The Canadian Journal of Statistics
Vol. xx, No. yy, 20??, Pages 1–25

La revue canadienne de statistique

Asymptotic Theory in Bipartite Graph Models with a Growing Number of Parameters

Yifan Fan[1], Binyan Jiang[2], Ting Yan[3] and Yuan Zhang[4*]

[1]*School of Mathematical Sciences, Guizhou Normal University*

[2]*Department of Applied Mathematics, Hong Kong Polytechnic University*

[3]*Department of Statistics, Central China Normal University*

[4]*Department of Statistics, The Ohio State University*

Key words and phrases: Bipartite graphs; asymptotic properties; node degrees; moment estimation.

MSC 2010: Primary 91D30; secondary 62E20

Abstract: Affiliation networks contain a set of actors and a set of events, where edges denote the affiliation relationships between actors and events. Here, we introduce a class of affiliation network models for modelling the degree heterogeneity, where two sets of degree parameters are used to measure the activeness of actors and the popularity of events, respectively. We develop the moment method to infer the degree parameters. We establish a unified theoretical framework in which the consistency and asymptotic normality of the moment estimator hold as the numbers of actors and events both go to infinity. We apply our results to several popular models with weighted edges, including the generalized β-model, Poisson model and Rayleigh model. We also carry out simulations and a real data application under the Poisson model to demonstrate the theoretical results.

* *Author to whom correspondence may be addressed. E-mail: yzhanghf@stat.osu.edu*

All authors contributed equally to this work.

1. INTRODUCTION

The affiliation relationships between a set of actors and a set of events can be conventionally represented by a bipartite graph, where edges only exist between nodes of distinct parties of this graph, i.e., actors and events. We use "actor" as a covering term that may stand for "actress", "author", "member" and the like. Correspondingly, "event" could denote "movie", "paper" and "club", where edges denote which movie actresses play, which paper authors write and with which club members are associated. In this paper, we study weighted networks and consider various edge-wise distributions. As increasing amounts of affiliation network data are collected, it is important to understand the generative mechanisms of these networks and to explore various characteristics of the network structures in a principled way. As a result, the analysis of affiliation networks has attracted great interests in recent years [e.g., [13, 9, 14, 1]]. In an interesting work [1], the authors propose a data generation procedure that produces new synthetic networks that match the degree distributions and the metamorphosis coefficient as the original network. Compared to our method, their approach is model-free while solely focuses on the two mentioned structural features. Their approach and ours come from different angles and make mutually complementing contributions to the available toolbox.

Node degrees carry admittedly important structural information [2] and play central roles in many network models [e.g., [11, 6, 3]]. A frequently observed phenomenon is that many nodes have low degrees while some others have high degrees, which is referred to as degree heterogeneity. Random graph models have been proposed to model the degree heterogeneity in undirected and directed networks, including the p_1-model [8], the β-model [6], the null model

[12], and the maximum entropy models [7], where each node is assigned one parameter to model the tendency of nodes to participate in network connection. Asymptotic theory for many of these models have also been derived [e.g., [6, 7, 18, 16, 17, 20]].

Despite the significant advances in degree-based network models for undirected and directed graphs, analogous results have not been established for bipartite graphs. In this paper, we introduce a class of bipartite graph models for modelling the degree heterogeneity in bipartite graphs and study their theoretical properties. Our main contributions are three-fold. First, we formulate a *general* model framework for ERGM's structures driven by the degree heterogeneity in bipartite graphs. Our framework significantly extends the scope of existing works such as [16] and [19]. Second, we develop a computationally feasible moment estimator. Our proposed estimator conveniently works for several popular models as special cases of our general framework, such as the β-model, the Poisson model and the Rayleigh model. Third, we present theoretical analysis of our proposed estimator and establish its consistency and asymptotic normality under mild conditions. These three aspects of main contributions are further supported by our numerical examples that demonstrate the utility of our method on both synthetic and real data sets.

The rest of this paper is organized as follows. In Section 2, we introduce a very general model for bipartite graphs, and propose a moment-equations based estimation framework. In Section 3, we present the key asymptotic properties of our estimator. In Section 4, we apply our general results in Section 3 to several popular bipartite network models that cover a wide range of settings, including continuous and discrete edge weights, and demonstrate the effective utility of our unified results. Section 5 presents simulation studies and a real data example under the

Poisson model. Section 6 contains discussion. All proofs are relegated to the Supplementary.

