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

BY TING YAN*, CHENLEI LENg AND Ji ZHU†

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

Although the asymptotic analyses for undirected network models based on the degree sequence have been greatly addressed recently, those for directed models remain open problems. 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 or non-binary weighted edges. We establish the uniform consistency and the asymptotic normality for the maximum likelihood estimate despite of a growing number of parameters and only one realized observation of a random graph. One key technical step in the proofs is approximating the inverse of the Fisher information matrix by using a simple matrix with high-quality approximate errors. Along the way, we establish a geometrically fast rate of convergence for the Newton iteration algorithm to solve the maximum likelihood estimate. 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 structure, and they are being collected in a variety of fields including social networks [18, 26], communication networks [1, 10, 2], biological networks [43, 3, 30], disease transmission networks [29, 40] 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 understand the uncertainty associated with observed outcomes [37, 38]. At the same time, data with network structure pose new challenges for statistical inference, in particular asymptotic analysis in which only one sample is observed and one is often interested in asymptotic background that the size of network becomes large [12].

The in- and out-degrees of vertices (or degrees for undirected networks) preliminarily summarize the information contained in network data and their distributions provide important insights for understanding the generative mechanism of networks. In the undirected case, the degree sequence has been extensively studied [8, 6, 31, 49, 36, 21]. Its distributions are explored under the framework of the exponential family parameterized by the so-called “potentials” of vertices recently, i.e., “β-model” by [8] for binary edges or “maximum entropy models” by [21] for weighted edges in which the degree sequence is the exclusively sufficient statistic. The asymptotic theories of the maximum likelihood estimate (MLE) in these models have not been derived until recently [8, 21, 47, 48]. In the directed case, how to construct and sample directed graphs with given in- and out-degree (sometimes referred to “bi-degree”) sequences and related problems are studied in mathematical and physical literatures [11, 9, 25], but statistical inference is little known, especially for asymptotic analysis. [39] study the distributions of the bi-degrees through empirical examples for social networks, but lack of substantial theoretical analysis.

Similar to the undirected case, we will study the distribution of the bi-degree sequence with it

---

*Research partially supported by the National Science Foundation of China (No. 11341001)
†Research partially supported by the National Science Foundation (DMS 0748389)
Primary 62F10, 62F12; secondary 62B05, 62E20, 05C80

Keywords and phrases: Bi-degree sequence, Central limit theorem, Consistency, Directed exponential random graph models, Fisher information matrix, Maximum likelihood estimation
as the exclusively sufficient statistic. The Koopman-Pitman-Darmois theorem or the principle of maximum entropy [45, 44] force its probability mass function to 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 or continuous) and make the maximum likelihood inference.

Notice that one out-degree parameter and one in-degree parameter are required to attach to the bi-degree of each vertex. As a result, the total number of parameters are twice of that of vertices. As the size of networks increases, the number of parameters goes to infinity. This makes asymptotic inference challenge. Establishing the uniform consistency and asymptotic normality of the MLE is the aim of this paper. To the best of our knowledge, it is the first time to derive such results in directed exponential random graph models with weighted edges. We remark that our proofs are non-trivial. One key feature of our proofs is approximating the inverse of the Fisher information matrix by using a simple matrix with high-quality approximation errors. It is utilized to derive the geometrically fast rate of convergence in the Newton iterative algorithm.

Next, we formally describe the models considered in this paper. Consider a directed graph $G$ 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 $G$. We consider three special 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, i.e., $a_{i,i} = 0$. Let $d_i = \sum_{j \neq i} a_{i,j}$ be the out-degree of vertex $i$ and $d = (d_1, \ldots, d_n)^\top$ be the out-degree sequence of the graph $G$. Similarly, define $b_j = \sum_{i \neq j} a_{i,j}$ as the in-degree of vertex $j$ and $b = (b_1, \ldots, b_n)^\top$ as the in-degree sequence. The pair $(b, d)$ or $\{(b_1, d_1), \ldots, (b_n, d_n)\}$ are the bi-degree sequence. The density or probability mass function on $G$ parameterized by exponential family distributions with respect to some canonical measure $\nu$ is

$$p(G) = \exp\left(\alpha^\top d + \beta^\top b - Z(\alpha, \beta)\right),$$

where $Z(\alpha, \beta)$ is the log-partition function, $\alpha = (a_1, \ldots, a_n)^\top$ is a parameter vector tied to the out-degree sequence, and $\beta = (\beta_1, \ldots, \beta_n)^\top$ is a parameter vector tied to the in-degree sequence. This model can be viewed as a generation of the $p_1$ model without reciprocity effect to weighted edges and a directed version of the $\beta$-model. It can be also represented as the log-linear model [14, 15, 16] and the algorithm developed for the log-linear model can be used to solve the MLE. As explained by [23], $\alpha_i$ quantifies the effect of an outgoing edge from vertex $i$ and $\beta_j$ quantifies the effect of an incoming edge connecting to vertex $j$. If $\alpha_i$ is large and positive, vertex $i$ will tend to have a relatively large out-degree. Similarly, if $\beta_j$ is large and positive, vertex $j$ tends to have a relatively large in-degree. Note that

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

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

$$Z(\alpha, \beta) = \sum_{i \neq j} Z_1(\alpha_i + \beta_j) := \sum_{i \neq j} \log \int_{\Omega} \exp((\alpha_i + \beta_j) a_{i,j}) \nu(da_{i,j}).$$
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 \((\alpha, \beta)\) to \((\alpha - c, \beta + c)\), the likelihood does not change. Therefore, for identifiability, constraints on \(\alpha\) or \(\beta\) are necessary. In this paper, we choose to set \(\beta_n = 0\). Other constraints are also possible, e.g., \(\sum_i \alpha_i = 0\) or \(\sum_j \beta_j = 0\). In total, there are \(2n - 1\) independent parameters and the natural parameter space becomes
\[
\Theta = \{ (\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_{n-1})^\top \in R^{2n-1} : Z(\alpha, \beta) < \infty \}.
\]
Model (1.1) can be served as null models for hypothesis testing such as [23, 15] and to reconstruct directed networks and make statistical inference in a situation in which the bi-degree sequence is only available due to privacy projection.

For the remainder of the paper, we proceed as follows. In Section 2, we first introduce notations 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 Subsections 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 Subsections 2.3 and 2.4 are similar to those in Subsection 2.2, we only show the proofs of the lemmas and theorems in Subsection 2.2 in Section 5 and the proofs in Subsections 2.3 and 2.4 as well as that of Proposition 1, are relegated to the Supplementary Material.

2. Main results.

2.1. Notations and Preparations. Let \(R_+ = (0, \infty)\), \(R_0 = [0, \infty)\), \(N = \{1, 2, \ldots\}\), \(N_0 = \{0, 1, 2, \ldots\}\). For a subset \(C \subset R^n\), let \(C^0\) and \(\overline{C}\) denote the interior and closure of \(C\), respectively. For a vector \(x = (x_1, \ldots, x_n)^\top \in R^n\), denote by \(|x|_\infty = \max_{1 \leq i \leq n} |x_i|\) the \(\ell_\infty\)-norm of \(x\). For an \(n \times n\) matrix \(J = (J_{i,j})\), let \(|J|_\infty\) denote the matrix norm induced by the \(\ell_\infty\)-norm on vectors in \(R^n\), i.e.
\[
|J|_\infty = \max_{x \neq 0} \frac{|Jx|_\infty}{|x|_\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |J_{i,j}|.
\]

In order to characterize the Fisher information matrix, we introduce a matrix class. 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 \(L_n(m, M)\) if the following holds:
\[
\begin{align*}
m &\leq v_{i,i} - \sum_{j=n+1}^{2n-1} v_{i,j} \leq M, \quad i = 1, \ldots, n - 1; \quad v_{n,n} = \sum_{j=n+1}^{2n-1} v_{n,j}, \\
v_{i,j} &= 0, \quad i, j = 1, \ldots, n, \quad i \neq j, \\
v_{i,j} &= 0, \quad i, j = n + 1, \ldots, 2n - 1, \quad i \neq j, \\
m &\leq v_{i,j} = v_{j,i} \leq M, \quad i = 1, \ldots, n, \quad j = n + 1, \ldots, 2n - 1, \quad j \neq n + i, \\
v_{i,n+i} &= v_{n+i,i} = 0, \quad i = 1, \ldots, n - 1, \\
v_{i,i} = \sum_{k=1}^{n} v_{i,k} = \sum_{k=1}^{n} v_{k,i}, \quad i = n + 1, \ldots, 2n - 1.
\end{align*}
\]

Clearly, if \(V \in L_n(m, M)\), then \(V\) is a \((2n - 1) \times (2n - 1)\) diagonally dominant, symmetric nonnegative matrix and \(V\) has the following structure:
\[
V = \begin{pmatrix}
V_{11} & V_{12} \\
V_{12} & V_{22}
\end{pmatrix},
\]
where \( V_{11} \) (\( n \) by \( n \)) and \( V_{22} \) (\( n-1 \) by \( n-1 \)) are diagonal matrices, \( V_{12} \) is a nonnegative matrix whose non-diagonal elements are positive and diagonal elements equal to zero.

