probability mass function (1.1) is also the likelihood function. Note that if one transforms $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ to $(\boldsymbol{\alpha} - c, \boldsymbol{\beta} + c)$, the likelihood does not change. Therefore, for identifiability, constraints on $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ are necessary. In this paper, we choose to set $\beta_n = 0$. Other constraints are also possible, for example, $\sum_i \alpha_i = 0$ or $\sum_j \beta_j = 0$. In total, there are $2n - 1$ independent parameters and the natural parameter space becomes

$$\Theta = \{(\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_{n-1})^\top \in R^{2n-1} : Z(\boldsymbol{\alpha}, \boldsymbol{\beta}) < \infty\}.$$

Note that model (1.1) can serve as the null model for hypothesis testing, for example, [17, 26], or be used to reconstruct directed networks and make statistical inference in a situation in which only the bi-degree sequence is available due to privacy consideration [24]. Moreover, many complex directed network models reply on the bi-degree sequences, indirectly or directly. Thus, model (1.1) can be used for preliminary analysis of network data for choosing suitable statistics in describing network configurations, for example, [41].

It is worth to note that the above discussions only consider independent edges. Exponential random graph models (ERGMs), sometimes referred as exponential-family random graph models, for example, [27, 44], can be more general. If dependent network configurations such as k-stars and triangles are included as sufficient statistics, then edges are not independent and such models incur "near-degeneracy" in the sense of [23], in which almost all realized graphs essentially either contain no edges or are complete [9, 23, 44]. It has been shown in [9] that most realizations from many ERGMs look similar to the results of a simple Erdős–Rényi model, which implies that many distinct models have essentially the same MLE, and it was also proved and characterized in [9] the degeneracy observed in the ERGM with the counts of edges and triangles as the exclusively sufficient statistics. Further, by assuming a finite dimension of the parameter space, it was shown in [45] that the MLE is not consistent in the ERGM when the sufficient statistics involve k-stars, triangles and motifs of k-nodes ($k \geq 2$), while it is consistent when edges are dyadic independent. In view of the model degeneracy and problematic properties of estimators in the ERGM for dependent network configurations, we choose not to consider dependent edges in this paper.

For the remainder of the paper, we proceed as follows. In Section 2, we first introduce notation and key technical propositions that will be used in the proofs. We establish asymptotic results in the cases of binary weights, continuous weights and discrete weights in Sections 2.2, 2.3 and 2.4, respectively. Simulation studies are presented in Section 3. We further discuss the results in Section 4. Since the technical proofs in Sections 2.3 and 2.4 are similar to those in Section 2.2, we show the proofs for the theorems in Section 2.2 in the Appendix, while the proofs for Sections 2.3 and 2.4, as well as those for Proposition 1, Theorem 7 and Lemmas 2 and 3 in Section 2.2 are relegated to the Online Supplementary Material [51].

2. Main results.

2.1. *Notation and preparations.* Let $\mathbb{R}_+ = (0, \infty)$, $\mathbb{R}_0 = [0, \infty)$, $\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. For a subset $C \subset \mathbb{R}^n$, let C^0 and $\overline{C}$ denote the interior and closure of C, respectively. For a vector $\mathbf{x} = (x_1, \ldots, x_n)^\top \in R^n$, denote by $\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$, the ℓ_∞-norm of $\mathbf{x}$. For an $n \times n$ matrix $J = (J_{i,j})$, let $\|J\|_\infty$ denote the matrix norm induced by the ℓ_∞-norm on vectors in $\mathbb{R}^n$, that is,

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

In order to characterize the Fisher information matrix, we introduce a class of matrices. Given two positive numbers m and M with $M \geq m > 0$, we say the $(2n-1) \times (2n-1)$ matrix $V = (v_{i,j})$ belongs to the class $\mathcal{L}_n(m, M)$ if the following holds:

$$m \leq v_{i,i} - \sum_{j=n+1}^{2n-1} v_{i,j} \leq M, \qquad i = 1, \ldots, n-1;$$

$$v_{n,n} = \sum_{j=n+1}^{2n-1} v_{n,j},$$

$$v_{i,j} = 0, \qquad i, j = 1, \ldots, n, i \neq j,$$

$$(2.1) \quad v_{i,j} = 0, \qquad i, j = n+1, \ldots, 2n-1, i \neq j,$$

$$m \leq v_{i,j} = v_{j,i} \leq M, \qquad i = 1, \ldots, n, j = n+1, \ldots, 2n-1, j \neq n+i,$$

$$v_{i,n+i} = v_{n+i,i} = 0, \qquad i = 1, \ldots, n-1,$$

$$v_{i,i} = \sum_{k=1}^{n} v_{k,i} = \sum_{k=1}^{n} v_{i,k}, \qquad i = n+1, \ldots, 2n-1.$$

Clearly, if $V \in \mathcal{L}_n(m, M)$, then V is a $(2n-1) \times (2n-1)$ diagonally dominant, symmetric nonnegative matrix and V has the following structure:

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{12}^{\top} & V_{22} \end{pmatrix},$$

where V_{11} (n by n) and V_{22} ($n-1$ by $n-1$) are diagonal matrices, V_{12} is a non-negative matrix whose nondiagonal elements are positive and diagonal elements equal to zero.

Define $v_{2n,i} = v_{i,2n} := v_{i,i} - \sum_{j=1; j \neq i}^{2n-1} v_{i,j}$ for $i = 1, \ldots, 2n-1$ and $v_{2n,2n} = \sum_{i=1}^{2n-1} v_{2n,i}$. Then $m \leq v_{2n,i} \leq M$ for $i = 1, \ldots, n-1$, $v_{2n,i} = 0$ for $i = n, n+1, \ldots, 2n-1$ and $v_{2n,2n} = \sum_{i=1}^{n} v_{i,2n} = \sum_{i=1}^{n} v_{2n,i}$. We propose to approximate the inverse of V, V^{-1}, by the matrix $S = (s_{i,j})$, which is defined as

$$s_{i,j} = \begin{cases} \dfrac{\delta_{i,j}}{v_{i,i}} + \dfrac{1}{v_{2n,2n}}, & i, j = 1, \ldots, n, \\[2mm] -\dfrac{1}{v_{2n,2n}}, & i = 1, \ldots, n, j = n+1, \ldots, 2n-1, \\[2mm] -\dfrac{1}{v_{2n,2n}}, & i = n+1, \ldots, 2n-1, j = 1, \ldots, n, \\[2mm] \dfrac{\delta_{i,j}}{v_{i,i}} + \dfrac{1}{v_{2n,2n}}, & i, j = n+1, \ldots, 2n-1, \end{cases}$$

where $\delta_{i,j} = 1$ when $i = j$ and $\delta_{i,j} = 0$ when $i \neq j$. Note that S can be rewritten as

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^\top & S_{22} \end{pmatrix},$$

where $S_{11} = 1/v_{2n,2n} + \mathrm{diag}(1/v_{1,1}, 1/v_{2,2}, \ldots, 1/v_{n,n})$, S_{12} is an $n \times (n-1)$ matrix whose elements are all equal to $-1/v_{2n,2n}$, and $S_{22} = 1/v_{2n,2n} + \mathrm{diag}(1/v_{n+1,n+1}, 1/v_{n+2,n+2}, \ldots, 1/v_{2n-1,2n-1})$.

To quantify the accuracy of this approximation, we define another matrix norm $\| \cdot \|$ for a matrix $A = (a_{i,j})$ by $\|A\| := \max_{i,j} |a_{i,j}|$. Then we have the following proposition, whose proof is given in the Online Supplementary Material [51].

PROPOSITION 1. *If $V \in \mathcal{L}_n(m, M)$ with $M/m = o(n)$, then for large enough n,*

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

where c_1 is a constant that does not depend on M, m and n.

Note that if M and m are bounded constants, then the upper bound of the above approximation error is on the order of n^{-2}, indicating that S is a high-accuracy approximation to V^{-1}. Further, based on the above proposition, we immediately have the following lemma.

LEMMA 1. *If $V \in \mathcal{L}_n(m, M)$ with $M/m = o(n)$, then for a vector $\mathbf{x} \in R^{2n-1}$,*

$$\|V^{-1}\mathbf{x}\|_\infty \leq \|(V^{-1} - S)\mathbf{x}\|_\infty + \|S\mathbf{x}\|_\infty$$

$$\leq \frac{2c_1(2n-1)M^2 \|\mathbf{x}\|_\infty}{m^3(n-1)^2} + \frac{|x_{2n}|}{v_{2n,2n}} + \max_{i=1,\ldots,2n-1} \frac{|x_i|}{v_{i,i}},$$

where $x_{2n} := \sum_{i=1}^n x_i - \sum_{i=n+1}^{2n-1} x_i$.

Let $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_{n-1})^\top$ and $\mathbf{g} = (d_1, \ldots, d_n, b_1, \ldots, b_{n-1})^\top$. Henceforth, we will use V to denote the Fisher information matrix of the parameter vector $\boldsymbol{\theta}$ and show $V \in \mathcal{L}_n(m, M)$. In the next three subsections, we will analyze three specific choices of the weight set: $\Omega = \{0, 1\}$, $\Omega = \mathbb{R}_0$, $\Omega = \mathbb{N}_0$, respectively. For each case, we specify the distribution of the edge weights $a_{i,j}$, the natural parameter space Θ, the likelihood equations, and prove the existence, uniqueness, consistency and asymptotic normality of the MLE. We defer the proofs for the results in Section 2.2 to the Appendix and all other proofs for Sections 2.3 and 2.4 to the Online Supplementary Material [51].

2.2. *Binary weights.* In the case of binary weights, that is, $\Omega = \{0, 1\}$, ν is the counting measure, and $a_{i,j}$, $1 \le i \ne j \le n$ are mutually independent Bernoulli random variables with

$$P(a_{i,j} = 1) = \frac{e^{\alpha_i + \beta_j}}{1 + e^{\alpha_i + \beta_j}}.$$

The log-partition function $Z(\boldsymbol{\theta})$ is $\sum_{i \ne j} \log(1 + e^{\alpha_i + \beta_j})$ and the likelihood equations are

$$d_i = \sum_{k=1, k \ne i}^{n} \frac{e^{\hat{\alpha}_i + \hat{\beta}_k}}{1 + e^{\hat{\alpha}_i + \hat{\beta}_k}}, \qquad i = 1, \ldots, n,$$

(2.2)

$$b_j = \sum_{k=1, k \ne j}^{n} \frac{e^{\hat{\alpha}_k + \hat{\beta}_j}}{1 + e^{\hat{\alpha}_k + \hat{\beta}_j}}, \qquad j = 1, \ldots, n-1,$$

where $\hat{\boldsymbol{\theta}} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_n, \hat{\beta}_1, \ldots, \hat{\beta}_{n-1})^\top$ is the MLE of $\boldsymbol{\theta}$ and $\hat{\beta}_n = 0$. Note that in this case, the likelihood equations are identical to the moment equations.