2. BIPARTITE NETWORK MODELS AND ESTIMATION

Let $\mathcal{G}(m,n)$ be a bipartite graph with m actors and n events. Define $[m] := \{1,\dots,m\}$ and $[n] := \{1,\dots,n\}$ by the set of the actors and the set of the events, respectively. Without loss of generality, we assume $n \le m$ hereafter in order to simplify the presentation of results and proofs. In this paper we study weighted edges and let $x_{i,j} \in \Lambda \subseteq \mathbb{R}$ be the edge weight between actor i and event j. Let $X = (x_{i,j})_{m\times n}$ be the bi-adjacency matrix of $\mathcal{G}(m,n)$. Define $\mathbf{d} = (d_1,\dots,d_m)^\top$ and $\mathbf{b} = (b_1,\dots,b_n)^\top$ to be the degrees of actors and events, where $d_i = \sum_{j=1}^n x_{i,j}$ and $b_j = \sum_{i=1}^m x_{i,j}$, respectively.

Edge weights could take discrete or continuous values. For instance, in an athlete-event network, we may use a binary weight to record the presence/absence of an athlete i in event j. In a bus-station network, an edge $a \in \mathbb{N}_0$ counts the number of buses arriving at station i in a day. In an insect–flower network, continuously-weighted edges $a \in \mathbb{R}^+$ represent the frequencies of insects picking flowers.

We introduce a general model framework for modeling the degree heterogeneity of bipartite graphs, which can be described as follows. Suppose that the probability density (mass) function of the edge weight $x_{i,j}$ between actor i and event j has the following form:

$$x_{i,j} = a|\alpha_i, \beta_j \sim f((\alpha_i + \beta_j)a), \quad i = 1,\dots,m, j = 1,\dots,n, \tag{1}$$

where $f(\cdot)$ is a probability density or mass function, α_i is the degree parameter of actor i measuring the activeness of actors, and β_j is the degree parameter of event j measuring the popularity of events. We further assume all edges are independently generated. The above model can be viewed as a generalization of a class of directed and undirected degree-based network models [e.g., [8, 6, 16]] to bipartite graphs. For example, a logistic $f(\cdot)$ corresponds to the bipartite version of the p_1 model for directed graphs in [8].

We note that the value of $f(\cdot)$ in (1) is invariant under the transforms $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ to $(\boldsymbol{\alpha} - c, \boldsymbol{\beta} + c)$ for a constant c. For model identification, without loss of generality we constrain $\beta_n = 0$.

To estimate the model parameters, we use a moment method instead of maximum likelihood estimation. When $f(\cdot)$ is an exponential family distribution, both methods are equivalent. Let $\mu(\cdot)$ denote the expectation of $f(\cdot)$. By definition, we have $\mathbb{E}(x_{i,j}) = \mu(\alpha_i + \beta_j)$. Equating population and sample versions of node degrees, we have the following moment equations:

$$\begin{aligned} d_i &= \textstyle\sum_{k=1}^{n} \mu(\alpha_i + \beta_k), \quad i = 1, \cdots, m, \\ b_j &= \textstyle\sum_{k=1}^{m} \mu(\alpha_k + \beta_j), \quad j = 1, \cdots, n-1. \end{aligned} \tag{2}$$

One can easily verify, $\sum_{i=1}^{m} d_i = \sum_{j=1}^{m} b_j$, therefore, the number of effective moment equations is $m + n - 1$ and we formulate the equations for $d_1, \ldots, d_m, b_1, \ldots, b_{n-1}$ in (2) and shall use them to estimate the $m + n - 1$ free model parameters. Denoted our moment estimator as the solution to (2) by $\widehat{\boldsymbol{\theta}} := (\widehat{\alpha}_1, \cdots, \widehat{\alpha}_m, \widehat{\beta}_1, \cdots, \widehat{\beta}_{n-1})$. We could use Newton-Raphson algorithm to solve $\widehat{\boldsymbol{\theta}}$.