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

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

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

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.

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

\[
\|V^{-1}-S\| \leq \frac{c_1M^2}{m^3(n-1)^2},
\]

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 \( x \in \mathbb{R}^{2n-1} \),

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

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

Let \( \theta = (\alpha_1,\ldots,\alpha_n,\beta_1,\ldots,\beta_{n-1})^T \) and \( g = (d_1,\ldots,d_n,b_1,\ldots,b_{n-1})^T \). Henceforth, we will use \( V \) to denote the Fisher information matrix of the parameter vector \( \theta \) 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 \( \Theta \), the likelihood equations, and prove the existence, uniqueness, consistency and asymptotic normality of the MLE. We defer the proofs of the results in Subsection 2.2 to Section 5 and all other proofs in Subsection 2.3 and 2.4 to the Online Supplementary Material.

### 2.2. Binary weights

In the case of binary weights, i.e. \( \Omega = \{0,1\} \), \( \nu \) is the counting measure, and \( a_{i,j}, 1 \leq i \neq j \leq n \) are mutually independent Bernoulli random variables with

\[
P(a_{i,j} = 1) = \frac{e^{\alpha_i+\beta_j}}{1+e^{\alpha_i+\beta_j}}.
\]
This probability distribution is the version of the $p_1$ model without reciprocal effect, originally proposed by [23]. The log-partition function $Z(\theta)$ is $\sum_{i\neq j} \log(1 + e^{\alpha_i + \beta_j})$ and the likelihood equations are

$$
\begin{align*}
    d_i &= \sum_{k=1, k \neq i}^{n} \frac{e^{\alpha_i + \beta_k}}{1 + e^{\alpha_i + \beta_k}}, \quad i = 1, \ldots, n, \\
    b_j &= \sum_{k=1, k \neq j}^{n} \frac{e^{\alpha_k + \beta_j}}{1 + e^{\alpha_k + \beta_j}}, \quad j = 1, \ldots, n - 1,
\end{align*}
$$

(2.2)

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

We first establish the existence and consistency of $\hat{\theta}$ by applying Theorem 8 in the Online Supplementary Material. Define a system of functions:

$$
\begin{align*}
    F_i(\theta) &= d_i - \sum_{k=1, k \neq i}^{n} \frac{e^{\alpha_i + \beta_k}}{1 + e^{\alpha_i + \beta_k}}, \quad i = 1, \ldots, n, \\
    F_{n+j}(\theta) &= b_j - \sum_{k=1, k \neq j}^{n} \frac{e^{\alpha_k + \beta_j}}{1 + e^{\alpha_k + \beta_j}}, \quad j = 1, \ldots, n - 1, \\
    F(\theta) &= (F_1(\theta), \ldots, F_{2n-1}(\theta))^\top.
\end{align*}
$$

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

$$
\begin{align*}
    \frac{\partial F_i}{\partial \alpha_l} &= 0, \quad l = 1, \ldots, n, \quad l \neq i; \quad \frac{\partial F_i}{\partial \alpha_i} = -\sum_{k=1, k \neq i}^{n} \frac{e^{\alpha_i + \beta_k}}{(1 + e^{\alpha_i + \beta_k})^2}, \\
    \frac{\partial F_i}{\partial \beta_j} &= -\frac{e^{\alpha_i + \beta_j}}{(1 + e^{\alpha_i + \beta_j})^2}, \quad j = 1, \ldots, n - 1, \quad j \neq i; \quad \frac{\partial F_i}{\partial \beta_i} = 0
\end{align*}
$$

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

$$
\begin{align*}
    \frac{\partial F_{n+j}}{\partial \alpha_l} &= -\frac{e^{\alpha_l + \beta_j}}{(1 + e^{\alpha_l + \beta_j})^2}, \quad l = 1, \ldots, n, \quad l \neq j; \quad \frac{\partial F_{n+j}}{\partial \alpha_j} = 0, \\
    \frac{\partial F_{n+j}}{\partial \beta_j} &= -\sum_{k=1, k \neq j}^{n} \frac{e^{\alpha_k + \beta_j}}{(1 + e^{\alpha_k + \beta_j})^2}; \quad \frac{\partial F_{n+j}}{\partial \beta_l} = 0, \quad l = 1, \ldots, n - 1.
\end{align*}
$$

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 $-F'(\theta) \in \mathcal{L}_n(m, M)$, thus Proposition 1 and Theorem 8 can be applied. Let $\theta^*$ denote the true parameter vector. The constants $K_1, K_2$ and $r$ in the upper bounds of Theorem 8 are given in the following lemma.

**Lemma 2.** Take $D = R^{2n-1}$ and $\theta^{(0)} = \theta^*$ in Theorem 8. Assume

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

Then we can choose the constants $K_1, K_2$ and $r$ in Theorem 8 as

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

where $c_{11}$ and $c_{12}$ are constants.
The following lemma assures that condition (2.3) holds with a large probability.

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

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

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

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

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

Further, if the MLE exists, it is unique.

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

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

where $g_i$ denotes the $i$th element of $g$. The identity between the likelihood equations and the moment estimating equations provides such a possibility. Specifically, if we apply Taylor’s expansion to each component of $g - \mathbb{E}(g)$, the second order term in the expansion is $V(\hat{\theta} - \theta)$, which implies that obtaining an expression of $\hat{\theta} - \theta$ crucially depends on the inverse of $V$. Note that $V = -F''(\theta) \in L_n(m, M)$ 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 Loève (1977, p. 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

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

In all, we have the following proposition.

**Proposition 2.** Assume that $A \sim \mathbb{P}_{\theta^*}$. If $e^{\|\theta^*\|_\infty} = o(n^{1/2})$, then for any fixed $k \geq 1$, as $n \rightarrow \infty$, the vector consisting of the first $k$ elements of $S\{g - \mathbb{E}(g)\}$ 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 $\hat{\theta} - \theta$ and $S\{g - \mathbb{E}(g)\}$ (See details in the Online Supplementary Material).


Theorem 2. Assume that $A \sim \mathbb{P}_{\theta^*}$. If $\|\theta^*\|_\infty \leq \tau \log n$, 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 $(\hat{\theta} - \theta^*)$ is asymptotically multivariate normal with mean $0$ 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 $1/n^{1/2}$. Since $(n - 1)e^{-2\|\theta^*\|_\infty}/4 \leq v_{i,i} \leq (n - 1)/4$, the rate of convergence is between $O(n^{-1/2}e^{\|\theta^*\|_\infty})$ 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. In the case of continuous weights, i.e. $\Omega = [0, \infty)$, $\nu$ is the Lebesgue measure, and $a_{i,j}, 1 \leq i \neq j \leq n$ are mutually independent exponential random variables with the density:

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

and the natural parameter space is

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

In communication networks, the value of an edge may be continuous. For example, if an edge denotes the talking time between two persons in a telephone network, then its weight is continuous. To follow the tradition that the rate parameters are positive in exponential families, we take the transformation $\bar{\theta} = -\theta$, $\bar{\alpha}_i = -\alpha_i$ and $\bar{\beta}_j = -\beta_j$. The corresponding natural parameter space then becomes

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

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

$$d_i = \sum_{k=1; k \neq i}^n (\bar{\alpha}_i + \bar{\beta}_k)^{-1}, \quad i = 1, \ldots, n,$$

$$b_j = \sum_{k=1; k \neq j}^n (\bar{\alpha}_k + \bar{\beta}_j)^{-1}, \quad j = 1, \ldots, n.$$ 

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

$$F_1(\bar{\theta}) = d_i - \sum_{k \neq i} (\bar{\alpha}_i + \bar{\beta}_k)^{-1}, \quad i = 1, \ldots, n,$$