We first establish the existence and consistency of $\hat{\boldsymbol{\theta}}$ by applying Theorem 7 in the Appendix. Define a system of functions:

$$F_i(\boldsymbol{\theta}) = d_i - \sum_{k=1; k \ne i}^{n} \frac{e^{\alpha_i + \beta_k}}{1 + e^{\alpha_i + \beta_k}}, \qquad i = 1, \ldots, n,$$

$$F_{n+j}(\boldsymbol{\theta}) = b_j - \sum_{k=1; k \ne j}^{n} \frac{e^{\alpha_k + \beta_j}}{1 + e^{\alpha_k + \beta_j}}, \qquad j = 1, \ldots, n,$$

$$F(\boldsymbol{\theta}) = \big(F_1(\boldsymbol{\theta}), \ldots, F_{2n-1}(\boldsymbol{\theta})\big)^\top.$$

Note the solution to the equation $F(\boldsymbol{\theta}) = 0$ is precisely the MLE. Then the Jacobian matrix $F'(\boldsymbol{\theta})$ of $F(\boldsymbol{\theta})$ can be calculated as follows. For $i = 1, \ldots, n$,

$$\frac{\partial F_i}{\partial \alpha_l} = 0, \qquad l = 1, \ldots, n, l \ne i; \qquad \frac{\partial F_i}{\partial \alpha_i} = -\sum_{k=1; k \ne i}^{n} \frac{e^{\alpha_i + \beta_k}}{(1 + e^{\alpha_i + \beta_k})^2},$$

$$\frac{\partial F_i}{\partial \beta_j} = -\frac{e^{\alpha_i + \beta_j}}{(1 + e^{\alpha_i + \beta_j})^2}, \qquad j = 1, \ldots, n-1, j \ne i; \qquad \frac{\partial F_i}{\partial \beta_i} = 0$$

and for $j = 1, \ldots, n-1$,

$$\frac{\partial F_{n+j}}{\partial \alpha_l} = -\frac{e^{\alpha_l + \beta_j}}{(1 + e^{\alpha_l + \beta_j})^2}, \qquad l = 1, \ldots, n, l \ne j; \qquad \frac{\partial F_{n+j}}{\partial \alpha_j} = 0,$$

$$\frac{\partial F_{n+j}}{\partial \beta_j} = -\sum_{k=1; k \ne j}^{n} \frac{e^{\alpha_k + \beta_j}}{(1 + e^{\alpha_k + \beta_j})^2}, \qquad \frac{\partial F_{n+j}}{\partial \beta_l} = 0, \qquad l = 1, \ldots, n-1.$$

First, note that since the Jacobian is diagonally dominant with nonzero diagonals, it is positive definite, implying that the likelihood function has a unique optimum.

Second, it is not difficult to verify that $-F'(\boldsymbol{\theta}) \in \mathcal{L}_n(m, M)$, thus Proposition 1 and Theorem 7 can be applied. Let $\boldsymbol{\theta}^*$ denote the true parameter vector. The constants K_1, K_2 and r in the upper bounds of Theorem 7 are given in the following lemma, whose proof is given in the Online Supplementary Material [51].

LEMMA 2.   *Take $D = R^{2n-1}$ and $\boldsymbol{\theta}^{(0)} = \boldsymbol{\theta}^*$ in Theorem 7. Assume*

$$\max\left\{ \max_{i=1,\ldots,n} |d_i - \mathbb{E}(d_i)|, \max_{j=1,\ldots,n} |b_j - \mathbb{E}(b_j)| \right\}$$

(2.3)

$$\leq \sqrt{(n-1)\log(n-1)}.$$

Then we can choose the constants K_1, K_2 and r in Theorem 7 as

$$K_1 = n - 1, \qquad K_2 = \frac{n-1}{2}, \qquad r \leq \frac{(\log n)^{1/2}}{n^{1/2}} \left(c_{11} e^{6\|\boldsymbol{\theta}^*\|_\infty} + c_{12} e^{2\|\boldsymbol{\theta}^*\|_\infty} \right),$$

where c_{11} and c_{12} are constants.

The following lemma assures that condition (2.3) holds with a large probability, whose proof is again given in the Online Supplementary Material [51].

LEMMA 3.   *With probability at least $1 - 4n/(n-1)^2$, we have*

$$\max\left\{ \max_i |d_i - \mathbb{E}(d_i)|, \max_j |b_j - \mathbb{E}(b_j)| \right\} \leq \sqrt{(n-1)\log(n-1)}.$$

Combining the above two lemmas, we have the result of consistency.

THEOREM 1.   *Assume that $\boldsymbol{\theta}^* \in \mathbb{R}^{2n-1}$ with $\|\boldsymbol{\theta}^*\|_\infty \leq \tau \log n$, where $0 < \tau < 1/24$ is a constant, and that $A \sim \mathbb{P}_{\boldsymbol{\theta}^*}$, where $\mathbb{P}_{\boldsymbol{\theta}^*}$ denotes the probability distribution (1.1) on A under the parameter $\boldsymbol{\theta}^*$. Then as n goes to infinity, with probability approaching one, the MLE $\hat{\boldsymbol{\theta}}$ exists and satisfies*

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty = O_p\left(\frac{(\log n)^{1/2} e^{8\|\boldsymbol{\theta}^*\|_\infty}}{n^{1/2}} \right) = o_p(1).$$

Further, if the MLE exists, it is unique.

Next, we establish asymptotic normality of $\hat{\boldsymbol{\theta}}$ and outline the main ideas in the following. Let $\ell(\boldsymbol{\theta}; A) = \sum_{i=1}^n \alpha_i d_i + \sum_{j=1}^{n-1} \beta_j b_j - \sum_{i \neq j} \log(1 + e^{\alpha_i + \beta_j})$ denote the log-likelihood function of the parameter vector $\boldsymbol{\theta}$ given the sample A. Note that $F'(\boldsymbol{\theta}) = \partial^2 \ell / \partial \boldsymbol{\theta}^2$, and $V = -F'(\boldsymbol{\theta})$ is the Fisher information matrix of the parameter vector $\boldsymbol{\theta}$. Clearly, $\hat{\boldsymbol{\theta}}$ does not have an explicit expression according to the system of likelihood equations (2.2). However, if $\hat{\boldsymbol{\theta}}$ can be approximately represented as a function of $\mathbf{g} = (d_1, \ldots, d_n, b_1, \ldots, b_{n-1})^\top$ with an explicit expression, then the central limit theorem for $\hat{\boldsymbol{\theta}}$ immediately follows by noting that under certain

regularity conditions

$$\frac{g_i - \mathbb{E}(g_i)}{v_{i,i}^{1/2}} \to N(0, 1), \qquad n \to \infty,$$

where g_i denotes the ith element of $\mathbf{g}$. The identity between the likelihood equations and the moment equations provides such a possibility. Specifically, if we apply Taylor's expansion to each component of $\mathbf{g} - \mathbb{E}(\mathbf{g})$, the second-order term in the expansion is $V(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, which implies that obtaining an expression of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ crucially depends on the inverse of V. Note that $V = -F'(\boldsymbol{\theta}) \in \mathcal{L}_n(m, M)$ according to the previous calculation. Although V^{-1} does not have a closed form, we can use S to approximate it and Proposition 1 establishes an upper bound on the error of this approximation, which is on the order of n^{-2} if M and m are bounded constants.

Regarding the asymptotic normality of $g_i - \mathbb{E}(g_i)$, we note that both $d_i = \sum_{k \neq i} a_{i,k}$ and $b_j = \sum_{k \neq j} a_{k,j}$ are sums of $n - 1$ independent Bernoulli random variables. By the central limit theorem for the bounded case in [31], page 289, we know that $v_{i,i}^{-1/2}(d_i - \mathbb{E}(d_i))$ and $v_{n+j,n+j}^{-1/2}(b_j - \mathbb{E}(b_j))$ are asymptotically standard normal if $v_{i,i}$ diverges. Since $e^x/(1 + e^x)^2$ is an increasing function on x when $x \geq 0$ and a decreasing function when $x \leq 0$, we have

$$\frac{(n-1)e^{2\|\boldsymbol{\theta}^*\|_\infty}}{(1 + e^{2\|\boldsymbol{\theta}^*\|_\infty})^2} \leq v_{i,i} \leq \frac{n-1}{4}, \qquad i = 1, \ldots, 2n.$$

In all, we have the following proposition.

PROPOSITION 2. *Assume that $A \sim \mathbb{P}_{\boldsymbol{\theta}^*}$. If $e^{\|\boldsymbol{\theta}^*\|_\infty} = o(n^{1/2})$, then for any fixed $k \geq 1$, as $n \to \infty$, the vector consisting of the first k elements of $S\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}$ is asymptotically multivariate normal with mean zero and covariance matrix given by the upper left $k \times k$ block of S.*

The central limit theorem is stated in the following and proved by establishing a relationship between $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ and $S\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}$ (see details in the Appendix and the Online Supplementary Material [51]).

THEOREM 2. *Assume that $A \sim \mathbb{P}_{\boldsymbol{\theta}^*}$. If $\|\boldsymbol{\theta}^*\|_\infty \leq \tau \log n$, where $\tau \in (0, 1/44)$ is a constant, then for any fixed $k \geq 1$, as $n \to \infty$, the vector consisting of the first k elements of $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ is asymptotically multivariate normal with mean $\mathbf{0}$ and covariance matrix given by the upper left $k \times k$ block of S.*

REMARK 1. By Theorem 2, for any fixed i, as $n \to \infty$, the convergence rate of $\hat{\theta}_i$ is $1/v_{i,i}^{1/2}$. Since $(n-1)e^{-2\|\boldsymbol{\theta}^*\|_\infty}/4 \leq v_{i,i} \leq (n-1)/4$, the rate of convergence is between $O(n^{-1/2}e^{\|\boldsymbol{\theta}^*\|_\infty})$ and $O(n^{-1/2})$.

In this subsection, we have presented the main ideas to prove the consistency and asymptotic normality of the MLE for the case of binary weights. In the next

two subsections, we apply similar ideas to the cases of continuous and discrete weights, respectively.