To discuss the existence and uniqueness of $\widehat{\boldsymbol{\theta}}$, define

$$\begin{aligned} F_i(\boldsymbol{\theta}) &= d_i - \sum_{k=1}^{n} \mu(\alpha_i + \beta_k), \quad i = 1, \cdots, m, \\ F_{m+j}(\boldsymbol{\theta}) &= b_j - \sum_{k=1}^{m} \mu(\alpha_k + \beta_j), \quad j = 1, \cdots, n-1. \end{aligned} \tag{3}$$

and let $F(\boldsymbol{\theta}) = (F_1(\boldsymbol{\theta}), \ldots, F_{m+n-1}(\boldsymbol{\theta}))^\top$. Generally speaking, one cannot always anticipate the Jacobian matrix $F'(\boldsymbol{\theta})$ to be invertible, which naturally leads to the existence and uniqueness of $\widehat{\theta}$, but fortunately, in the next section, we will show that $\widehat{\boldsymbol{\theta}}$ exists with probability approaching one under mild conditions.

3. ASYMPTOTIC PROPERTIES

3.1. Notation and preliminaries

Let $\mathbb{R}^+ = (0, \infty)$, $\mathbb{R}_0 = [0, \infty)$, $\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and $\|\mathbf{x}\|_\infty = \max_{1\le i\le n} |x_i|$ be the ℓ_∞-norm of $\mathbf{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$. Let $\Omega(\mathbf{x}, r)$ denote the ball $\{\mathbf{y} : \|\mathbf{x} - \mathbf{y}\| \le r\}$. Define the $\|\cdot\|_\infty$ matrix norm for a matrix $J \in \mathbb{R}^{n\times n}$ as

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

Denote the $\|\cdot\|_{\max}$ matrix norm by $\|J\|_{\max} := \max_{i,j} |J_{i,j}|$. Define the true parameter values by $\boldsymbol{\theta}^* = (\alpha_1^*, \ldots, \alpha_m^*, \beta_1^*, \ldots, \beta_{n-1}^*)$.

The asymptotic behaviors of the moment estimator crucially depend on the Jacobian matrix of $F(\boldsymbol{\theta})$. It turns out that this Jacobian is structured, and we characterize its structure by the notion of a general matrix class as follows. For $Q \ge q > 0$, we say that a $(m+n-1)\times(m+n-1)$-dimensional matrix $V = (v_{i,j})$ belongs to the matrix class $\mathcal{L}_{m,n}(q, Q)$ if the following conditions hold:

$$
\begin{aligned}
& q \le v_{i,i} - \textstyle\sum_{j=m+1}^{m+n-1} v_{i,j} \le Q, i = 1, \ldots, m, \\
& v_{i,j} = 0, i, j = 1, \ldots, m, i \neq j, \\
& v_{i,j} = 0, i, j = m+1, \ldots, m+n-1, i \neq j, \\
& q \le v_{i,j} = v_{j,i} \le Q, i = 1, \ldots, m, j = m+1, \ldots, m+n-1, \\
& v_{i,i} = \textstyle\sum_{k=1}^{m} v_{k,i} = \sum_{k=1}^{m} v_{i,k}, i = m+1, \ldots, m+n-1.
\end{aligned} \tag{4}
$$

One can easily verify the following properties of V in $\mathcal{L}_{m,n}(q,Q)$: it is symmetric, element-wise nonnegative and diagonally dominant, and hence V is positive semidefinite. Further, V is strictly positive definite, since for any $(z_1, \ldots, z_{m+n-1}) \in \mathbb{R}^{m+n-1}$, $z^\top V z = 0$ implies $z = \mathbf{0}$. Moreover, one can verify that $F'(\boldsymbol{\theta}) \in \mathcal{L}_{m,n}(q,Q)$.