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

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

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

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

$$\frac{\partial F_i}{\partial \beta_j} = \frac{1}{(\bar{\alpha}_i + \bar{\beta}_j)^2}, \quad j = 1, \ldots, n - 1, \quad j \neq i; \quad \frac{\partial F_i}{\partial \beta_i} = 0,$$
and for $j = 1, \ldots, n - 1$,
\[
\frac{\partial F_{n+j}}{\partial \alpha_l} = \frac{1}{(\alpha_l + \beta_j)^2}, \quad l = 1, \ldots, n, \ l \neq j; \quad \frac{\partial F_{n+j}}{\partial \beta_j} = 0, \quad l = 1, \ldots, n - 1, \ l \neq j.
\]

Further, take \( \bar{q} \) and satisfies
\[
\max \{ \text{max}_{i=1,\ldots,n} |d_i - \mathbb{E}(d_i)|, \ \text{max}_{j=1,\ldots,n} |b_j - \mathbb{E}(b_j)| \} \leq \sqrt{\frac{8(n - 1) \log n}{\gamma q_n^2}}.
\]

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

Further, take \( \bar{\theta}^{(0)} = \bar{\theta}^* \) and \( D = \Omega(\bar{\theta}^*, 2r) \) in Theorem 8, i.e. an open ball \( \{ \theta : ||\theta - \bar{\theta}^*||_\infty < 2r \} \). 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
\[
\max \{ \text{max}_{i} |d_i - \mathbb{E}(d_i)|, \ \text{max}_{j} |b_j - \mathbb{E}(b_j)| \} \leq \sqrt{\frac{8(n - 1) \log n}{\gamma q_n^2}}.
\]

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

**Theorem 3.** Assume that \( \bar{\theta}^* \) satisfies \( q_n \leq \bar{\alpha}_i^* + \bar{\beta}_j^* \leq Q_n \) and \( A \sim P_{\bar{\theta}} \). If \( Q_n/q_n = o((n/\log n)^{1/3}) \), then as \( n \) goes to infinity, with probability approaching one, the MLE \( \hat{\theta} \) exists and satisfies
\[
||\hat{\theta} - \bar{\theta}^*||_\infty \leq O_p(\frac{Q_n^9 (\log n)^{1/2}}{n^{1/2} q_n^{9/2}}) = o_p(1).
\]

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 \( V = F'(\bar{\theta}^*) \in L_n(q, M) \) is the Fisher information matrix of \( \bar{\theta} \). It is not difficult to show that the third moment of the exponential random variable with rate parameter \( \lambda \) is \( 6\lambda^{-3} \). Under the assumption of \( 0 < q_n \leq \bar{\alpha}_i^* + \bar{\beta}_j^* \leq Q_n \), we have
\[
\frac{\sum_{j=1, j \neq i}^{n} \mathbb{E}(a_{i,j}^3)}{v_{i,i}^{3/2}} \leq \frac{6Q_n}{(n - 1)^{1/2}}, \quad \text{for } i = 1, \ldots, n,
\]
and
\[
\frac{\sum_{i=1, i \neq j}^{n} \mathbb{E}(a_{i,j}^3)}{v_{i,i}^{3/2}} \leq \frac{6Q_n}{(n - 1)^{1/2}}, \quad \text{for } j = 1, \ldots, n.
\]
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, \( v^{-1/2}_i (d_i - E(d_i)) \) is asymptotically standard normal if \( Q_n/q_n = o(n^{1/2}) \). Similarly, \( v^{-1/2}_{n+j, n+j} (b_j - E(b_j)) \) is also asymptotically standard normal under the same condition. Noting that \( |S(g - E(g))|_i = v^{-1}_i (g_i - E(g_i)) + v^{-1}_{2n, 2n} (b_n - E(b_n)) \), 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 \( S(g - E(g)) \) 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 \( \hat{\theta} - \bar{\theta}^* \) and \( S(g - E(g)) \), we have the central limit theorem for the MLE \( \hat{\theta} \).

**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 \( \hat{\theta} - \bar{\theta}^* \) 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 \( 1/v_i^{1/2} \). Since \( (n - 1)/Q_n^2 \leq v_i \leq (n - 1)/q_n^2 \), 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, i.e. \( \Omega = \mathbb{N}_0 \), \( \nu \) 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:

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

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

\[
\bar{\Theta} = \{ \bar{\theta} : \bar{\alpha}_i + \bar{\beta}_j > 0 \}.
\]

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

\[
d_i = \sum_{k \neq i} e^{-(\bar{\alpha}_i + \bar{\beta}_j)} e^{(\bar{\alpha}_i + \bar{\beta}_j)} = \frac{1}{e^{(\bar{\alpha}_i + \bar{\beta}_j)} - 1}, \quad i = 1, \ldots, n,
\]

\[
b_j = \sum_{k \neq j} e^{-(\bar{\alpha}_i + \bar{\beta}_j)} e^{(\bar{\alpha}_i + \bar{\beta}_j)} = \frac{1}{e^{(\bar{\alpha}_i + \bar{\beta}_j)} - 1}, \quad j = 1, \ldots, n - 1.
\]

We first establish the existence and consistency of \( \hat{\theta} \) by applying Theorem 8. Define a system of functions:

\[
F_i(\bar{\theta}) = d_i - \sum_{k \neq i} \frac{1}{e^{(\bar{\alpha}_i + \bar{\beta}_j)} - 1}, \quad i = 1, \ldots, n,
\]

\[
F_{n+j}(\bar{\theta}) = b_j - \sum_{k \neq j} \frac{1}{e^{(\bar{\alpha}_i + \bar{\beta}_j)} - 1}, \quad j = 1, \ldots, n,
\]

\[
F(\bar{\theta}) = (F_1(\bar{\theta}), \ldots, F_{2n-1}(\bar{\theta}))^\top.
\]

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

\[
\frac{\partial F_i}{\partial \bar{\alpha}_l} = 0, \quad l = 1, \ldots, n, \quad l \neq i; \quad \frac{\partial F_i}{\partial \bar{\alpha}_i} = \sum_{k=1, k \neq i}^{n} \frac{e^{(\bar{\alpha}_i + \bar{\beta}_k)} - 1}{[e^{(\bar{\alpha}_i + \bar{\beta}_k)} - 1]^2}.
\]
\[
\frac{\partial F_i}{\partial \beta_j} = \frac{e^{(\alpha_i+\beta_j)} - 1}{[e^{(\alpha_i+\beta_j)} - 1]^2}, \quad j = 1, \ldots, n - 1, j \neq i; \quad \frac{\partial F_i}{\partial \alpha_j} = 0,
\]
and for \( j = 1, \ldots, n - 1, \)
\[
\frac{\partial F_{n+j}}{\partial \alpha_l} = \frac{e^{(\alpha_k+\beta_j)} - 1}{[e^{(\alpha_k+\beta_j)} - 1]^2}, \quad l = 1, \ldots, n, l \neq j; \quad \frac{\partial F_{n+j}}{\partial \alpha_j} = 0,
\]
\[
\frac{\partial F_{n+j}}{\partial \beta_j} = \sum_{k \neq j} \frac{e^{(\alpha_k+\beta_j)} - 1}{[e^{(\alpha_k+\beta_j)} - 1]^2}, \quad \frac{\partial F_{n+j}}{\partial \beta_l} = 0, \quad l = 1, \ldots, n - 1, l \neq j.
\]

Let \( \hat{\theta}^* \) be the true parameter vector. It is not difficult to see \( F'(\hat{\theta}^*) \in L_n(m, M) \) so that Proposition 1 can be applied. The constants in the upper bounds of Theorem 8 are given in the following lemma.

**Lemma 6.** Assume that \( \hat{\theta}^* \) satisfies \( q_n \leq \hat{\alpha}_i + \hat{\beta}_j \leq Q_n \) for all \( i \neq j, A \sim \mathbb{P}_{\hat{\theta}^*} \) and

\[
\max_{i=1,\ldots,n} |d_i - \mathbb{E}(d_i)|, \quad \max_{j=1,\ldots,n} |b_j - \mathbb{E}(b_j)| \leq \sqrt{\frac{8(n-1)\log n}{\gamma q_n^2}}.
\]

Then we have
\[
r = ||[F'(\hat{\theta}^*)]^{-1}F(\hat{\theta}^*)||_{\infty} \leq O \left( q_n^{-1}(e^{3Q_n}(1+q_n^{-4}) + e^{Q_n}) \sqrt{\frac{\log n}{n}} \right).
\]

Further, take \( \bar{\theta}^{(0)} = \hat{\theta}^* \) and \( D = \Omega(\hat{\theta}^*; 2r) \) in Theorem 8, i.e. an open ball \( \{ \theta : \| \theta - \hat{\theta}^* \|_\infty < 2r \} \).