2.3. *Continuous weights.* Another important case of model (1.1) is when the weight of the edge is continuous. For example, in communication networks, if an edge denotes the talking time between two people in a telephone network, then its weight is continuous. In the case of continuous weights, that is, $\Omega = [0, \infty)$, ν is the Borel measure and $a_{i,j}$, $1 \le i \ne j \le n$ are mutually independent exponential random variables with the density

$$f_{\boldsymbol{\theta}}(a) = \frac{1}{-(\alpha_i + \beta_j)} e^{(\alpha_i + \beta_j)a}, \qquad \alpha_i + \beta_j < 0,$$

and the natural parameter space is

$$\Theta = \{\boldsymbol{\theta} : \alpha_i + \beta_j < 0\}.$$

To follow the tradition that the rate parameters are positive in exponential families, we take the transformation $\bar{\boldsymbol{\theta}} = -\boldsymbol{\theta}$, $\bar{\alpha}_i = -\alpha_i$ and $\bar{\beta}_j = -\beta_j$. The corresponding natural parameter space then becomes

$$\overline{\Theta} = \{\bar{\boldsymbol{\theta}} : \bar{\alpha}_i + \bar{\beta}_j > 0\}.$$

Here, we denote by $\hat{\boldsymbol{\theta}}$ the MLE of $\bar{\boldsymbol{\theta}}$. The log-partition $Z(\bar{\boldsymbol{\theta}})$ is $\sum_{i \ne j} \log(\bar{\alpha}_i + \bar{\beta}_j)$ and the likelihood equations are

$$
\begin{aligned}
d_i &= \sum_{k=1; k \ne i}^{n} (\hat{\alpha}_i + \hat{\beta}_k)^{-1}, \qquad i = 1, \ldots, n, \\
b_j &= \sum_{k=1; k \ne j}^{n} (\hat{\alpha}_k + \hat{\beta}_j)^{-1}, \qquad j = 1, \ldots, n.
\end{aligned}
$$

(2.4)

Similar to Section 2.2, we define a system of functions:

$$F_i(\bar{\boldsymbol{\theta}}) = d_i - \sum_{k \ne i} (\bar{\alpha}_i + \bar{\beta}_k)^{-1}, \qquad i = 1, \ldots, n,$$

$$F_{n+j}(\bar{\boldsymbol{\theta}}) = b_j - \sum_{k \ne j} (\bar{\alpha}_k + \bar{\beta}_j)^{-1}, \qquad j = 1, \ldots, n-1,$$

$$F(\bar{\boldsymbol{\theta}}) = \left(F_1(\bar{\boldsymbol{\theta}}), \ldots, F_{2n-1}(\bar{\boldsymbol{\theta}}) \right)^{\top}.$$

The solution to the equation $F(\bar{\boldsymbol{\theta}}) = 0$ is the MLE, and the Jacobian matrix $F'(\bar{\boldsymbol{\theta}})$ of $F(\bar{\boldsymbol{\theta}})$ can be calculated as follows. For $i = 1, \ldots, n$,

$$\frac{\partial F_i}{\partial \bar{\alpha}_l} = 0, \qquad l = 1, \ldots, n, l \ne i; \qquad \frac{\partial F_i}{\partial \bar{\alpha}_i} = \sum_{k \ne i} \frac{1}{(\bar{\alpha}_i + \bar{\beta}_k)^2},$$

$$\frac{\partial F_i}{\partial \bar{\beta}_j} = \frac{1}{(\bar{\alpha}_i + \bar{\beta}_j)^2}, \qquad j = 1, \ldots, n-1, j \ne i; \qquad \frac{\partial F_i}{\partial \bar{\beta}_i} = 0,$$

and for $j = 1, \ldots, n-1$,

$$\frac{\partial F_{n+j}}{\partial \bar{\alpha}_l} = \frac{1}{(\bar{\alpha}_l + \bar{\beta}_j)^2}, \qquad l = 1, \ldots, n, l \neq j; \qquad \frac{\partial F_{n+j}}{\partial \bar{\alpha}_j} = 0,$$

$$\frac{\partial F_{n+j}}{\partial \bar{\beta}_j} = \sum_{k \neq j} \frac{1}{(\bar{\alpha}_j + \bar{\beta}_j)^2}; \qquad \frac{\partial F_{n+j}}{\partial \bar{\beta}_l} = 0, \qquad l = 1, \ldots, n-1, l \neq j.$$

It is not difficult to see that $F'(\bar{\boldsymbol{\theta}}^*) \in \mathcal{L}_n(m, M)$ such that Proposition 1 can be applied, and the constants in the upper bounds of Theorem 7 are given in the following lemma.

LEMMA 4. *Assume that $\bar{\boldsymbol{\theta}}^*$ satisfies $q_n \leq \bar{\alpha}_i^* + \bar{\beta}_j^* \leq Q_n$ for any $1 \leq i \neq j \leq n$ and*

$$(2.5) \quad \max\left\{ \max_{i=1,\ldots,n} |d_i - \mathbb{E}(d_i)|, \max_{j=1,\ldots,n} |b_j - \mathbb{E}(b_j)| \right\} \leq \sqrt{\frac{8(n-1)\log n}{\gamma q_n^2}},$$

where γ is an absolute constant. Then we have

$$r = \|[F'(\bar{\boldsymbol{\theta}}^*)]^{-1} F(\bar{\boldsymbol{\theta}}^*)\|_\infty \leq \left(\frac{2c_1 Q_n^6}{n q_n^4} + \frac{1}{(n-1)q_n^2} \right) \sqrt{\frac{8(n-1)\log n}{\gamma q_n^2}}.$$

Further, take $\bar{\boldsymbol{\theta}}^{(0)} = \bar{\boldsymbol{\theta}}^*$ and $D = \Omega(\bar{\boldsymbol{\theta}}^*, 2r)$ in Theorem 7, that is, an open ball $\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}^*\|_\infty < 2r\}$. If $q_n - 4r > 0$, then we can choose $K_1 = 2(n-1)/(q_n - 4r)^3$ and $K_2 = (n-1)/(q_n - 4r)^3$.

The following lemma assures condition (2.5) holds with a large probability.

LEMMA 5. *With probability at least $1 - 4/n$, we have*

$$\max\left\{ \max_i |d_i - \mathbb{E}(d_i)|, \max_j |b_j - \mathbb{E}(b_j)| \right\} \leq \sqrt{\frac{8(n-1)\log n}{\gamma q_n^2}}.$$

Combining the above two lemmas, we have the result of consistency.

THEOREM 3. *Assume that $\bar{\boldsymbol{\theta}}^*$ satisfies $q_n \leq \bar{\alpha}_i^* + \bar{\beta}_j^* \leq Q_n$ and $A \sim P_{\bar{\boldsymbol{\theta}}^*}$. If $Q_n/q_n = o\{(n/\log n)^{1/18}\}$, then as n goes to infinity, with probability approaching one, the MLE $\hat{\boldsymbol{\theta}}$ exists and satisfies*

$$\|\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}^*\|_\infty = O_p\left(\frac{Q_n^9 (\log n)^{1/2}}{n^{1/2} q_n^9} \right) = o_p(1).$$

Further, if the MLE exists, it is unique.

Again, note that both $d_i = \sum_{k \neq i} a_{i,k}$ and $b_j = \sum_{k \neq j} a_{k,j}$ are sums of $n-1$ independent exponential random variables, and $V = F'(\bar{\boldsymbol{\theta}}^*) \in \mathcal{L}_n(m, M)$ is the Fisher information matrix of $\bar{\boldsymbol{\theta}}$. It is not difficult to show that the third moment of the exponential random variable with rate parameter λ is $6\lambda^{-3}$. Under the assumption of $0 < q_n \leq \bar{\alpha}_i^* + \bar{\beta}_j^* \leq Q_n$, we have

$$\frac{\sum_{j=1; j \neq i}^{n} \mathbb{E}(a_{i,j}^3)}{v_{i,i}^{3/2}} = \frac{6\sum_{j=1; j \neq i}^{n} (\bar{\alpha}_i^* + \bar{\beta}_j^*)^{-1}}{v_{i,i}^{1/2}} \leq \frac{6Q_n/q_n}{(n-1)^{1/2}} \qquad \text{for } i = 1, \ldots, n$$

and

$$\frac{\sum_{i=1; i \neq j}^{n} \mathbb{E}(a_{i,j}^3)}{v_{n+j,n+j}^{3/2}} = \frac{6\sum_{i=1; i \neq j}^{n} (\bar{\alpha}_i^* + \bar{\beta}_j^*)^{-1}}{v_{n+j,n+j}^{1/2}} \leq \frac{6Q_n/q_n}{(n-1)^{1/2}} \qquad \text{for } j = 1, \ldots, n.$$

Note that if $Q_n/q_n = o(n^{1/2})$, the above expression goes to zero. This implies that the condition for the Lyapunov's central limit theorem holds. Therefore, $v_{i,i}^{-1/2}(d_i - \mathbb{E}(d_i))$ is asymptotically standard normal if $Q_n/q_n = o(n^{1/2})$. Similarly, $v_{n+j,n+j}^{-1/2}(b_j - \mathbb{E}(b_j))$ is also asymptotically standard normal under the same condition. Noting that $[S(\mathbf{g} - \mathbb{E}(\mathbf{g}))]_i = v_{i,i}^{-1}(g_i - \mathbb{E}(g_i)) + v_{2n,2n}^{-1}(b_n - \mathbb{E}(b_n))$, we have the following proposition.

PROPOSITION 3.   *If $Q_n/q_n = o(n^{1/2})$, then for any fixed $k \geq 1$, as $n \to \infty$, the vector consisting of the first k elements of $S(\mathbf{g} - \mathbb{E}(\mathbf{g}))$ is asymptotically multivariate normal with mean zero and covariance matrix given by the upper $k \times k$ block of the matrix S.*

By establishing a relationship between $\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}^*$ and $S\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}$, we have the central limit theorem for the MLE $\hat{\boldsymbol{\theta}}$.

THEOREM 4.   *If $Q_n/q_n = o(n^{1/50}/(\log n)^{1/25})$, then for any fixed $k \geq 1$, as $n \to \infty$, the vector consisting of the first k elements of $\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}^*$ is asymptotically multivariate normal with mean zero and covariance matrix given by the upper $k \times k$ block of the matrix S.*

REMARK 2.   By Theorem 4, for any fixed i, as $n \to \infty$, the convergence rate of $\hat{\theta}_i$ is $1/v_{i,i}^{1/2}$. Since $(n-1)/Q_n^2 \leq v_{i,i} \leq (n-1)/q_n^2$, the rate of convergence is between $O(n^{-1/2}Q_n)$ and $O(n^{-1/2}q_n)$.