For narration convenience, we also define $v_{m+n,i} = v_{i,m+n} := v_{i,i} - \sum_{j=1;j\neq i}^{m+n-1} v_{i,j}$, where $i = 1, \ldots, m+n-1$, and $v_{m+n,m+n} := \sum_{i=1}^{m+n-1} v_{m+n,i}$. Generally, the inversion V^{-1} does not have a closed form expression. To estimate V^{-1} for inference purposes, we devise a matrix $S = (s_{i,j})$ to approximate V^{-1}, constructed as follows:

$$
s_{i,j} = \begin{cases}
\frac{\delta_{i,j}}{v_{i,i}} + \frac{1}{v_{m+n,m+n}}, i, j = 1, \ldots, m, \\
-\frac{1}{v_{m+n,m+n}}, \qquad i = 1, \ldots, m, \ j = m+1, \ldots, m+n-1, \\
-\frac{1}{v_{m+n,m+n}}, \qquad i = m+1, \ldots, m+n-1, j = 1, \ldots, m, \\
\frac{\delta_{i,j}}{v_{i,i}} + \frac{1}{v_{m+n,m+n}}, i, j = m+1, \ldots, m+n-1,
\end{cases} \tag{5}
$$

and $\delta_{i,j} = 1$ if $i = j$ and is otherwise zero. The upper bound of the approximation error is presented in Lemma S1.1 of Supplementary material.

3.2. Consistency, uniqueness and asymptotic normality

Suppose for some $q_{m,n}$ and $Q_{m,n}$, we have $q_{m,n} < \alpha_i^* + \beta_j^* < Q_{m,n}$ for all i, j. We will first show that there is a unique solution in the neighborhood of $\boldsymbol{\theta}^*$:

$$\mathbb{D} = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty \leq 2r\} \subset \{\boldsymbol{\theta} : q_{m,n} - 4r \leq \alpha_i + \beta_j \leq Q_{m,n} + 4r,\ i \in [m], j \in [n]\}, \tag{6}$$

where $r = \|[F'(\boldsymbol{\theta}^*)]^{-1} F(\boldsymbol{\theta}^*)\|_\infty$, and further show that this local solution is also the unique global solution. Assume that $\mu(\cdot)$ is second order differentiable and satisfies the following two regularity conditions

- Condition (1): When $0 < q_{m,n} \leq u \leq Q_{m,n}$, there are three positive numbers $b_{m,n,0}, b_{m,n,1}$, and $b_{m,n,2}$ such that

$$b_{m,n,0} \leq |\mu'(u)| \leq b_{m,n,1}, \tag{7}$$

$$|\mu''(u)| \leq b_{m,n,2}, \tag{8}$$

  where $b_{m,n,0}$, $b_{m,n,1}$ and $b_{m,n,2}$ may depend on $q_{m,n}$ and $Q_{m,n}$.

- Condition (2): $x_{i,j}$ is sub-exponential in that $\mathbb{E}[x_{i,j}^p]^{1/p} \leq \lambda_{i,j} p$ for all $p \geq 1$ for some parameter $\lambda_{i,j} > 0$, and there exists a constant $\widetilde{Q}_{m,n} > 0$ such that $\sup_{i,j} \lambda_{i,j} \leq \widetilde{Q}_{m,n}$.

We notice that the assumption (7) guarantees that $\mu'(u)$ is always positive (or always negative) in $[q_{m,n}, Q_{m,n}]$ and thus bounded away from zero. We would use this fact later. A well-known corollary of Condition (2) is the sub-exponential Bernstein's inequality as follows. A proof can be found in Corollary 2.8.3 of [15].

Lemma 1. *Under Condition (2), for a sufficiently large constant $b_{m,n,3} > 0$, there exists a constant $C_{\widetilde{Q}_{m,n}, b_{m,n,3}}$ such that with probability at least $1 - C_{\widetilde{Q}_{m,n}, b_{m,n,3}}(m+n)^{-1}$, it holds that*

$$\max\{\max_{i=1,\cdots,m} |d_i - \mathbb{E}(d_i)|, \max_{j=1,\cdots,n} |b_j - \mathbb{E}(b_j)|\} \leq b_{m,n,3}(\sqrt{m\log m} + \sqrt{n\log n}). \tag{9}$$

Since we earlier assumed $n \leq m$, the right hand side of (9) can be replaced by $b_{m,n,3}\sqrt{m\log m}$ with $C_{\widetilde{Q}_{m,n}, b_{m,n,3}}$ taking a different constant value. Now, we are ready to present our first main result.