If \( q_n - 4r > 0 \) then we can choose \( K_1 = 2(n-1)e^{q_n-4r}(1 + e^{q_n-4r})e^{q_n-4r} - 1 \) and \( K_2 = (n-1)e^{q_n-4r}(1 + e^{q_n-4r})e^{q_n-4r} - 1 \).

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
\[
\max_{i} |d_i - \mathbb{E}(d_i)|, \quad \max_{j} |b_j - \mathbb{E}(b_j)| \leq \sqrt{\frac{8(n-1)\log n}{\gamma q_n^2}}.
\]

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

**Theorem 5.** Assume that \( \hat{\theta}^* \) satisfies \( q_n \leq \hat{\alpha}_i + \hat{\beta}_j \leq Q_n \) for all \( i \neq j, A \sim \mathbb{P}_{\hat{\theta}^*} \). If \( (1 + q_n^{-1})e^{6Q_n} = o(n^{1/2}/(\log n)^{1/2}) \) then as \( n \) goes to infinity, with probability approaching one, the MLE \( \hat{\theta} \) exists and satisfies
\[
\| \hat{\theta} - \hat{\theta}^* \|_{\infty} \leq O_p \left( e^{3Q_n}(1 + \frac{1}{q_n^2}) \sqrt{\frac{\log n}{n}} \right) = o_p(1).
\]

Further, if the MLE exists, it is unique.

Note that both \( d_i = \sum_{j \neq i} a_{i,j} \) and \( b_j = \sum_{i \neq j} a_{i,j} \) are sums of \( n - 1 \) independent geometric random variables. Also note that \( q_n \leq \hat{\alpha}_i + \hat{\beta}_j \leq Q_n \) and \( V = F'(\hat{\theta}^*) \in L_n(m, M) \), thus we have
\[
\frac{e^{Q_n}}{(e^{Q_n} - 1)^2} \leq v_{i,j} \leq \frac{e^{q_n}}{(e^{q_n} - 1)^2}, \quad i = 1, \ldots, n, j = n+1, \ldots, 2n, j \neq n+i,
\]
\[
\frac{(n - 1)e^{Q_n}}{(e^{Q_n} - 1)^2} \leq v_{i,i} \leq \frac{(n - 1)e^{q_n}}{(e^{q_n} - 1)^2}, \quad i = 1, \ldots, 2n.
\]

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 \( p_{i,j} = 1 - e^{-(\alpha_i^* + \beta_j^*)} \). By simple calculations, we also have

\[
\mathbb{E}(a_{i,j}^3) = v_{i,j} (6 + \frac{e^{\alpha_i^* + \beta_j^*} - 1}{e^{\alpha_i^* + \beta_j^*}} + \frac{6}{e^{\alpha_i^* + \beta_j^*} - 1}).
\]

It then follows

\[
\frac{\sum_{j \neq i} \mathbb{E}(a_{i,j}^3)}{v_{i,i}^{5/2}} \leq \frac{7 + 6(e^{q_n} - 1)^{-1}}{v_{i,i}^{1/2}} \leq \frac{[7 + 6(e^{q_n} - 1)^{-1}]e^{Q_n} - 1}{n^{1/2}e^{Q_n}/2}.
\]

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 \( S\{g - \mathbb{E}(g)\} \) 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 \( \hat{\theta} \) is stated as follows.

**Theorem 6.** If \( e^{Q_n}(1 + q_n^{-1/2}) = o(n^{1/2}/\log n) \), then for any fixed \( k \geq 1 \), as \( n \to \infty \), the vector consisting of the first \( k \) elements of \( \theta - \hat{\theta} \) 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 the work of [19, 20], Haberman (1977) proved the 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. Haberman (1981, page 60) indicated “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 \( X_{ii} \) unobserved”. Therefore he claimed that the method of his 1977 paper can be extended to derive the consistency and asymptotic normality of the MLE in the \( p_1 \) model without reciprocity, but he didn’t give a formal proof. These conclusions are premature. There are several reasons. First, in the item response experiments, the total \( g \) persons give answers (0 or 1) to the total \( h \) items. The experiments outcomes construct a bipartite directed graph since there are no responses between any two persons or two items while there exists edges in every pair of vertices of the directed graph considered in the present paper. Second, each vertex in the Rasch model is only assigned one parameter measuring the out-degree effect for persons or the in-degree effect for
items while in the model (1.1), there exists these two parameters for each vertex simultaneously. Therefore, the model (1.1) can not be viewed as the equivalent Rasch model. Conversely, the Rasch model can be considered as a special case of the model (1.1) for the bipartite directed random graph. We also noted that Fischer [17] pointed out that the Rasch model can be considered as the Bradley-Terry model [7] for incomplete paired comparisons. [41] proved the uniform consistency and asymptotic normality of the MLE in the Bradley-Terry model with a diverging number of parameters. Third, in contrast to Haberman’s (1977) proofs, our methods mainly utilize the approximate inverse of the Fisher information matrix and don’t require the assumption of the constant upper bound on all the parameters, while his methods are based on the classical exponential family theory of [4, 5]. Therefore, we concluded that Haberman’s (1977) work cannot be extended into the model (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 \( \alpha_0^* = \alpha^* = 1 \) for \( i, j \); for the case with discrete weights, we set \( \alpha_0^* = \alpha^* = 0.2 \) for \( i, j \). In both cases, we considered four different values for \( L = 0, \log(\log n), (\log n)^{1/2} \) and \( \log n \). For the case with continuous weights, we set \( \alpha_0^* = 1 + (n - 1 - i) L / (n - 1) \) for \( i, j \). The results for \( \hat{\theta} \) and \( \hat{\beta} \) for all cases are similar, thus only the results of \( \hat{\theta} \), \( \hat{\theta}_n \) and \( \hat{\beta}_n \) are considered: \( L = 0, \log(\log(n)), \log(n) \) and \( n^{1/2} \). For the parameter values \( n \), let \( \hat{\beta}_1^* = \alpha^*_i, i = 1, \ldots, n - 1 \) for simplicity and \( \beta_n^* = 0 \) by default.

Note that by Theorems 2, 4 and 6, \( \hat{\xi}_{i,j} = [\hat{\alpha}_i - \hat{\alpha}_j - (\hat{\alpha}_i^* - \hat{\alpha}_j^*)] / (1/\hat{v}_{i,j} + 1/\hat{v}_{j,i})^{1/2}, \) (1.1) for \( i, j; \hat{\xi}_{i,j} = [\hat{\alpha}_i + \hat{\beta}_j - \hat{\alpha}_j^* - \hat{\beta}_j^*] / (1/\hat{v}_{i,j} + 1/\hat{v}_{j,i}+\hat{v}_{n,j,n,j})^{1/2}, \) and \( \hat{\eta}_{i,j} = [\hat{\beta}_i - \hat{\beta}_j - (\hat{\beta}_i^* - \hat{\beta}_j^*)] / (1/\hat{v}_{n+i,n+i}+1/\hat{v}_{n+i,n+i+j})^{1/2} \) are all asymptotically distributed as standard normal random variables, where \( \hat{v}_{i,j} \) is the estimate of \( v_{i,j} \) by replacing \( \hat{\theta}^* \) with \( \hat{\theta} \). Therefore, we assess the asymptotic normality of \( \hat{\xi}_{i,j}, \hat{\xi}_{i,j} \) 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 \( \hat{\xi}_{i,j}, \hat{\xi}_{i,j} \) 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 simulate two values 100 and 200 for \( n \) and find that the QQ-plots for them are similar. Therefore, we only shows the QQ-plots under \( n = 200 \) 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 for continuous weights, when \( L = \log n, n^{1/2} \), the empirical quantiles coincide with the theoretical ones very well (The QQ plots when \( L = 0, \log(\log n) \) 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 \( L = (\log n)^{1/2} \). 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 agrees 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 pair (1, 2) are higher than the nominal level; when \( L = \log n \), the MLE did not exist for all 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 for a class of weighted edges including binary, continuous and infinite discrete
weights when 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 infinite 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:

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

and the likelihood equations become:

$$d_i = \sum_j \frac{1}{1 - e^{-(\alpha_i + \beta_j)q}} \sum_k^{q-1} e^{-k(\alpha_i + \beta_j)},$$

$$b_j = \sum_i \left[ \frac{1}{e^{\alpha_i + \beta_j} - 1} - \frac{q}{e^{\alpha_i + \beta_j} - 1} \right].$$

It can be shown that the Fisher information matrix of $\theta$ is also in the class of matrices $L_n(m, M)$ under certain conditions. Therefore, except for some more complex calculations in contrast with the binary case, there are 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 expected. Further, our general results can also be applied to the Poisson model where $a_{i,j}$ follows:

$$P(a_{i,j} = a) = \frac{(\alpha_i + \beta_j)^a}{a!} e^{-(\alpha_i + \beta_j)}, \quad a \in \mathbb{N}_0, \quad \alpha_i + \beta_j > 0,$$

in which the Fisher information matrix is a constant matrix.