2.4. *Discrete weights.*   In the case of discrete weights, that is, $\Omega = \mathbb{N}_0$, ν is the counting measure and $a_{i,j}$, $1 \leq i \neq j \leq n$ are mutually independent geometric random variables with the probability mass function

$$P(a_{i,j} = a) = (1 - e^{(\alpha_i + \beta_j)})e^{(\alpha_i + \beta_j)a}, \qquad a = 0, 1, 2, \ldots,$$

where $\alpha_i + \beta_j < 0$. The natural parameter space is $\Theta = \{\boldsymbol{\theta} : \alpha_i + \beta_j < 0\}$. Again, we take the transformation $\bar{\boldsymbol{\theta}} = -\boldsymbol{\theta}$, $\bar{\alpha}_i = -\alpha_i$ and $\bar{\beta}_j = -\beta_j$, and the corresponding natural parameter space becomes

$$\overline{\Theta} = \{\bar{\boldsymbol{\theta}} : \bar{\alpha}_i + \bar{\beta}_j > 0\}.$$

The log-partition $Z(\bar{\boldsymbol{\theta}})$ is $\sum_{i \neq j} \log(1 - e^{-(\bar{\alpha}_i + \bar{\beta}_j)})$ and the likelihood equations are

$$(2.6) \quad d_i = \sum_{k \neq i} \frac{e^{-(\hat{\alpha}_i + \hat{\beta}_k)}}{1 - e^{-(\hat{\alpha}_i + \hat{\beta}_k)}} = \sum_{k \neq i} \frac{1}{e^{(\hat{\alpha}_i + \hat{\beta}_k)} - 1}, \qquad i = 1, \ldots, n,$$

$$(2.7) \quad b_j = \sum_{k \neq j} \frac{e^{-(\hat{\alpha}_k + \hat{\beta}_j)}}{1 - e^{-(\hat{\alpha}_k + \hat{\beta}_j)}} = \sum_{k \neq j} \frac{1}{e^{(\hat{\alpha}_k + \hat{\beta}_j)} - 1}, \qquad j = 1, \ldots, n-1.$$

We first establish the existence and consistency of $\hat{\boldsymbol{\theta}}$ by applying Theorem 7. Define a system of functions:

$$F_i(\bar{\boldsymbol{\theta}}) = d_i - \sum_{k \neq i} \frac{1}{e^{(\bar{\alpha}_i + \bar{\beta}_k)} - 1}, \qquad i = 1, \ldots, n,$$

$$F_{n+j}(\bar{\boldsymbol{\theta}}) = b_j - \sum_{k \neq j} \frac{1}{e^{(\bar{\alpha}_k + \bar{\beta}_j)} - 1}, \qquad j = 1, \ldots, n,$$

$$F(\bar{\boldsymbol{\theta}}) = \left(F_1(\bar{\boldsymbol{\theta}}), \ldots, F_{2n-1}(\bar{\boldsymbol{\theta}})\right)^{\top}.$$

The solution to the equation $F(\bar{\boldsymbol{\theta}}) = 0$ is the MLE, and the Jacobian matrix $F'(\bar{\boldsymbol{\theta}})$ of $F(\bar{\boldsymbol{\theta}})$ can be calculated as follows: for $i = 1, \ldots, n$,

$$\frac{\partial F_i}{\partial \bar{\alpha}_l} = 0, \qquad l = 1, \ldots, n, l \neq i; \qquad \frac{\partial F_i}{\partial \bar{\alpha}_i} = \sum_{k=1; k \neq i}^{n} \frac{e^{(\bar{\alpha}_i + \bar{\beta}_k)} - 1}{(e^{(\bar{\alpha}_i + \bar{\beta}_k)} - 1)^2},$$

$$\frac{\partial F_i}{\partial \bar{\beta}_j} = \frac{e^{(\bar{\alpha}_i + \bar{\beta}_j)} - 1}{(e^{(\bar{\alpha}_i + \bar{\beta}_j)} - 1)^2}, \qquad j = 1, \ldots, n-1, j \neq i; \qquad \frac{\partial F_i}{\partial \bar{\beta}_i} = 0,$$

and for $j = 1, \ldots, n-1$,

$$\frac{\partial F_{n+j}}{\partial \bar{\alpha}_l} = \frac{e^{(\bar{\alpha}_l + \bar{\beta}_j)} - 1}{[e^{(\bar{\alpha}_l + \bar{\beta}_j)} - 1]^2}, \qquad l = 1, \ldots, n, l \neq j; \qquad \frac{\partial F_{n+j}}{\partial \bar{\alpha}_j} = 0,$$

$$\frac{\partial F_{n+j}}{\partial \bar{\beta}_j} = \sum_{k \neq j} \frac{e^{(\bar{\alpha}_k + \bar{\beta}_j)} - 1}{[e^{(\bar{\alpha}_k + \bar{\beta}_j)} - 1]^2}; \qquad \frac{\partial F_{n+j}}{\partial \bar{\beta}_l} = 0, \qquad l = 1, \ldots, n-1, l \neq j.$$

Let $\bar{\boldsymbol{\theta}}^*$ be the true parameter vector. It is not difficult to see $F'(\bar{\boldsymbol{\theta}}^*) \in \mathcal{L}_n(m, M)$ so that Proposition 1 can be applied. The constants in the upper bounds of Theorem 7 are given in the following lemma.

LEMMA 6.   *Assume that $\bar{\boldsymbol{\theta}}^*$ satisfies $q_n \le \bar{\alpha}_i^* + \bar{\beta}_j^* \le Q_n$ for all $i \ne j$, $A \sim \mathbb{P}_{\bar{\boldsymbol{\theta}}^*}$ and*

$$(2.8)\quad \max\left\{\max_{i=1,\dots,n}|d_i - \mathbb{E}(d_i)|,\ \max_{j=1,\dots,n}|b_j - \mathbb{E}(b_j)|\right\} \le \sqrt{\frac{8(n-1)\log n}{\gamma q_n^2}},$$

where γ is an absolute constant. Then we have

$$r = \|[F'(\bar{\boldsymbol{\theta}}^*)]^{-1} F(\bar{\boldsymbol{\theta}}^*)\|_\infty \le O\left(q_n^{-1}(e^{3Q_n}(1+q_n^{-4}) + e^{Q_n})\sqrt{\frac{\log n}{n}}\right).$$

Further, take $\bar{\boldsymbol{\theta}}^{(0)} = \bar{\boldsymbol{\theta}}^*$ and $D = \Omega(\bar{\boldsymbol{\theta}}^*, 2r)$ in Theorem 7, that is, an open ball $\{\boldsymbol{\theta}: \|\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}^*\|_\infty < 2r\}$. If $q_n - 4r > 0$, then we can choose $K_1 = 2(n-1)e^{q_n-4r}(1 + e^{q_n-4r})(e^{q_n-4r} - 1)^{-2}$ and $K_2 = (n-1)e^{q_n-4r}(1 + e^{q_n-4r})(e^{q_n-4r} - 1)^{-2}$.

The following lemma assures that the condition in the above lemma holds with a large probability.

LEMMA 7.   *With probability at least $1 - 4n/(n-1)^2$, we have*

$$\max\left\{\max_i|d_i - \mathbb{E}(d_i)|,\ \max_j|b_j - \mathbb{E}(b_j)|\right\} \le \sqrt{\frac{8(n-1)\log n}{\gamma q_n^2}}.$$

Combining the above two lemmas, we have the result of consistency.

THEOREM 5.   *Assume that $\bar{\boldsymbol{\theta}}^*$ satisfies $q_n \le \bar{\alpha}_i^* + \bar{\beta}_j^* \le Q_n$ for all $i \ne j$ and $A \sim P_{\bar{\boldsymbol{\theta}}^*}$. If $(1 + q_n^{-11})e^{6Q_n} = o(n^{1/2}/(\log n)^{1/2})$ then as n goes to infinity, with probability approaching one, the MLE $\hat{\boldsymbol{\theta}}$ exists and satisfies*

$$\|\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}^*\|_\infty = O_p\left(e^{3Q_n}\left(1 + \frac{1}{q_n^5}\right)\sqrt{\frac{\log n}{n}}\right) = o_p(1).$$

Further, if the MLE exists, it is unique.

Note that both $d_i = \sum_{j \ne i} a_{i,j}$ and $b_j = \sum_{i \ne j} a_{i,j}$ are sums of $n-1$ independent geometric random variables. Also note that $q_n \le \bar{\alpha}_i^* + \bar{\beta}_j^* \le Q_n$ and $V = F'(\bar{\boldsymbol{\theta}}^*) \in \mathcal{L}_n(m, M)$, thus we have

$$\frac{e^{Q_n}}{(e^{Q_n} - 1)^2} \le v_{i,j} \le \frac{e^{q_n}}{(e^{q_n} - 1)^2}, \qquad i = 1,\dots,n,\ j = n+1,\dots,2n,\ j \ne n+i,$$

$$\frac{(n-1)e^{Q_n}}{(e^{Q_n} - 1)^2} \le v_{i,i} \le \frac{(n-1)e^{q_n}}{(e^{q_n} - 1)^2}, \qquad i = 1,\dots,2n.$$

Using the moment-generating function of the geometric distribution, it is not difficult to verify that

$$\mathbb{E}(a_{i,j}^3) = \frac{1 - p_{i,j}}{p_{i,j}} + \frac{6(1 - p_{i,j})}{p_{i,j}^2} + \frac{6(1 - p_{i,j})^2}{p_{i,j}^3},$$

where $p_{i,j} = 1 - e^{-(\bar{\alpha}_i^* + \bar{\beta}_j^*)}$. By simple calculations, we also have

$$\mathbb{E}(a_{i,j}^3) = v_{i,j}\left(6 + \frac{e^{\bar{\alpha}_i^* + \bar{\beta}_j^*} - 1}{e^{\bar{\alpha}_i^* + \bar{\beta}_j^*}} + \frac{6}{e^{\bar{\alpha}_i^* + \bar{\beta}_j^*} - 1}\right).$$

It then follows

$$\frac{\sum_{j \neq i} \mathbb{E}(a_{i,j}^3)}{v_{i,i}^{3/2}} \leq \frac{7 + 6(e^{q_n} - 1)^{-1}}{v_{i,i}^{1/2}} \leq \frac{[7 + 6(e^{q_n} - 1)^{-1}](e^{Q_n} - 1)}{n^{1/2} e^{Q_n/2}}.$$

Note that if $e^{Q_n/2}/q_n = o(n^{1/2})$, the above expression goes to zero, which implies that the condition for the Lyapunov's central limit theorem holds. Therefore, for $i = 1, \ldots, n$, $v_{i,i}^{-1/2}(d_i - \mathbb{E}(d_i))$ is asymptotically standard normal if $e^{Q_n/2}/q_n = o(n^{1/2})$. Similarly, for $i = 1, \ldots, n$, $v_{n+i,n+i}^{-1/2}(b_i - \mathbb{E}(b_i))$ is also asymptotically standard normal if $e^{Q_n/2}/q_n = o(n^{1/2})$. Therefore, we have the following proposition.