Theorem 1. *Suppose Conditions (1) and (2) hold, $m/n = O(1)$, and*

$$\frac{b_{m,n,1}^2 b_{m,n,3}}{b_{m,n,0}^3} = o\left(\sqrt{\frac{n}{\log m}}\right), \tag{10}$$

$$\frac{b_{m,n,1}^4 b_{m,n,2} b_{m,n,3}}{b_{m,n,0}^6} = o\left(\sqrt{\frac{n}{\log m}}\right). \tag{11}$$

(i) Then with probability at least $1 - C_{\widetilde{Q}_{m,n}, b_{m,n,3}}(m+n)^{-1}$, the moment estimator $\widehat{\boldsymbol{\theta}}$ exists and is unique in the following neighborhood:

$$\left\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty \leq \frac{C b_{m,n,0}^3 \cdot n}{b_{m,n,1}^2 b_{m,n,2} \cdot m}\right\} \tag{12}$$

where C is some global constant, and moreover, for this $\widehat{\boldsymbol{\theta}}$, we also have

$$\| \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \|_\infty = O\left(\frac{b_{m,n,1}^2 b_{m,n,3}}{b_{m,n,0}^3}\sqrt{\frac{\log m}{n}}\right) = o(1). \tag{13}$$

(ii) The unique solution $\widehat{\boldsymbol{\theta}}$ in the neighborhood (12) is also the unique solution in $\{\boldsymbol{\theta} : q_{m,n} \leq \alpha_i + \beta_j \leq Q_{m,n},\ i \in [m], j \in [n]\}$.

Remark 1. The main technique that we used to establish the main results is the analysis of an oracle Newton iteration initiated at $\boldsymbol{\theta}^*$, which provably converges to $\widehat{\boldsymbol{\theta}}$. Here we briefly explain the roles of the assumptions in this theorem. In Condition (1), the equation (7) ensures that the Jacobian matrix $F'(\boldsymbol{\theta})$ belongs to the matrix class $\mathcal{L}_{m,n}(b_{m,n,0}, b_{m,n,1})$ (or $-F'(\boldsymbol{\theta}) \in \mathcal{L}_{m,n}(b_{m,n,0}, b_{m,n,1})$. For simplicity, in this paper, we only focus on the case where $\mu'(u) > 0$ and $F \in \mathcal{L}$. Proof for the case where $\mu'(u) < 0$ and $-F \in \mathcal{L}$ can be established similarly). (8) guarantees that the Jacobian matrix $F'(\boldsymbol{\theta})$ is Lipschitz continuous, which would be a key in bounding the errors of the Newton iterations and eventually the error in $\widehat{\boldsymbol{\theta}}$. Condition (2) implies Lemma 1 and some important concentration results dependent on node degrees in our analysis, see details in Lemmas S1.4, S1.6, S1.8 in the Supplementary material. The apparently mild assumptions (10) and (11) will be needed in some technical proof steps. In fact, our results allow the complexity of the population model to increase with the sample sizes m, n. In most existing models, the population distribution is fixed, and in this case, $q_{m,n}, Q_{m,n}, b_{m,n,0}, b_{m,n,1}, b_{m,n,2}$ and $b_{m,n,3}$ will all be global constants, and our assumptions (10) and (11) would trivially hold.

Remark 2. If $b_{m,n,2}/b_{m,n,0}$ goes to zero slowly enough, there exists a small constant $\epsilon > 0$ such that the solution to the moment functions (3) exists and is unique in $\Omega(\boldsymbol{\theta}^*, \epsilon)$. Moreover, if the initial point $\boldsymbol{\theta}^{(0)}$ is close enough to the true value. that is, $\boldsymbol{\theta}^{(0)} \in \Omega(\boldsymbol{\theta}^*, \epsilon)$, the iterative sequence provably converges to $\widehat{\boldsymbol{\theta}}$. For more details, see conclusion (iii) of Lemma S1.2 in the Supplementary material. For cleanness, in this paper we only present theoretical guarantees for the estimator obtained by the Newton iterations initiated at $\boldsymbol{\theta}^{(0)} = \boldsymbol{\theta}^*$, and the analysis for $\boldsymbol{\theta}^{(0)} \in \Omega(\boldsymbol{\theta}^*, \epsilon)$ is similar but with much more involved formulation.