It is also worth to note that the conditions imposed on $q_n$ and $Q_n$ may not be the best possible. In particular, the conditions guaranteeing the asymptotic normality seem stronger than that 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.

Finally, we discuss some issues related to the original Holland and Leinhardt’s $p_1$ model. In their model, there are two other parameters ($\rho$ and $\lambda$) that measure the strength of the reciprocal effect and the density of edges, and the sufficient statistic of the density parameter $\lambda$ is a linear combination of the in-degrees of vertices or the out-degrees of vertices. Therefore, 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 $\sum_i (\alpha_i + \lambda + \beta_n) d_i + \sum_j (\beta_j - \beta_n) b_j$. By making 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) when there is no reciprocal parameter $\rho$. Therefore, there is no difference in the effect of model fitting with and without the density parameter. If only the reciprocal parameter is incorporated into (1.1), the simulation results in [46] indicate that the MLEs still enjoy the properties of the uniform consistency and asymptotic normality, in which the asymptotic variances of the MLEs are the correspondingly diagonal elements of the Fisher information matrix. Since the Fisher information matrix for the parameter vector $(\rho, \alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_{n-1})$ is the extension of that for $(\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n)$ with only a new row and a column, by the Sherman-Morrison formula, we can also obtain an approximate inverse but with a very complex form. In order to see whether there are simple forms of asymptotic variances similar to the diagonal elements of the approximate matrix $S$, we did some additional simulations and the results show that the asymptotic variances for the MLE $\hat{\alpha}_i, \hat{\beta}_j$ are still those defined in this paper and that of $\hat{\rho}$ is $1/\text{var}(\sum_{i \neq j} a_{i,j} a_{i,j})$. It would be of interest to investigate whether this conclusion holds and to see whether the current methods of proofs can be expanded.
discussed how to test the fit of the $p_1$ model. Since the likelihood ratio test for comparing the fit of two models crucially depends on the properties of the MLE, the asymptotic behavior of the MLE has an impact on the asymptotic distribution of the likelihood ratio test. In view of that the MLE keeps good asymptotic properties in model (1.1), we believe the conjectures on the asymptotic distribution of the likelihood ratio test in these references for testing the fit of $p_1$ model are reasonable.

5. Proofs of lemmas and theorems. In this section we give proofs for the lemmas and theorems presented in Section 2.

5.1. Preliminaries. We first present the Hoeffding’s inequality for bounded independent random variables, the concentration inequality for independent sub-exponential random variables and 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.

5.1.1. Hoeffding’s inequality. 

THEOREM 7 (Hoeffding’s inequality). Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in [a_i, b_i]$ almost surely. Hoeffding (1963) proved

$$P(\sum_{i}(X_i - E(X_i)) \geq x) \leq 2 \exp(-\frac{2x^2}{\sum_{i}(b_i - a_i)^2}).$$

5.1.2. Uniqueness of the MLE. Let $\sigma_{\Omega}$ be a $\sigma$-algebra over the set of weight values $\Omega$ and $\nu$ be a canonical $\sigma$-finite probability measure on $(\Omega, \sigma_{\Omega})$. In this paper, $\nu$ is the Lebesgue measure in the case of continuous weight and the counting measure in the case of discrete weight. Denote $\nu_n$ by the product measure on $\Omega_n$. Let $P$ be all the probability distributions on $\Omega$ that are absolutely continuous with respective to $\nu_n$. Define the mean parameter space $M$ to be the set of expected degree vectors tied to $\theta$ from all distributions $P \in P$:

$$M = \{E_P g : P \in P\}.$$ 

Since a convex combination of probability distributions in $\Psi$ is also a probability distribution in $\Psi$, the set $M$ 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 $\Theta$ is open, then $\Psi$ is regular. By the general theory for a regular and minimal exponential family distribution (Theorem 3.3 of Wainwright and Jordan, 2008), the gradient of the log-partition function maps the natural parameter space $\Theta$ to the interior of the mean parameter space $M$, and this mapping

$$\nabla Z : \Theta \rightarrow M^\circ$$

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

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

5.1.3. Newton iterative theorem. Let $D$ be an open convex subset of $\mathbb{R}^{2n-1}$, $\Omega(x, r)$ denote the open ball $\{y \in \mathbb{R}^{2n-1} : \|x - y\|_\infty < r\}$ and $\overline{\Omega(x, r)}$ be its closure, where $x \in \mathbb{R}^{2n-1}$. 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 [24,
32, 33, 42, 35]. 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.

**Theorem 8.** Define a system of equations:

\[
F_i(\theta) = d_i - \sum_{k=1, k \neq i}^n f(\alpha_i + \beta_k), \quad i = 1, \ldots, n,
\]

\[
F_n + (\theta) = b_j - \sum_{k=1, k \neq j}^n f(\alpha_k + \beta_j), \quad j = 1, \ldots, n - 1,
\]

\[
F'(\theta) = (F_1(\theta), \ldots, F_n(\theta), F_{n+1}(\theta), \ldots, F_{2n-1}(\theta))^T,
\]

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 \( x, y, v \in D \), we have

\[
\|F'(x) - F'(y)\|_\infty \leq K_1 \|x - y\|_\infty \|v\|_\infty,
\]

\[
\max_{i=1, \ldots, 2n-1} \|F'_i(x) - F'_i(y)\|_\infty \leq K_2 \|x - y\|_\infty,
\]

where \( F'(\theta) \) is the Jacobian matrix of \( F \) on \( \theta \) and \( F'_i(\theta) \) is the gradient function of \( F_i \) on \( \theta \). Consider \( \theta_0(0) \in D \) with \( \Omega(\theta_0(0), 2r) \subset D \), where \( r = \|[F'(\theta_0(0))]^{-1} F(\theta_0(0))\|_\infty \). For any \( \theta \in \Omega(\theta_0(0), 2r) \), we assume

\[
F'(\theta) \in \mathcal{L}_n(m, M) \quad \text{or} \quad -F'(\theta) \in \mathcal{L}_n(m, M).
\]

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

\[
\rho = \frac{c_1(2n-1)M^2K_1}{2m^3n^2} + \frac{K_2}{(n-1)m}. \tag{5.4}
\]

If \( \rho r < 1/2 \), then \( \theta^{(k)} \in \Omega(\theta_0(0), 2r), k = 1, 2, \ldots \), are well-defined and satisfy

\[
\|\theta^{(k+1)} - \theta^{(0)}\|_\infty \leq r/(1 - \rho r). \tag{5.5}
\]

Further, \( \lim_{k \to \infty} \theta^{(k)} \) exists and the limiting point is precisely the solution of \( F(\theta) = 0 \) in the range of \( \theta \in \Omega(\theta_0(0), 2r) \).

**Proof.** We prove \( \theta^{(k)} \in \Omega(\theta_0(0), 2r) \) by induction on \( k \). Note that

\[
\|\theta^{(1)} - \theta^{(0)}\|_\infty = \|[F'(\theta^{(0)})]^{-1} F(\theta^{(0)})\|_\infty = r,
\]

thus the result holds for the base case \( k = 1 \). Now assume that for \( i = 1, \ldots, k \), we have \( \theta^{(i)} \in \Omega(\theta_0(0), 2r) \). According to the definition, \( \theta^{(k+1)} \) is well-defined. By (5.3), Proposition 1 can be applied to approximate \( [F'(\theta^{(k)})]^{-1} \). Let \( [F'(\theta^{(k)})]^{-1} = W + S \), then

\[
-(\theta^{(k+1)} - \theta^{(k)}) = [F'(\theta^{(k)})]^{-1} F(\theta^{(k)}) = WF(\theta^{(k)}) + SF(\theta^{(k)}). \tag{5.6}
\]