PROPOSITION 4. *If $e^{Q_n/2}/q_n = o(n^{1/2})$, then for any fixed $k \geq 1$, as $n \to \infty$, the vector consisting of the first k elements of $S\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}$ is asymptotically multivariate normal with mean zero and covariance matrix given by the upper $k \times k$ block of the matrix S.*

The central limit theorem for the MLE $\hat{\boldsymbol{\theta}}$ is stated as follows.

THEOREM 6. *If $e^{9Q_n}(1 + q_n^{-15}) = o\{n^{1/2}/\log n\}$, then for any fixed $k \geq 1$, as $n \to \infty$, the vector consisting of the first k elements of $\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}^*$ is asymptotically multivariate normal with mean zero and covariance matrix given by the upper $k \times k$ block of the matrix S.*

REMARK 3. By Theorem 6, for any fixed i, as $n \to \infty$, the convergence rate of $\hat{\theta}_i$ is $1/v_{i,i}^{1/2}$. Since $(n - 1)e^{Q_n}(e^{Q_n} - 1)^{-2} \leq v_{i,i} \leq (n - 1)e^{q_n}(e^{q_n} - 1)^{-2}$, the rate of convergence is between $O(n^{-1/2}e^{Q_n/2})$ and $O(n^{-1/2}e^{q_n/2})$.

Comparison to [21, 22]. It is worth to note that [21] proved uniform consistency and asymptotic normality of the MLE in the Rasch model for item response theory under the assumption that all unknown parameters are bounded by a constant. Further, Haberman ([22], page 60) wrote that "Since Holland and Leinhardt's p_1 model is an example of an exponential response model..." and "The situation in the Holland–Leinhardt model is very similar, for their model under $\rho = 0$ is

mathematically equivalent to the incomplete Rasch model with $g = h$ and X_{ii} un-observed." Consequently, it was claimed that the method in [21] can be extended to derive the consistency and asymptotic normality of the MLE of the p_1 model without reciprocity, but a formal proof was not given. However, these conclusions seem premature due to the following reasons. First, in an item response experiment, a total of g people give answers (0 or 1) to a total of h items. The outcomes of the experiment naturally form a bipartite undirected graph, for example, [7], while model (1.1) is directed. Second, each vertex in the Rasch model is only assigned one parameter measuring either the out-degree effect for people or the in-degree effect for items, while there are two parameters in model (1.1), one for the in-degree and the other for the out-degree, for each vertex simultaneously. Therefore, model (1.1) cannot be simply viewed as an equivalent Rasch model. We also note that [19] pointed out that the Rasch model can be considered as the Bradley–Terry model [8] for incomplete paired comparisons, for which [46] proved uniform consistency and asymptotic normality for the MLE with a diverging number of parameters. Third, in contrast to the proofs in [21], our methods utilize an approximate inverse of the Fisher information matrix, requiring no upper bound on the parameters, while the methods in [21] were based on the classical exponential family theory of [4, 5]. Therefore, we conjecture that the methods in [21] cannot be extended to study the model in (1.1).

3. Simulation studies.   In this section, we evaluate the asymptotic results for model (1.1) through numerical simulations. The settings of parameter values take a linear form. Specifically, for the case with binary weights, we set $\alpha^*_{i+1} = (n - 1 - i)L/(n - 1)$ for $i = 0, \ldots, n - 1$; for the case with discrete weights, we set $\bar{\alpha}^*_{i+1} = 0.2 + (n - 1 - i)L/(n - 1)$ for $i = 0, \ldots, n - 1$. In both cases, we considered four different values for L, $L = 0$, $\log(\log n)$, $(\log n)^{1/2}$ and $\log n$, respectively. For the case with continuous weights, we set $\bar{\alpha}^*_{i+1} = 1 + (n - 1 - i)L/(n - 1)$ for $i = 0, \ldots, n - 1$ and also four values of L are considered: $L = 0$, $\log(\log(n))$, $\log(n)$ and $n^{1/2}$. For the parameter values of $\bar{\boldsymbol{\beta}}$, let $\bar{\beta}^*_i = \bar{\alpha}^*_i$, $i = 1, \ldots, n - 1$ for simplicity and $\bar{\beta}^*_n = 0$ by default.

Note that by Theorems 2, 4 and 6, $\hat{\xi}_{i,j} = [\hat{\alpha}_i - \hat{\alpha}_j - (\bar{\alpha}^*_i - \bar{\alpha}^*_j)]/(1/\hat{v}_{i,i} + 1/\hat{v}_{j,j})^{1/2}$, $\hat{\zeta}_{i,j} = (\hat{\alpha}_i + \hat{\beta}_j - \bar{\alpha}^*_i - \bar{\beta}^*_j)/(1/\hat{v}_{i,i} + 1/\hat{v}_{n+j,n+j})^{1/2}$, and $\hat{\eta}_{i,j} = [\hat{\beta}_i - \hat{\beta}_j - (\bar{\beta}^*_i - \bar{\beta}^*_j)]/(1/\hat{v}_{n+i,n+i} + 1/\hat{v}_{n+j,n+j})^{1/2}$ are all asymptotically distributed as standard normal random variables, where $\hat{v}_{i,i}$ is the estimate of $v_{i,i}$ by replacing $\bar{\boldsymbol{\theta}}^*$ with $\hat{\boldsymbol{\theta}}$. Therefore, we assess the asymptotic normality of $\hat{\xi}_{i,j}$, $\hat{\zeta}_{i,j}$ and $\hat{\eta}_{i,j}$ using the quantile–quantile (QQ) plot. Further, we also record the coverage probability of the 95% confidence interval, the length of the confidence interval and the frequency that the MLE does not exist. The results for $\hat{\xi}_{i,j}$, $\hat{\zeta}_{i,j}$ and $\hat{\eta}_{i,j}$ are similar, thus only the results of $\hat{\xi}_{i,j}$ are reported. Each simulation is repeated 10,000 times.

We consider two values for n, $n = 100$ and 200 and find that the QQ-plots for them are similar. Therefore, we only show the QQ-plots when $n = 200$ in Figure 1 to save space. In this figure, the horizontal and vertical axes are the theoretical and empirical quantiles, respectively, and the straight lines correspond to the reference line $y = x$. In Figure 1(b), we can see that when the weights are continuous and $L = \log n$ and $n^{1/2}$, the empirical quantiles coincide with the theoretical ones very well [the QQ-plots when $L = 0$ and $\log(\log n)$ are similar to those of $L = \log n$ and not shown]. On the other hand, for binary and discrete weights, when $L = 0$ and $\log(\log n)$, the empirical quantiles agree well with the theoretical ones while there are notable deviations when $L = (\log n)^{1/2}$; again, to save space, the QQ-plots for $L = 0$ in the case of binary weights and for $L = \log(\log n)$ in the case of discrete weights are not shown. When $L = \log n$, the MLE did not exist in all repetitions (see Table 1, thus the corresponding QQ-plot could not be shown).

Table 1 reports the coverage probability of the 95% confidence interval for $\alpha_i - \alpha_j$, the length of the confidence interval, and the frequency that the MLE did not exist. As we can see, the length of the confidence interval increases as L increases and decreases as n increases, which qualitatively agree with the theory. In the case of continuous weights, the coverage frequencies are all close to the nominal level, while in the case of binary and discrete weights, when $L = (\log n)^{1/2}$ (conditions in Theorem 6 no longer hold), the MLE often does not exist and the coverage frequencies for the $(1, 2)$ pair are higher than the nominal level; when $L = \log n$, the MLE did not exist in any of the repetitions.

4. Summary and discussion. In this paper, we have derived the uniform consistency and asymptotic normality of MLEs in the directed ERGM with the bi-degree sequence as the sufficient statistics; the edge weights are allowed to be binary, continuous or infinitely discrete and the number of vertices goes to infinity. In this class of models, a remarkable characterization is that the Fisher information matrix of the parameter vector is symmetric, nonnegative and diagonally dominant such that an approximately explicit expression of the MLE can be obtained.

In the case of discrete weights, only binary and infinitely countable values have been considered. In the finite discrete case, we may assume $a_{i,j}$ takes values in the set $\Omega = \{0, 1, \ldots, q - 1\}$, where q is a fixed constant. By (1.1), it can be shown that the probability mass function of $a_{i,j}$ is of the form

$$\mathbb{P}(a_{i,j} = a) = \frac{1 - e^{-(\alpha_i + \beta_j)}}{1 - e^{-(\alpha_i + \beta_j)q}} \times e^{-(\alpha_i + \beta_j)a}, \qquad a = 0, \ldots, q - 1,$$

and the likelihood equations become

$$d_i = \sum_{j \neq i} \frac{1 - e^{-(\alpha_i + \beta_j)}}{1 - e^{-(\alpha_i + \beta_j)q}} \sum_{k=0}^{q-1} e^{-k(\alpha_i + \beta_j)},$$

$$b_j = \sum_{i \neq j} \left(\frac{1}{e^{\hat{\alpha}_i + \hat{\beta}_j} - 1} - \frac{q}{e^{(\hat{\alpha}_i + \hat{\beta}_j)q} - 1} \right).$$
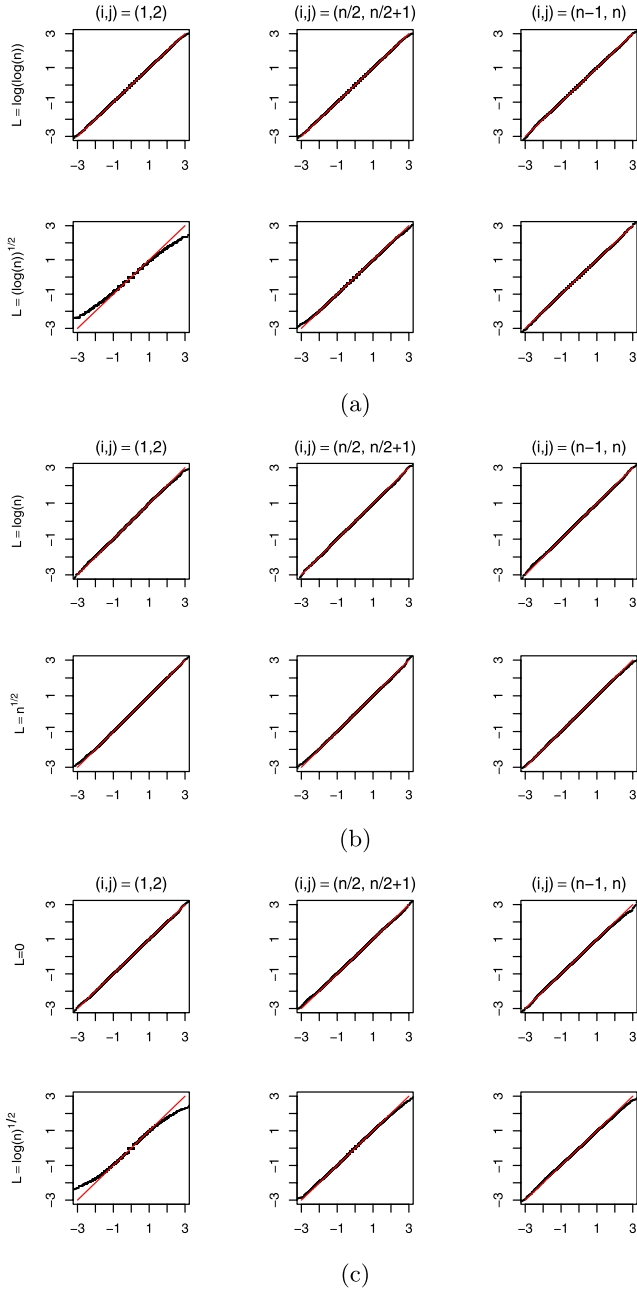
FIG. 1. *The QQ-plots of* $\hat{\xi}_{i,j}$ *(*$n = 200$*).* (a) *Binary weights.* (b) *Continuous weights.* (c) *Infinite discrete weights.*