Our second main result describes the asymptotic normality of $\widehat{\boldsymbol{\theta}}$. It turns out that the asymptotic variance of $\widehat{\boldsymbol{\theta}}$ can be characterized by the covariance structure of observed node degrees. Let $\mathbf{g} = (d_1, \cdots, d_m, b_1, \cdots, b_{n-1})^T$ and $g_{m+n} = b_n$ denote the observed degree sequence. Define $U = (u_{i,j}) = \mathrm{Cov}\,\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}$. Now we assume the following mild regularity condition.

Condition (3) For some $0 < \eta_{m,n,4} < \eta_{m,n,5}$, we have

$$\eta_{m,n,4} \le \min_{i,j} \mathrm{Var}(x_{i,j}) \le \max_{i,j} \mathrm{Var}(x_{i,j}) \le \eta_{m,n,5}. \tag{14}$$

We have the following relationship between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{g}$:

Lemma 2. *Under Condition (3) and*

$$\frac{b_{m,n,1}^6 b_{m,n,3}^2 b_{m,n,2} m^{1/2} \log m}{b_{m,n,0}^9 n} = o(1), \tag{15}$$

$$b_{m,n,1}^2 \eta_{m,n,5} / b_{m,n,0}^3 \eta_{m,n,4}^{1/2} = o(n^{1/2}),$$

then we have

$$(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)_i = [S^*\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}]_i + o_p(m^{-1/2}). \tag{16}$$

where S^* is the designed approximation in (5) to the inverse matrix of $V^* = F'(\boldsymbol{\theta}^*)$, V^* denotes the Jacobian matrix of $F(\boldsymbol{\theta})$ at the true point $\boldsymbol{\theta}^*$, and we borrow the $o_p(\cdot)$ notion from the commonly-used probability theory notion system.

Notice that S^* in (16) is a non-random coefficient matrix, and $\mathbf{g} - \mathbb{E}(\mathbf{g})$ is asymptotically normal. Now we characterize the distribution of $\mathbf{g}$ via its covariance matrix. Under assumption (14), for $1 \le i \ne i' \le m$ and $m+1 \le j \ne j' \le m+n-1$, we have $u_{i,i'} = \text{Cov}(\sum_{j'} x_{i,j'}, \sum_{j'} x_{i',j'}) = 0$, and similarly $u_{j,j'} = 0$, and further, $u_{i,j} = \text{Cov}(\sum_{j'} x_{i,j'}, \sum_{i'} x_{i',j}) = \text{Var}(x_{i,j})$. It is easy to check the remaining conditions to verify that indeed $U \in \mathcal{L}_{m,n}(\eta_{m,n,4}, \eta_{m,n,5})$. Define $u_{m+n,m+n} = \text{Var}(b_n)$, and define

$$Z = u_{m+n,m+n}/v^2_{m+n,m+n} + \text{diag}(u_{1,1}/v^2_{1,1}, \cdots, u_{m+n-1,m+n-1}/v^2_{m+n-1,m+n-1}).$$

Since $d_1/u_{1,1}, \ldots, d_r/u_{r,r}$ and $b_1/u_{m+1,m+1}, \ldots, b_s/u_{m+s,m+s}$ are asymptotically independent, we have the following characterization of g's asymptotic behavior.

Proposition 1. *Under Condition (2), if $m/n = O(1)$, then*

$$(g_i - \mathbb{E}(g_i))/u_{i,i}^{1/2} \xrightarrow{d} N(0,1), m \to \infty,$$

where g_i is the ith element of $\mathbf{g}$. Also, for any fixed $k \ge 1$, as $m \to \infty$, the vector consisting of the first k elements of $S^*\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}$ is asymptotically normal with mean zero and covariance matrix given by the upper left $k \times k$ submatrix of Z.