Without loss of generality, we assume \( F'(\theta) \in \mathcal{L}_n(m, M) \). (In the case of \( -F'(\theta) \in \mathcal{L}_n(m, M) \), the proof is similar.) Note that \( WF(\theta^{(k+1)}) + F'(\theta^{(k+1)}) (\theta^{(k)} - \theta^{(k-1)}) = 0 \). By the mean value theorem, Proposition 1 and (5.1), we have

\[
\|WF(\theta^{(k)})\|_\infty \leq \|W\|_\infty \|F'(\theta^{(k)}) - F'(\theta^{(k-1)})\|_\infty \|\theta^{(k)} - \theta^{(k-1)}\|_\infty \leq \frac{c_1(2n-1)M^2}{m^3n^2} \times \|F'(\theta^{(k)} + (1-t)\theta^{(k-1)}) - F'(\theta^{(k-1)})\|_\infty \|\theta^{(k)} - \theta^{(k-1)}\|_\infty \leq \frac{c_1(2n-1)M^2}{m^3n^2} \times \frac{K_1}{2} \|\theta^{(k)} - \theta^{(k-1)}\|_\infty^2. \tag{5.7}
\]
Note that $F_{2n}(\theta) = \sum_{i=1}^{n} F_i(\theta) - \sum_{j=1}^{n-1} F_{n+j}(\theta)$. Therefore, we have

$$
0 = (1_{n}^{\top}, -1_{n-1}^{\top}) F(\theta^{(k-1)}) + (1_{n}^{\top}, -1_{n-1}^{\top}) F'(\theta^{(k-1)})(\theta^{(k)} - \theta^{(k-1)})
$$

$$
= F_{2n}(\theta^{(k-1)}) + F_{2n}'(\theta^{(k-1)})(\theta^{(k)} - \theta^{(k-1)}).
$$

By (5.2), it follows that

$$
\|F_{2n}(\theta^{(k)})\|_{\infty} = \|F_{2n}(\theta^{(k)}) - F_{2n}(\theta^{(k-1)}) - F_{2n}'(\theta^{(k-1)})(\theta^{(k)} - \theta^{(k-1)})\|_{\infty}
$$

$$
= \| \int_{0}^{1} [F_{2n}'(t\theta^{(k)} + (1-t)\theta^{(k-1)}) - F_{2n}'(\theta^{(k-1)})](\theta^{(k)} - \theta^{(k-1)}) dt\|_{\infty}
$$

$$
\leq \frac{K_2}{2} \|\theta^{(k)} - \theta^{(k-1)}\|_{\infty}^2.
$$

Similarly, for $i = 1, \ldots, 2n - 1$, we have

$$
\|F_i(\theta^{(k)})\|_{\infty} \leq \frac{K_2}{2} \|\theta^{(k)} - \theta^{(k-1)}\|_{\infty}^2.
$$

Let $V$ denote $F'(\theta)$. Since $V \in L_n(m, M)$, $(n-1)m \leq v_{i,i} \leq (n-1)M$ for $i = 1, \ldots, 2n - 1$. Thus,

$$
(5.8) \quad \|SF(\theta^{(k)})\|_{\infty} \leq \max_{i=1, \ldots, 2n-1} \frac{|F_i(\theta^{(k)})|}{v_{i,i}} + \frac{|F_{2n}(\theta^{(k)})|}{v_{2n,2n}} \leq \frac{K_2}{(n-1)m} \|\theta^{(k)} - \theta^{(k-1)}\|_{\infty}^2.
$$

In view of (5.6), (5.7) and (5.8), we have

$$
(5.9) \quad \|\theta^{(k+1)} - \theta^{(k)}\|_{\infty} \leq \rho \|\theta^{(k)} - \theta^{(k-1)}\|_{\infty}^2,
$$

where $\rho$ is defined in (5.4). By repeating the above process, we will obtain:

$$
\|\theta^{(k+1)} - \theta^{(k)}\|_{\infty} \leq \rho \|\theta^{(k)} - \theta^{(k-1)}\|_{\infty}^2 \leq \rho^{1+2} \|\theta^{(k-1)} - \theta^{(k-2)}\|_{\infty}^4
$$

$$
\leq \cdots
$$

$$
\leq \rho^{1+\cdots+2k} \|\theta^{(1)} - \theta^{(0)}\|_{\infty}^{2k+1} = r(\rho r)^{2k+1-1}.
$$

Consequently, if $\rho r < 1/2$, then

$$
\|\theta^{(k+1)} - \theta^{(0)}\|_{\infty} \leq \sum_{i=0}^{k} \|\theta^{(i+1)} - \theta^{(i)}\|_{\infty}
$$

$$
\leq \sum_{i=0}^{k} r(\rho r)^{2i-1} \leq \sum_{i=0}^{k} r(\rho r)^{i}
$$

$$
= \frac{r(1 - (\rho r)^{k+1})}{1 - \rho r} \leq 2r.
$$

This shows $\theta^{(k+1)} \in \Omega(\theta^{(0)}, 2r)$ and (5.5) holds.

By (5.9), $\{\theta^{(k+1)}\}_{k=0}^{\infty}$ is a Cauchy sequence. Thus, $\hat{\theta} := \lim_{k \to \infty} \theta^{(k)}$ exists and lies in $\Omega(\theta^{(0)}, 2r)$. The result that $\hat{\theta}$ is a solution follows in the usual way from

$$
\|F(\theta^{(k)})\|_{\infty} = \|F'(\theta^{(k)})(\theta^{(k+1)} - \theta^{(k)})\|_{\infty}
$$

$$
\leq \|F'(\theta^{(0)})\|_{\infty} + \|F'(\theta^{(0)}) - F'(\theta^{(k)})\|_{\infty} \|\theta^{(k)} - \theta^{(k+1)}\|_{\infty} \to 0.
$$

$\square$
5.2. Proofs in the case of binary weights.

5.2.1. Proof of Lemma 2. For fixed $n$, we first derive $K_1$ and $K_2$ in the inequalities (5.1) and (5.2), respectively. Let $x, y \in \mathbb{R}^{2n-1}$ and

$$F'_i(\theta) = (F'_{i,1}(\theta), \ldots, F'_{i,2n-1}(\theta)) := \left( \frac{\partial F_i}{\partial \alpha_1}, \ldots, \frac{\partial F_i}{\partial \alpha_n}, \frac{\partial F_i}{\partial \beta_1}, \ldots, \frac{\partial F_i}{\partial \beta_{n-1}} \right).$$

Then, for $i = 1, \ldots, n$, we have

$$\frac{\partial^2 F_i}{\partial \alpha_s \partial \alpha_l} = 0, \ s \neq l; \ \ \frac{\partial^2 F_i}{\partial \alpha_i^2} = -\sum_{k \neq i} \frac{e^{\alpha_i + \beta_k} (1 - e^{\alpha_i + \beta_k})}{(1 + e^{\alpha_i + \beta_k})^3},$$

$$\frac{\partial^2 F_i}{\partial \beta_s \partial \alpha_i} = \frac{-e^{\alpha_i + \beta_s} (1 - e^{\alpha_i + \beta_s})}{(1 + e^{\alpha_i + \beta_s})^3}, \ s = 1, \ldots, n - 1, \ s \neq i; \ \ \frac{\partial^2 F_i}{\partial \beta_i \partial \alpha_i} = 0,$$

$$\frac{\partial^2 F_i}{\partial \beta_j^2} = \frac{-e^{\alpha_i + \beta_j} (1 - e^{\alpha_i + \beta_j})}{(1 + e^{\alpha_i + \beta_j})^3}, \ j = 1, \ldots, n - 1; \ \ \frac{\partial^2 F_i}{\partial \beta_s \partial \beta_l} = 0, \ s \neq l.$$

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

$$F'_i(x) - F'_i(y) = J^{(i)}(x - y),$$

where

$$J^{(i)}_{s,l} = \int_0^1 \frac{\partial F'_i}{\partial \theta_1} ((x + (1 - t)y)dt, \ s, l = 1, \ldots, 2n - 1.$$

Note that

$$\left| \frac{e^{\alpha_i + \beta_j} (1 - e^{\alpha_i + \beta_j})}{(1 + e^{\alpha_i + \beta_j})^3} \right| \leq \frac{e^{\alpha_i + \beta_j}}{(1 + e^{\alpha_i + \beta_j})^2} \leq \frac{1}{4}.$$

Therefore,

$$\max_s \sum_{l=1}^{2n-1} |J^{(i)}| \leq (n - 1)/2, \ \ \sum_{s,l} |J^{(i)}_{s,l}| \leq n - 1.$$

Similarly, for $i = n+1, \ldots, 2n - 1$, we also have $F'_i(x) - F'_i(y) = J^{(i)}(x - y)$ and $\sum_{s,l} |J^{(i)}_{s,l}| \leq n - 1$. Consequently,

$$\|F'_i(x) - F'_i(y)\|_{\infty} = \|J^{(i)}\|_{\infty} \|x - y\|_{\infty} \leq (n - 1)/2, \ i = 1, \ldots, 2n - 1,$$

and for any vector $v \in \mathbb{R}^{2n-1},$

$$\left\| [F(x) - F(y)]v \right\|_{\infty} = \max_i \left| \sum_{j=1}^{2n-1} (F'_{i,j}(x) - F'_{i,j}(y))v_j \right|$$

$$= \max_i \left| (x - y)J^{(i)}v \right|$$

$$\leq \|x - y\|_{\infty} \|v\|_{\infty} \sum_{k,l} \left| J^{(i)}_{k,l} \right| \leq (n - 1)\|x - y\|_{\infty} \|v\|_{\infty}$$

Therefore, we could choose $K_1 = n - 1$ and $K_2 = (n - 1)/2$.