TABLE 1
The reported values are the coverage frequency ($\times 100\%$) for $\alpha_i - \alpha_j$ for a pair (i, j)/the length of the confidence interval/the frequency ($\times 100\%$) that the MLE did not exist

n	(i, j)	$L = 0$	$L = \log(\log n)$	$L = (\log(n))^{1/2}$	$L = \log(n)$
			Binary weights		
100	(1, 2)	94.81/0.57/0	95.63/0.10/0.30	98.60/1.46/15.86	NA/NA/100
	(50, 51)	94.78/0.57/0	95.18/0.76/0.30	95.41/0.93/15.86	NA/NA/100
	(99, 100)	94.87/0.57/0	95.02/0.63/0.30	94.97/0.68/15.86	NA/NA/100
200	(1, 2)	95.35/0.40/0	95.50/0.75/0	98.13/1.10/1.02	NA/NA/100
	(100, 101)	95.03/0.40/0	95.08/0.55/0	95.23/0.68/1.02	NA/NA/100
	(199, 200)	95.28/0.40/0	95.32/0.45/0	95.26/0.48/1.02	NA/NA/100
			Continuous weights		
100	(1, 2)	95.46/1.12/0	95.32/2.37/0	95.55/4.82/0	95.16/9.09/0
	(50, 51)	95.28/1.12/0	95.44/1.93/0	95.71/3.48/0	95.51/6.13/0
	(99, 100)	95.38/1.12/0	95.63/1.50/0	95.81/2.07/0	95.72/2.83/0
200	(1, 2)	95.25/0.79/0	95.04/1.74/0	95.42/3.78/0	95.01/8.71/0
	(100, 101)	95.10/0.79/0	95.21/1.41/0	95.31/2.68/0	95.39/5.73/0
	(199, 200)	95.53/0.79/0	95.62/1.07/0	95.40/1.52/0	95.21/2.28/0
			Discrete weights		
100	(1, 2)	95.22/0.23/0	96.83/1.98/0.54	99.72/3.29/56.83	NA/NA/100
	(50, 51)	95.72/0.23/0	95.93/1.15/0.54	96.18/1.66/56.83	NA/NA/100
	(99, 100)	95.49/0.23/0	95.73/0.52/0.54	95.63/0.61/56.83	NA/NA/100
200	(1, 2)	95.08/0.16/0	96.02/1.51/0	98.26/2.56/12.63	NA/NA/100
	(100, 101)	95.31/0.16/0	95.55/0.87/0	95.43/1.23/12.63	NA/NA/100
	(199, 200)	95.28/0.16/0	95.54/0.38/0	95.31/0.44/12.63	NA/NA/100

It can be shown that the Fisher information matrix of $\boldsymbol{\theta}$ is also in the class of matrices $\mathcal{L}_n(m, M)$ under certain conditions. Therefore, except for some more complex calculations in contrast with the binary case, there is no extra difficulty to show that the conditions of Theorem 1 hold, and the consistency and asymptotic normality of the MLE in the finite discrete case can also be established.

It is worth noting that the conditions imposed on q_n and Q_n may not be best possible. In particular, the conditions guaranteeing the asymptotic normality seem stronger than those guaranteeing the consistency. For example, in the case of continuous weights, the consistency requires $Q_n/q_n = (n/\log n)^{1/18}$, while the asymptotic normality requires $Q_n/q_n = n^{1/50}/(\log n)^{1/25}$. Simulation studies suggest that the conditions on q_n and Q_n might be relaxed. We will investigate this in future studies and note that the asymptotic behavior of the MLE depends not only on q_n and Q_n, but also on the configuration of the parameters.

Regarding the p_1 model by [26], which is related to model (1.1), one of the key features of the p_1 model is to measure the dyad-dependent reciprocation by the reciprocity parameter ρ. In the p_1 model, there is also another parameter (λ) that

measures the density of edges, and the sufficient statistic of the density parameter λ is a linear combination of the in-degrees of vertices and the out-degrees of vertices. Specifically, the item $\lambda \sum_{i \neq j} a_{i,j} + \sum_i \alpha_i d_i + \sum_j \beta_j b_j$ in the p_1 model can be rewritten as $\sum_i (\alpha_i + \lambda + \beta_n) d_i + \sum_j (\beta_j - \beta_n) b_j$. Therefore, when there is no reciprocity parameter ρ, by taking the transformation of parameters $\tilde{\alpha}_i = \alpha_i + \lambda + \beta_n$ and $\tilde{\beta}_j = \beta_j - \beta_n$, we obtain the model (1.1). If the reciprocity parameter is incorporated into model (1.1), the induced Fisher information matrix is no longer diagonally dominant and Proposition 1 cannot be applied. However, simulation results in [52] indicate that the MLEs still enjoy the properties of uniform consistency and asymptotic normality, in which the asymptotic variances of the MLEs are the corresponding diagonal elements of the inverse of the Fisher information matrix. In order to extend the current work to study the reciprocity parameter, a new approximate matrix to the inverse of the Fisher information matrix is needed. We plan to investigate this problem in further work.

Finally, we note that the results in this paper can be potentially used to test the fit of the p_1 model. For example, the issue of testing the fit of the p_1 model has been discussed in several previous work, including [15, 17, 26, 37], but mostly in heuristic ways. In view of the result in this paper that the MLE enjoys good asymptotic properties in model (1.1), the conjectures in the above references on the asymptotic distribution of the likelihood ratio test for testing the fit of p_1 model seem reasonable. For example, to test $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$, the likelihood ratio test proposed by [26] is likely well approximated by the chi-square distribution with one degree of freedom.

APPENDIX: PROOFS OF THEOREMS

In this section, we give proofs for the theorems presented in Section 2.

A.1. Preliminaries. We first present the interior mapping theorem of the mean parameter space, and establish the geometric rate of convergence for the Newton iterative algorithm to solve a system of likelihood equations that will be used in this section.

A.1.1. *Uniqueness of the MLE.* Let σ_Ω be a σ-algebra over the set of weight values Ω and ν be a canonical σ-finite probability measure on (Ω, σ_Ω). In this paper, ν is the Borel measure in the case of continuous weight and the counting measure in the case of discrete weight. Denote $\nu^{n(n-1)}$ by the product measure on $\Omega^{n(n-1)}$. Let $\mathfrak{P}$ be all the probability distributions on $\Omega^{\binom{n}{2}}$ that are absolutely continuous with respective to $\nu^{\binom{n}{2}}$. Define the mean parameter space $\mathcal{M}$ to be the set of expected degree vectors tied to $\boldsymbol{\theta}$ from all distributions $\mathbb{P} \in \mathfrak{P}$:

$$\mathcal{M} = \{\mathbb{E}_\mathbb{P} \mathbf{g} : \mathbb{P} \in \mathfrak{P}\}.$$

Since a convex combination of probability distributions in $\mathfrak{P}$ is also a probability distribution in $\mathfrak{P}$, the set $\mathcal{M}$ is necessarily convex. If there is no linear combination of the sufficient statistics in an exponential family distribution that is constant, then the exponential family distribution is minimal. It is true for the probability distribution (1.1). If the natural parameter space Θ is open, then $\mathbb{P}$ is regular. By the general theory for a regular and minimal exponential family distribution (Theorem 3.3 of [49]), the gradient of the log-partition function maps the natural parameter space Θ to the interior of the mean parameter space $\mathcal{M}$, and this mapping

$$\nabla Z : \Theta \rightarrow \mathcal{M}^\circ$$

is bijective. Note that the solution to $\nabla Z(\boldsymbol{\theta}) = \mathbf{g}$ is precisely the MLE of $\boldsymbol{\theta}$. Thus, we have established the following.

PROPOSITION 5.   *Assume Θ is open. Then there exists a solution $\boldsymbol{\theta} \in \Theta$ to the MLE equation $\nabla Z(\boldsymbol{\theta}) = \mathbf{g}$ if and only if $\mathbf{g} \in \mathcal{M}^\circ$, and if such a solution exists, it is also unique.*

A.1.2. *Newton iterative theorem.*   Let D be an open convex subset of $\mathbb{R}^{2n-1}$, $\Omega(\mathbf{x}, r)$ denote the open ball $\{\mathbf{y} \in \mathbb{R}^{2n-1} : \|\mathbf{x} - \mathbf{y}\|_\infty < r\}$ and $\overline{\Omega(\mathbf{x}, r)}$ be its closure, where $\mathbf{x} \in \mathbb{R}^{2n-1}$. We will use Newton's iterative sequence to prove the existence and consistency of the MLE. Convergence properties of the Newton's iterative algorithm have been studied by many mathematicians [28, 35, 36, 38, 47]. For the ad-hoc system of likelihood equations considered in this paper, we establish a fast geometric rate of convergence for the Newton's iterative algorithm given in the following theorem, whose proof is given in Online Supplementary Materials [51].