Remark 3. When the maximum likelihood equations of f coincides with the moment equations, as is the case in maximum entropy models, Poisson model and β-model, then (14) implies (7), and it can be shown by a order-one Taylor expansion that the Jacobian matrix of the parameter vector and the covariance matrix of the degree sequence coincide. If $x_{i,j}$ is a sub-exponential random variable with parameter $\lambda_{i,j}$, $\mathbb{E}|x_{i,j}|^3 \le (3\lambda_{i,j})^3$. For $i \in [m]$, $u_{i,i}^{-3/2} \sum_{j=1}^n \mathbb{E}|x_{i,j}|^3 \le C_0 m^{-1/2} \eta_{m,n,4}^{-3/2} \to 0$, as $m \to \infty$, where C_0 is some absolute constant. By Lyapunov Central Limit Theorem [[5], page 362], we get that $u_{i,i}^{-1/2}(d_i - \mathbb{E}(d_i)) \xrightarrow{d} N(0,1), m \to \infty$. Similarly, we also obtain that for $j \in [n]$, $u_{m+j,m+j}^{-1/2}(b_j - \mathbb{E}(b_j)) \xrightarrow{d} N(0,1), m \to \infty$.

With the above preparations, we are now ready to present our second main result on the asymptotic normality of $\widehat{\boldsymbol{\theta}}$.

Theorem 2. *Suppose (8), (13)–(15) and $b_{m,n,1}^2\eta_{m,n,5}/b_{m,n,0}^3\eta_{m,n,4}^{1/2} = o(n^{1/2})$ hold, then for any fixed $k \geq 1$, as $m \to \infty$, we have*

$$\widehat{\boldsymbol{\theta}}_{[1:k]} - \boldsymbol{\theta}_{[1:k]} \xrightarrow{d} N(0, Z_{[1:k,1:k]}) \tag{17}$$

where $\widehat{\boldsymbol{\theta}}_{[1:k]}$ and $\boldsymbol{\theta}_{[1:k]}$ are the first k elements of the corresponding vector and $Z_{[1:k,1:k]}$ pertains to the upper left $k \times k$ submatrix of Z.

Remark 4. Theorem 2 implies that for any fixed k, asymptotically, the standard deviation of $\widehat{\theta}_i$ is $Z_{i,i} = u_{i,i}^{1/2}/v_{i,i}$, satisfying $(\eta_{m,n,4}/mb_{m,n,1}^2)^{1/2} \preceq u_{i,i}^{1/2}/v_{i,i} \preceq (\eta_{m,n,5}/nb_{m,n,0}^2)^{1/2}$. Using Theorem 2, we can conveniently construct approximate marginal and joint confidence intervals for estimating $\boldsymbol{\theta}^*$. For example, an approximate $1-\alpha$ confidence interval for $\theta_i - \theta_j$ is $\widehat{\theta}_i - \widehat{\theta}_j \pm Z_{1-\alpha/2}(\widehat{u}_{i,i}/\widehat{v}_{i,i}^2 + \widehat{u}_{j,j}/\widehat{v}_{j,j}^2)^{1/2}$, where $Z_{1-\alpha/2}$ is the $1-\alpha/2$-quantile of the standard normal distribution, and $\widehat{v}_{i,i}$ and $\widehat{v}_{j,j}$ are the moment estimates of $v_{i,i}$ and $v_{j,j}$ by replacing all θ_i with their moment estimates. Here, $\widehat{u}$ and $\widehat{v}$ are the estimated covariance matrix $\mathrm{Cov}\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}$ and the estimated Jacobian matrix $F'(\widehat{\boldsymbol{\theta}})$, respectively.

4. APPLICATIONS

4.1. Generalized β-model

The β-model [6] is an exponential random graph model with the degree sequence as the exclusively sufficient statistic. Here, we generalize it to bipartite graphs. For simplicity, we assume that edges belong to the sample space $\Lambda = \{0, 1, \ldots, q-1\}$, where $q \geq 2$ is a constant. Assume the edges $x_{i,j}$'s are independently generated with the following probability mass function:

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

In this example, we consider $\mathbb{D}$ defined in (6) and set $q_{m,n} = -Q_{m,n}$. By definition, we have

$$\mu(\alpha_i + \beta_j) = \sum_{a=0}^{q-1} \frac{a e^{a(\alpha_i+\beta_j)}}{\sum_{k=0}^{q-1} e^{k(\alpha_i+\beta_j)}}.$$

Now we check the conditions of Theorem 1. For Condition (1), straight calculation shows that,

$$\mu'(\alpha_i + \beta_j) = \frac{\sum_{0\leq k<l\leq q-1} (k-l)^2 e^{(k+l)(\alpha_i+\beta_j)}}{(\sum_{a=0}^{q-1} e^{a(\alpha_i+\beta_j)})^2}.$$