Clearly, $-F'(\theta^*) \in \mathcal{L}_{2n-1}(m_*, M_*)$, where

$$M_* = 1/4, \ m_* = \frac{e^{2\|\theta^*\|_{\infty}}}{(1 + e^{2\|\theta^*\|_{\infty}})^2}.$$
Note that 
\[ F(\theta^*) = (d_1 - \mathbb{E}(d_1), \ldots, d_n - \mathbb{E}(d_n), b_1 - \mathbb{E}(b_1), \ldots, b_{n-1} - \mathbb{E}(b_{n-1})). \]

By the assumption of (2.3) and Lemma 1, we have
\[
 r = \| [F'(\theta^*)]^{-1} F(\theta^*) \|_\infty \leq \frac{2c_1(2n - 1)M^2}{m^2(n - 1)^2} + \max_{i = 1, \ldots, 2n-1} \frac{|F_i(\theta^*)| + |F_{2n}(\theta^*)|}{v_{i,n}} \\
\leq (n \log n)^{1/2} \left( \frac{c_{11} e^{6\|\theta^*\|_\infty}}{n} + \frac{c_{12} e^{2\|\theta^*\|_\infty}}{n} \right),
\]
where \( c_{11} \) and \( c_{12} \) are constants.

5.2.2. Proof of Lemma 3. By Hoeffding’s inequality, we have
\[
P(|d_i - \mathbb{E}(d_i)| \geq \sqrt{(n - 1) \log(n - 1)}) \leq 2 \exp\{-2(n - 1) \log(n - 1)/(n - 1)\} = \frac{2}{(n - 1)^2}.
\]
Therefore,
\[
P(\max_i |d_i - \mathbb{E}(d_i)| \geq \sqrt{(n - 1) \log(n - 1)}) \leq n \times \frac{2}{(n - 1)^2}.
\]
Similarly, we have
\[
P(\max_i |b_i - \mathbb{E}(b_i)| \geq \sqrt{(n - 1) \log(n - 1)}) \leq n \times \frac{2}{(n - 1)^2}.
\]
Consequently,
\[
P(\max_i |d_i - \mathbb{E}(d_i)|, \max_j |b_j - \mathbb{E}(b_j)| \geq \sqrt{(n - 1) \log(n - 1)}) \\
\leq P(\max_i |d_i - \mathbb{E}(d_i)| \geq \sqrt{(n - 1) \log(n - 1)}) + P(\max_j |b_j - \mathbb{E}(b_j)| \geq \sqrt{(n - 1) \log(n - 1)}) \\
\leq \frac{4n}{(n - 1)^2}.
\]
This is equivalent to Lemma 3.

5.2.3. Proof of Theorem 1. Assume that condition (2.3) holds. Recall the Newton’s iterates \( \theta^{(k+1)} = \theta^{(k)} - [F'(\theta^{(k)})]^{-1} F(\theta^{(k)}) \) with \( \theta^{(0)} = \theta^* \). If \( \theta \in \Omega(\theta^*, 2r) \), then \(-F'(\theta) \in \mathcal{L}_n(m, M)\) with
\[
M = \frac{1}{4}, \quad m = \frac{e^{2\|\theta^*\|_\infty+2r}}{(1 + e^{2\|\theta^*\|_\infty+2r})^2}.
\]
If \( \|\theta^*\|_\infty \leq \tau \log n \) with the constant \( \tau \) satisfying \( 0 < \tau < 1/16 \), then as \( n \to \infty \), \( n^{-1/2} (\log n)^{1/2} e^{8\|\theta^*\|} \leq n^{-1/2+8\tau} (\log n)^{1/2} \to 0 \). By Lemma 2 and condition (2.3), for sufficiently small \( r \),
\[
\rho r \leq \left[ \frac{c_1(2n - 1)M^2(n - 1)}{2m^3n^2} + \frac{(n - 1)}{2m(n - 1)} \right] \times \left( \frac{(\log n)^{1/2}}{n^{1/2}} \right) \left( \frac{c_{11} e^{6\|\theta^*\|_\infty}}{n^{1/2}} + \frac{c_{12} e^{2\|\theta^*\|_\infty}}{n^{1/2}} \right) \\
\leq \frac{O\left( (\log n)^{1/2} e^{12\|\theta^*\|_\infty} \times \frac{n^{1/2}}{n^{1/2}} \right)}{n^{1/2}} + \frac{O\left( (\log n)^{1/2} e^{8\|\theta^*\|_\infty} \times \frac{n^{1/2}}{n^{1/2}} \right)}{n^{1/2}}.
\]
Therefore, if \( \|\theta^*\|_\infty \leq \tau \log n \), then \( \rho r \to 0 \) as \( n \to \infty \). Consequently, by Theorem 8, \( \lim_{n \to \infty} \hat{\theta}^{(n)} \) exists. Denote the limit as \( \hat{\theta} \), then it satisfies
\[
\|\hat{\theta} - \theta^*\|_\infty \leq 2r = O\left( \frac{\log n)^{1/2} e^{8\|\theta^*\|_\infty}}{n^{1/2}} \right) = o(1).
\]
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.
5.2.4. Proof of Theorem 2. Before proving Theorem 2, we first establish two lemmas.

**Lemma 8.** Let $R = V^{-1} - S$ and $U = \text{Cov}(R[g - \mathbb{E}g])$. Then

$$
||U|| \leq ||V^{-1} - S|| + \frac{(1 + e^{2||\theta^*||})^4}{4e^4||\theta^*||(n - 1)^2}.
$$

**Proof.** Note that

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

where $I$ is a $(2n - 1) \times (2n - 1)$ diagonal matrix, and by (??), we have

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

Thus,

$$||U|| \leq ||V^{-1} - S|| + ||S(I_{2n-1} - VS)|| \leq ||V^{-1} - S|| + \frac{3(1 + e^{2||\theta^*||})^4}{4e^4||\theta^*||(n - 1)^2}.$$

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

$$
\hat{\theta}_i - \theta^*_i = [V^{-1}\{g - \mathbb{E}(g)\}]_i + o_p(n^{-1/2}).
$$

**Proof.** By Theorem 1, we have

$$
\hat{\mu}_n := \max_{1 \leq i \leq 2n - 1} |\hat{\theta}_i - \theta^*_i| = O_p(\frac{(\log n)^{1/2}e^{8||\theta^*||}}{n^{1/2}}).
$$

Let $\tilde{\gamma}_{i,j} = \tilde{\alpha}_i + \tilde{\beta}_j - \alpha_i - \beta_j$. By Taylor expansion, for any $1 \leq i \neq j \leq n$,

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

where

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

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

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

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

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

Equivalently,

$$\hat{\theta} - \theta^* = V^{-1}(g - \mathbb{E}(g)) + V^{-1}h.$$
Since $|e^x(1 - e^x)/(1 + e^x)^3| \leq 1$, we have
\[
|h_{i,j}| \leq |s_{i,j}^2|/2 \leq 2\rho_{n}^2, \quad |h_i| \leq \sum_{j \neq i} |h_{i,j}| \leq 2(n - 1)\rho_{n}^2.
\]