THEOREM 7.   *Define a system of equations*

$$F_i(\boldsymbol{\theta}) = d_i - \sum_{k=1, k \neq i}^{n} f(\alpha_i + \beta_k), \qquad i = 1, \ldots, n,$$

$$F_{n+j}(\boldsymbol{\theta}) = b_j - \sum_{k=1, k \neq j}^{n} f(\alpha_k + \beta_j), \qquad j = 1, \ldots, n - 1,$$

$$F(\boldsymbol{\theta}) = \big(F_1(\boldsymbol{\theta}), \ldots, F_n(\boldsymbol{\theta}), F_{n+1}(\boldsymbol{\theta}), \ldots, F_{2n-1}(\boldsymbol{\theta})\big)^\top,$$

where $f(\cdot)$ is a continuous function with the third derivative. Let $D \subset \mathbb{R}^{2n-1}$ be a convex set and assume for any $\mathbf{x}, \mathbf{y}, \mathbf{v} \in D$, we have

(A.1) $$\big\|[F'(\mathbf{x}) - F'(\mathbf{y})]\mathbf{v}\big\|_\infty \leq K_1 \|\mathbf{x} - \mathbf{y}\|_\infty \|\mathbf{v}\|_\infty,$$

(A.2) $$\max_{i=1,\ldots,2n-1} \big\|F_i'(\mathbf{x}) - F_i'(\mathbf{y})\big\|_\infty \leq K_2 \|\mathbf{x} - \mathbf{y}\|_\infty,$$

where $F'(\boldsymbol{\theta})$ is the Jacobin matrix of F on $\boldsymbol{\theta}$ and $F_i'(\boldsymbol{\theta})$ is the gradient function of F_i on $\boldsymbol{\theta}$. Consider $\boldsymbol{\theta}^{(0)} \in D$ with $\Omega(\boldsymbol{\theta}^{(0)}, 2r) \subset D$, where $r = \|[F'(\boldsymbol{\theta}^{(0)})]^{-1} F(\boldsymbol{\theta}^{(0)})\|_\infty$. For any $\boldsymbol{\theta} \in \Omega(\boldsymbol{\theta}^{(0)}, 2r)$, we assume

(A.3) $$F'(\boldsymbol{\theta}) \in \mathcal{L}_n(m, M) \quad or \quad -F'(\boldsymbol{\theta}) \in \mathcal{L}_n(m, M).$$

For $k = 1, 2, \ldots$, define the Newton iterates $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - [F'(\boldsymbol{\theta}^{(k)})]^{-1} F(\boldsymbol{\theta}^{(k)})$. Let

(A.4) $$\rho = \frac{c_1(2n-1)M^2 K_1}{2m^3 n^2} + \frac{K_2}{(n-1)m}.$$

If $\rho r < 1/2$, then $\boldsymbol{\theta}^{(k)} \in \Omega(\boldsymbol{\theta}^{(0)}, 2r)$, $k = 1, 2, \ldots$, are well defined and satisfy

(A.5) $$\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(0)}\|_\infty \le r/(1 - \rho r).$$

Further, $\lim_{k \to \infty} \boldsymbol{\theta}^{(k)}$ exists and the limiting point is precisely the solution of $F(\boldsymbol{\theta}) = 0$ in the range of $\boldsymbol{\theta} \in \Omega(\boldsymbol{\theta}^{(0)}, 2r)$.

A.2. Proofs of Theorems 1 and 2.

A.2.1. *Proof of Theorem* 1. Assume that condition (2.3) holds. Recall the Newton's iterates $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - [F'(\boldsymbol{\theta}^{(k)})]^{-1} F(\boldsymbol{\theta}^{(k)})$ with $\boldsymbol{\theta}^{(0)} = \boldsymbol{\theta}^*$. If $\boldsymbol{\theta} \in \Omega(\boldsymbol{\theta}^*, 2r)$, then $-F'(\boldsymbol{\theta}) \in \mathcal{L}_n(m, M)$ with

$$M = \frac{1}{4}, \qquad m = \frac{e^{2(\|\boldsymbol{\theta}^*\|_\infty + 2r)}}{(1 + e^{2(\|\boldsymbol{\theta}^*\|_\infty + 2r)})^2}.$$

If $\|\boldsymbol{\theta}^*\|_\infty \le \tau \log n$ with the constant τ satisfying $0 < \tau < 1/16$, then as $n \to \infty$, $n^{-1/2}(\log n)^{1/2} e^{8\|\boldsymbol{\theta}^*\|} \le n^{-1/2+8\tau}(\log n)^{1/2} \to 0$. By Lemma 2 and condition (2.3), for sufficiently small r,

$$\rho r \le \left[\frac{c_1(2n-1)M^2(n-1)}{2m^3 n^2} + \frac{(n-1)}{2m(n-1)} \right]$$
$$\times \frac{(\log n)^{1/2}}{n^{1/2}} (c_{11} e^{6\|\boldsymbol{\theta}^*\|_\infty} + c_{12} e^{2\|\boldsymbol{\theta}^*\|_\infty})$$
$$\le O\left(\frac{(\log n)^{1/2} e^{12\|\boldsymbol{\theta}^*\|_\infty}}{n^{1/2}} \right) + O\left(\frac{(\log n)^{1/2} e^{8\|\boldsymbol{\theta}^*\|_\infty}}{n^{1/2}} \right).$$

Therefore, if $\|\boldsymbol{\theta}^*\|_\infty \le \tau \log n$, then $\rho r \to 0$ as $n \to \infty$. Consequently, by Theorem 7, $\lim_{n \to \infty} \hat{\boldsymbol{\theta}}^{(n)}$ exists. Denote the limit as $\hat{\boldsymbol{\theta}}$, then it satisfies

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty \le 2r = O\left(\frac{(\log n)^{1/2} e^{8\|\boldsymbol{\theta}^*\|_\infty}}{n^{1/2}} \right) = o(1).$$

By Lemma 3, condition (2.3) holds with probability approaching one, thus the above inequality also holds with probability approaching one. The uniqueness of the MLE comes from Proposition 5.

A.2.2. *Proof of Theorem 2.* Before proving Theorem 2, we first establish two lemmas.

LEMMA 8. *Let $R = V^{-1} - S$ and $U = \mathrm{Cov}[R\{\mathbf{g} - \mathbb{E}\mathbf{g}\}]$. Then*

$$(A.6) \qquad \|U\| \le \|V^{-1} - S\| + \frac{(1 + e^{2\|\theta^*\|_\infty})^4}{4e^{4\|\theta^*\|_\infty}(n-1)^2}.$$

PROOF. Note that

$$U = WVW^\top = (V^{-1} - S) - S(I - VS),$$

where I is a $(2n - 1) \times (2n - 1)$ diagonal matrix, and by inequality (C3) in [51], we have

$$\left|\{S(I - VS)\}_{i,j}\right| = |w_{i,j}| \le \frac{3(1 + e^{2\|\theta^*\|_\infty})^4}{4e^{4\|\theta^*\|_\infty}(n-1)^2}.$$

Thus,

$$\|U\| \le \|V^{-1} - S\| + \|S(I_{2n-1} - VS)\|$$

$$\le \|V^{-1} - S\| + \frac{3(1 + e^{2\|\theta^*\|_\infty})^4}{4e^{4\|\theta^*\|_\infty}(n-1)^2}. \qquad \square$$

LEMMA 9. *Assume that the conditions in Theorem 1 hold. If $\|\theta^*\|_\infty \le \tau \log n$ and $\tau < 1/40$, then for any i,*

$$(A.7) \qquad \hat{\theta}_i - \theta_i^* = \left[V^{-1}\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}\right]_i + o_p(n^{-1/2}).$$

PROOF. By Theorem 1, we have

$$\hat{\rho}_n := \max_{1 \le i \le 2n-1}|\hat{\theta}_i - \theta_i^*| = O_p\left(\frac{(\log n)^{1/2}e^{8\|\theta\|_\infty}}{n^{1/2}}\right).$$

Let $\hat{\gamma}_{i,j} = \hat{\alpha}_i + \hat{\beta}_j - \alpha_i - \beta_j$. By Taylor's expansion, for any $1 \le i \ne j \le n$,

$$\frac{e^{\hat{\alpha}_i + \hat{\beta}_j}}{1 + e^{\hat{\alpha}_i + \hat{\beta}_j}} - \frac{e^{\alpha_i^* + \beta_j^*}}{1 + e^{\alpha_i^* + \beta_j^*}} = \frac{e^{\alpha_i^* + \beta_j^*}}{(1 + e^{\alpha_i^* + \beta_j^*})^2}\hat{\gamma}_{i,j} + h_{i,j},$$

where

$$h_{i,j} = \frac{e^{\alpha_i^* + \beta_j^* + \phi_{i,j}\hat{\gamma}_{i,j}}(1 - e^{\alpha_i^* + \beta_j^* + \phi_{i,j}\hat{\gamma}_{i,j}})}{2(1 + e^{\alpha_i^* + \beta_j^* + \phi_{i,j}\hat{\gamma}_{i,j}})^3}\hat{\gamma}_{i,j}^2,$$

and $0 \le \phi_{i,j} \le 1$. By the likelihood equations (2.2), we have

$$\mathbf{g} - \mathbb{E}(\mathbf{g}) = V(\hat{\theta} - \theta^*) + \mathbf{h},$$

where $\mathbf{h} = (h_1, \ldots, h_{2n-1})^\top$ and,

$$h_i = \sum_{k=1, k \neq i}^{n} h_{i,k}, \qquad i = 1, \ldots, n,$$

$$h_{n+i} = \sum_{k=1, k \neq i}^{n} h_{k,i}, \qquad i = 1, \ldots, n-1.$$

Equivalently,

(A.8) $$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* = V^{-1}(\mathbf{g} - \mathbb{E}(\mathbf{g})) + V^{-1}\mathbf{h}.$$

Since $|e^x(1 - e^x)/(1 + e^x)^3| \leq 1$, we have

$$|h_{i,j}| \leq |\hat{\gamma}_{i,j}^2|/2 \leq 2\hat{\rho}_n^2, \qquad |h_i| \leq \sum_{j \neq i} |h_{i,j}| \leq 2(n-1)\hat{\rho}_n^2.$$

Note that $(S\mathbf{h})_i = h_i/v_{i,i} + (-1)^{1_{\{i>n\}}} h_{2n}/v_{2n,2n}$, and $(V^{-1}\mathbf{h})_i = (S\mathbf{h})_i + (R\mathbf{h})_i$. By direct calculations, we have

$$|(S\mathbf{h})_i| \leq \frac{|h_i|}{v_{i,i}} + \frac{|h_{2n}|}{v_{2n,2n}} \leq \frac{16\hat{\rho}_n^2(1 + e^{2\|\boldsymbol{\theta}^*\|_\infty})^2}{e^{2\|\boldsymbol{\theta}^*\|_\infty}} \leq O\left(\frac{e^{20\|\boldsymbol{\theta}^*\|_\infty} \log n}{n}\right),$$

and by Proposition 1, we have

$$|(R\mathbf{h})_i| \leq \|R\|_\infty \times \left[(2n-1) \max_i |h_i|\right] \leq O\left(\frac{e^{22\|\boldsymbol{\theta}^*\|_\infty} \log n}{n}\right).$$

If $\|\boldsymbol{\theta}^*\|_\infty \leq \tau \log n$ and $\tau < 1/44$, then

$$|(V^{-1}h)_i| \leq |(Sh)_i| + |(Rh)_i| = o(n^{-1/2}).$$

This completes the proof. □

PROOF OF THEOREM 2.   By (A.8), we have

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})_i = \left[S\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}\right]_i + \left[R\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}\right]_i + (V^{-1}\mathbf{h})_i.$$

By Lemmas 8 and 9, if $\|\boldsymbol{\theta}^*\|_\infty \leq \tau \log n$ and $\tau < 1/44$, then

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})_i = \left[S\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}\right]_i + o_p(n^{-1/2}).$$

Therefore, Theorem 2 follows directly from Proposition 2. □

SUPPLEMENTARY MATERIAL

Supplement to "Asymptotics in directed exponential random graph models with an increasing bi-degree sequence." (DOI: 10.1214/15-AOS1343SUPP; .pdf). The supplemental material contains proofs for the lemmas in Section 2.2, the theorems and lemmas in Sections 2.3 and 2.4, Proposition 1 and Theorem 7.

REFERENCES

[1] ADAMIC, L. A. and GLANCE, N. (2005). The political blogosphere and the 2004 US Election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery* 36–43. ACM, New York.

[2] AKOGLU, L., VAZ DE MELO, P. O. S. and FALOUTSOS, C. (2012). Quantifying reciprocity in large weighted communication networks. *Advances in Knowledge Discovery and Data Mining*, *Lecture Notes in Computer Science* **7302** 85–96.

[3] BADER, G. D. and HOGUE, C. W. V. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4** 2–27.

[4] BARNDORFF-NIELSEN, O. (1973). Exponential families and conditioning. Ph.D. thesis, Univ. of Copenhagen.

[5] BERK, R. H. (1972). Consistency and asymptotic normality of MLE's for exponential models. *Ann. Mat. Statist.* **43** 193–204. MR0298810

[6] BICKEL, P. J., CHEN, A. and LEVINA, E. (2011). The method of moments and degree distributions for network models. *Ann. Statist.* **39** 2280–2301. MR2906868

[7] BOLLA, M. and ELBANNA, A. (2014). Estimating parameters of a multipartite loglinear graph model via the EM algorithm. Preprint. Available at arXiv:1411.7934.

[8] BRADLEY, R. A. and TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika* **39** 324–345. MR0070925

[9] CHATTERJEE, S. and DIACONIS, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.* **41** 2428–2461. MR3127871

[10] CHATTERJEE, S., DIACONIS, P. and SLY, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. MR2857452

[11] CHEN, N. and OLVERA-CRAVIOTO, M. (2013). Directed random graphs with given degree distributions. *Stoch. Syst.* **3** 1–40.

[12] DIESNER, J. and CARLEY, K. M. (2005). Exploration of communication networks from the Enron email corpus. In *Proceedings of Workshop on Link Analysis*, *Counterterrorism and Security*, *SIAM International Conference on Data Mining* 3–14. SIAM, Philadelphia, PA.

[13] ERDŐS, P. L., MIKLÓS, I. and TOROCZKAI, Z. (2010). A simple Havel–Hakimi type algorithm to realize graphical degree sequences of directed graphs. *Electron. J. Combin.* **17** Research Paper 66, 10. MR2644852

[14] FIENBERG, S. E. (2012). A brief history of statistical models for network analysis and open challenges. *J. Comput. Graph. Statist.* **21** 825–839. MR3005799

[15] FIENBERG, S. E., PETROVIĆ, S. and RINALDO, A. (2011). Algebraic statistics for p_1 random graph models: Markov bases and their uses. In *Looking Back. Lect. Notes Stat. Proc.* (N. J. Dorans and S. Sinharay, eds.) **202** 21–38. Springer, New York. MR2856692

[16] FIENBERG, S. E. and RINALDO, A. (2012). Maximum likelihood estimation in log-linear models. *Ann. Statist.* **40** 996–1023. MR2985941

[17] FIENBERG, S. E. and WASSERMAN, S. (1981). An exponential family of probability distributions for directed graphs: Comment. *J. Amer. Statist. Assoc.* **76** 54–57.

[18] FIENBERG, S. E. and WASSERMAN, S. S. (1981). Categorical data analysis of single sociometric relations. *Sociol. Method.* **1981** 156–192.

[19] FISCHER, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika* **46** 59–77. MR0655008

[20] GIRVAN, M. and NEWMAN, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **99** 7821–7826 (electronic). MR1908073

[21] HABERMAN, S. J. (1977). Maximum likelihood estimates in exponential response models. *Ann. Statist.* **5** 815–841. MR0501540

[22] HABERMAN, S. J. (1981). An exponential family of probability distributions for directed graphs: Comment. *J. Amer. Statist. Assoc.* **76** 60–61.

[23] HANDCOCK, M. S. (2003). Assessing degeneracy in statistical models of social networks, Working Paper 39. Technical report, Center for Statistics and the Social Sciences, Univ. Washington, Seattle, WA.

[24] HELLERINGER, S. and KOHLER, H.-P. (2007). Sexual network structure and the spread of HIV in Africa: Evidence from Likoma Island, Malawi. *AIDS* **21** 2323–2332.

[25] HILLAR, C. and WIBISONO, A. (2013). Maximum entropy distributions on graphs. Preprint. Available at arXiv:1301.3321.

[26] HOLLAND, P. W. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176

[27] HUNTER, D. R. and HANDCOCK, M. S. (2006). Inference in curved exponential family models for networks. *J. Comput. Graph. Statist.* **15** 565–583. MR2291264

[28] KANTOROVIČ, L. V. (1948). On Newton's method for functional equations. *Dokl. Akad. Nauk SSSR* **59** 1237–1240. MR0024700

[29] KIM, H., DEL GENIO, C. I., BASSLER, K. E. and TOROCZKAI, Z. (2012). Constructing and sampling directed graphs with given degree sequences. *New J. Phys.* **14** 023012.

[30] KOSSINETS, G. and WATTS, D. J. (2006). Empirical analysis of an evolving social network. *Science* **311** 88–90. MR2192483

[31] LOÈVE, M. (1977). *Probability Theory. I*, 4th ed. Springer, New York. MR0651017

[32] NEPUSZ, T., YU, H. and PACCANARO, A. (2012). Detecting overlapping protein complexes in protein–protein interaction networks. *Nat. Methods* **18** 471–472.

[33] NEWMAN, M. E. J. (2002). Spread of epidemic disease on networks. *Phys. Rev. E (3)* **66** 016128, 11. MR1919737

[34] OLHEDE, S. C. and WOLFE, P. J. (2012). Degree-based network models. Preprint. Available at arXiv:1211.6537.

[35] ORTEGA, J. M. (1968). The Newton–Kantorovich theorem. *Amer. Math. Monthly* **75** 658–660. MR0231218

[36] ORTEGA, J. M. and RHEINBOLDT, W. C. (1970). *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York. MR0273810

[37] PETROVIĆ, S., RINALDO, A. and FIENBERG, S. E. (2010). Algebraic statistics for a directed random graph model with reciprocation. In *Algebraic Methods in Statistics and Probability II. Contemp. Math.* **516** (M. A. G. Vianaand and H. P. Wynn, eds.) 261–283. Amer. Math. Soc., Providence, RI. MR2730754

[38] POLYAK, B. T. (2004). Newton–Kantorovich method and its global convergence. *J. Math. Sci.* **133** 1513–1523.

[39] RINALDO, A., PETROVIĆ, S. and FIENBERG, S. E. (2013). Maximum likelihood estimation in the β-model. *Ann. Statist.* **41** 1085–1110. MR3113804

[40] ROBINS, G. and PATTISON, P. (2007). An introduction to exponential random graph (p^*) models for social networks. *Soc. Netw.* **29** 173–191.

[41] ROBINS, G., PATTISON, P. and WANG, P. (2009). Closure, connectivity and degree distributions: Exponential random graph (p^*) models for directed social networks. *Soc. Netw.* **31** 105–117.

[42] ROBINS, G. L., SNIJDERS, T. A. B., WANG, P., HANDCOCK, M. and PATTISON, P. (2007). Recent developments in exponential random graph (p*) models for social networks. *Soc. Netw.* **29** 192–215.

[43] SALATHÉA, M., KAZANDJIEVAB, M., LEEB, J. W., LEVISB, P., MARCUS, FELDMAN, M. W. and JONES, J. H. (2010). A high-resolution human contact network for infectious disease transmission. *Proc. Natl. Acad. Sci. USA* **107** 22020–22025.

[44] SCHWEINBERGER, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *J. Amer. Statist. Assoc.* **106** 1361–1370. MR2896841

[45] SHALIZI, C. R. and RINALDO, A. (2013). Consistency under sampling of exponential random graph models. *Ann. Statist.* **41** 508–535. MR3099112

[46] SIMONS, G. and YAO, Y.-C. (1999). Asymptotics when the number of parameters tends to infinity in the Bradley–Terry model for paired comparisons. *Ann. Statist.* **27** 1041–1060. MR1724040

[47] TAPIA, R. A. (1971). Classroom Notes: The Kantorovich theorem for Newton's method. *Amer. Math. Monthly* **78** 389–392. MR1536290

[48] VON MERING, C., KRAUSE, R., SNEL, B., CORNELL, M., OLIVER, S. G., FIELDS, S. and BORK, P. (2002). Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417** 399–403.

[49] WAINWRIGHT, M. and JORDAN, M. I. (2008). Graphical models, exponential families, and variational inference. *Faund. Trends Mach. Learn.* **1** 1–305.

[50] WU, N. (1997). *The Maximum Entropy Method*. Springer, Berlin. MR1456995

[51] YAN, T., LENG, C. and ZHU, J. (2015). Supplement to "Asymptotics in directed exponential random graph models with an increasing bi-degree sequence." DOI:10.1214/15-AOS1343SUPP.

[52] YAN, T. and LENG, C. (2015). A simulation study of the p_1 model for directed random graphs. *Stat. Interface* **8** 255–266. MR3341325

[53] 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. MR3068452

[54] 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. MR3282018

[55] 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. MR3059083

T. YAN
DEPARTMENT OF STATISTICS
CENTRAL CHINA NORMAL UNIVERSITY
WUHAN, 430079
CHINA
E-MAIL: tingyanty@mail.ccnu.edu.cn

C. LENG
DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY, CV4 7AL
UNITED KINGDOM
E-MAIL: C.leng@warwick.ac.uk

J. ZHU
DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
439 WEST HALL
1085 S. UNIVERSITY AVE
ANN ARBOR, MICHIGAN 48109-1107
USA
E-MAIL: jizhu@umich.edu