Note that $(Sh)_i = h_i/v_i + (-1)^{i(i^+)h_{2n}/v_{2n,2n}}$, and $(V^{-1}h)_i = (Sh)_i + (Rh)_i$. By direct calculations, we have
\[
|(Sh)_i| \leq \frac{|h_i| + |h_{2n}|}{v_{i,i}} \leq \frac{16\rho_{n}^2(1 + e^2\|\theta^*\|_\infty)^2}{e^{22\|\theta^*\|_\infty} \log n} \leq O\left(\frac{e^{22\|\theta^*\|_\infty} \log n}{n}\right),
\]
and by Proposition 1, we have
\[
|(Rh)_i| \leq \|R\|_\infty \times (2n - 1) \max_i |h_i| \leq O\left(\frac{e^{22\|\theta^*\|_\infty} \log n}{n}\right).
\]
If $\|\theta^*\|_\infty \leq \tau \log n$ and $\tau < 1/44$, then $|(V^{-1}h)_i| \leq |(Sh)_i| + |(Rh)_i| = o(n^{-1/2})$. This completes the proof.

\[\square\]

**PROOF OF THEOREM 2.** By (5.12), we have
\[
(\hat{\theta} - \theta)_i = [S\{g - \mathbb{E}(g)\}]_i + [R\{g - \mathbb{E}(g)\}]_i + (V^{-1}h)_i.
\]

By Lemmas 8 and 9, if $\|\theta^*\|_\infty \leq \tau \log n$ and $\tau < 1/44$, then
\[
(\hat{\theta} - \theta)_i = [S\{g - \mathbb{E}(g)\}]_i + o(n^{-1/2}).
\]

Theorem 2 follows directly from Proposition 2. \[\square\]

**SUPPLEMENTARY MATERIAL**

**Supplement to “Asymptotics in directed exponential random graph models with an increasing bi-degree sequence”:**

The supplemental material containing proofs of the theorems and lemmas in Subsections 2.3 and 2.4, and Proposition 1, is available by sending emails to tingyanty@mail.cnu.edu.cn.

**References.**

[1] Adamic L. A. and Glance N. (2005) The political blogosphere and the 2004 US Election. Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem.
[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, doi:10.1186/1471-2105-4:2.
[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. Math. Statist.*, **43**, 193–204.
[6] Bickel, P. J., Chen, A. and Levina, E. (2012). The method of moments and degree distributions for network models. *Ann. Statist.*, **39**, 2280–2301.
[7] Bradley, R. A. and Terry, M. E. (1952). The rank analysis of incomplete block designs I. The method of paired comparisons. *Biometrika* **39** 324–345.
[8] Chatterjee S., Diaconis P., and Sly A. (2011). Random graphs with a given degree sequence. *Annals of Applied Probability*, **21**, 1400–1435.
[9] Chen N. and Olvera-Cravioto M. (2013). Directed random graphs with given degree distributions. *Stochastic Systems*, **3**, 1–40.
[10] Diesner J. and Carley K. M. (2005). Exploration of Communication Networks from the Enron Email Corpus. Proceedings of Workshop on Link Analysis, Counterterrorism and Security, SIAM International Conference on Data Mining, pp. 3–14.
[11] Erdős, P. L., Péter L., Miklós, I., and Toroczkai, Z. (2010) A simple Havel-Hakimi type algorithm to realize graphical degree sequences of directed graphs. The Electronic Journal of Combinatorics. Research Paper R66.

[12] Fienberg S. E. (2012) A brief history of statistical models for network analysis and open challenges. Journal of Computational and Graphical Statistics, 21, 825–839.

[13] Fienberg, S. E., Petrović, S. and Rinaldo, A. (2011). Algebraic statistics for random graph models: markov bases and their uses. in Looking Back: Proceedings of a Conference in Honor of Paul W. Holland (Vol. 202 of Lecture Notes in Statistics), eds. N. J. Dorans and S. Sinharay, New York: Springer, pp. 21C-38.

[14] Fienberg, S. E. and Wasserman, S. S. (1981a). Categorical data analysis of single sociometric relations. Sociological Methodology, 1981, 156–192.

[15] Fienberg S. E. and Wasserman S. (1981b). An exponential family of probability distributions for directed graphs: comment. Journal of the American Statistical Association, 76(373), 54–57.

[16] Fienberg S. E. and Rinaldo A. (2012). Maximum likelihood estimation in log-linear models. Annals of Statistics, 40(2), 996–1023.

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

[18] Girvan M. and Newman M. E. J. (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America. 99, 7821–7826.

[19] Haberman, S. J. (1977) Maximum likelihood estimates in exponential response models. The Annals of Statistics, 5, 815–841.

[20] Haberman, S. J. (1981). An exponential family of probability distributions for directed graphs: comment. Journal of the American Statistical Association, 76(373), 60–61.

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

[22] Hoefding W. (1963). Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association 58, 13–30.

[23] Holland, P. W., and Leinhardt, S., (1981). An exponential family of probability distributions for directed graphs (with discussion). Journal of the American Statistical Association 76, 33–65.

[24] Kantorovich, L. V. (1948) On Newton's method for functional equations, Dokl. Akad. Nauk. SSSR, 59 (1948), 1257-1260.

[25] Kim, H., Genio, C. I. Del., Bassler, K. E., and Toroczkai, Z. (2012). Constructing and sampling directed graphs with given degree sequences. New Journal of Physics. 14, 023012.

[26] Koskinet, G. and Watts, D. J. (2006). Empirical Analysis of an Evolving Social Network. Science, 311, 88–90.

[27] Lang, S. (1993). Real and Functional Analysis. Springer.

[28] Loeve, M. (1977). Probability Theory. 4th Ed. New York: Springer-Verlag.

[29] Newman M. E. J. (2002). The spread of epidemic disease on networks. Phys. Rev. E, 66(016128).

[30] Nepusz T., Yu H. and Paccanaro A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. Nature methods. 18, 471–472.

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

[32] Ortega, J. M. (1968). The Newton-Kantorovich Theorem. The American Mathematical Monthly, 75, 658–660.

[33] Ortega, J. M., and Rheinboldt, W. C., (1970) Iterative solution of nonlinear equations in several variables, New York: Academic Press.

[34] 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 (Vol. 516), eds. M. A. G. Vianaand H. P. Wynn, Providence, RI: American Mathematical Society, pp. 261-C283.

[35] Polvak, B. T. (2006). Newton-Kantorovich method and its global convergence, Journal of Mathematical Science, 133, 1513–1523.

[36] Rinaldo A., Petrović S., and Fienberg S. E. (2013). Maximum likelihood estimation in the β-model Ann. Statist. 41(3), 1085–1110.

[37] Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007a). An introduction to exponential random graph (p*) models for social networks. Social Networks, 29, 173–191.

[38] Robins, G. L., Sniders, T. A. B., Wang, P., Handcock, M., and Pattison, P. (2007b). Recent developments in exponential random graph (p*) models for social networks. Social Networks, 29, 192–215.

[39] Robins G., Pattison P., and Wang P. (2009). Closure, connectivity and degree distributions: Exponential random graph (p*) models for directed social networks. Social Networks, 31, 105–117.

[40] Salathia M., Kazandjievab M., Leeb J. W., Levisb P., Marcus W. 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.

[41] 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.

Tapia, R. A., (1971). The kantorovich theorem for newton’s method, The American Mathematical Monthly 78, 389–392.

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.

Wainwright M. and Jordan M. I. (2008). Graphical models, exponential families, and variational inference. Foundations and Trends in Machine Learning, 1(1–2), 1–305.

Wu N. (1997). The maximum entropy method. New York, Springer.

Yan T. and Leng C. (2013). A simulation study of the \( p_1 \) model. Statistics and Its Interface, To appear.

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

Yan T., Zhao Y. and Qin H. (2013). Asymptotic normality in the maximum entropy models on graphs with an increasing number of parameters. Available at http://arxiv.org/abs/1308.1768.

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
University of Warwick
Coventry, CV4 7AL, UK.
E-mail: C.leng@warwick.ac.uk

Department of Statistics
University of Michigan
Ann Arbor, USA.
E-mail: jizhu@umich.edu
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 (×100%) for $\alpha_i - \alpha_j$ for a pair (i, j) / the length of the confidence interval / the frequency (×100%) that the MLE did not exist.

### Binary Weights

| n   | (i, j)  | $L = 0$       | $L = \log(\log n)$ | $L = (\log(n))^{1/2}$ | $L = \log(n)$ |
|-----|---------|---------------|---------------------|------------------------|--------------|
| 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   |
|     | (50,51) | 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

| n   | (i, j)  | $L = 0$       | $L = \log(\log n)$ | $L = \log(n)$ |
|-----|---------|---------------|---------------------|--------------|
| 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

| n   | (i, j)  | $L = 0$       | $L = \log(\log n)$ | $L = (\log(n))^{1/2}$ | $L = \log(n)$ |
|-----|---------|---------------|---------------------|------------------------|--------------|
| 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   |