This yields

$$\frac{1}{2(1+e^{Q_{m,n}+2r})} \leq |\mu'(\alpha_i + \beta_j)| \leq \frac{q^2}{2}, \quad \text{and} \quad \frac{1}{2(1+e^{Q_{m,n}})} \leq |\mu'(\alpha_i^* + \beta_j^*)| \leq \frac{q^2}{2}.$$

Thus $F'(\boldsymbol{\theta}^*) \in \mathcal{L}_{m,n}(b^*_{m,n,0}, b^*_{m,n,1})$, where $b^*_{m,n,1} = \frac{q^2}{2}, b^*_{m,n,0} = \frac{1}{2(1+e^{Q_{m,n}})}$.

Next, Condition (2) is easy to verify since the distribution is discrete with a finite sample space, it is certainly sub-exponential. Thus, we choose $b_{m,n,3} = q-1$. Again, by Lemma S1.1

in the Supplementary material, we have

$$r = O\left(e^{3Q_{m,n}}\sqrt{\frac{\log m}{n}}\right).$$

If $e^{Q_{m,n}} = o\left((n/\log m)^{1/6}\right)$ and $m/n = O(1)$, we have $r \to 0$ as $m \to \infty$. For any $\boldsymbol{\theta} \in \Omega(\boldsymbol{\theta}^*, 2r)$, we have $F'(\boldsymbol{\theta}) \in \mathcal{L}_{m,n}(b_{m,n,0}, b_{m,n,1})$, where

$$b_{m,n,1} = \frac{q^2}{2},\ \ b_{m,n,0} = \frac{1}{2(1+e^{Q_{m,n}+2r})}. \tag{18}$$

On the other hand, for any $i \in [m], j \in [n]$,

$$\mu''(\alpha_i+\beta_j) = \frac{(1/2)\sum_{k\neq l,a}(k-l)^2(k+l-2a)e^{(k+l+a)(\alpha_i+\beta_j)}}{(\sum_{a=0}^{q-1} e^{a(\alpha_i+\beta_j)})^3}.$$

Since $\sum_{k\neq l,a} e^{(k+l+a)(\alpha_i+\beta_j)} \le (\sum_{a=0}^{q-1} e^{a(\alpha_i+\beta_j)})^3$, we can choose $b_{m,n,2} = (q-1)^3$. If $e^{Q_{m,n}} = o\left((n/\log m)^{1/12}\right)$,

$$\frac{b_{m,n,1}^4 b_{m,n,2} b_{m,n,3}}{b_{m,n,0}^6}\sqrt{\frac{\log m}{n}} = O\left(e^{6Q_{m,n}}\sqrt{\frac{\log m}{n}}\right) = o(1),$$

where $b_{m,n,1}$ and $b_{m,n,0}$ are given in (18), then (13) is satisfied. By Theorem 1, the uniform consistency of $\widehat{\boldsymbol{\theta}}$ holds as follows.

Corollary 1. *If* $e^{Q_{m,n}} = o\left((n/\log m)^{1/12}\right)$, $m/n = O(1)$, *then as* m *goes to infinity, with probability approaching one, the moment estimator* $\widehat{\boldsymbol{\theta}}$ *exists and is unique in* $\{\boldsymbol{\theta} : q_{m,n} \le \alpha_i + \beta_j \le Q_{m,n},\ i \in [m], j \in [n]\}$. *Furthermore, we have*

$$\| \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \|_\infty = O\left(e^{3Q_{m,n}}\sqrt{\frac{\log m}{n}}\right) = o(1).$$

Note that the moment equations of the maximum entropy distributions are equal to the maximum likelihood equations, then the covariance matrix of $\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}$ is $V = F'(\boldsymbol{\theta}) \in \mathcal{L}_{m,n}(b_{m,n,0}, b_{m,n,1})$, such that

$$\frac{n}{2(1+e^{Q_{m,n}})} \leq v_{i,i} \leq \frac{mq^2}{2},\ i = 1, \cdots, m+n-1.$$

By the central limit theorem for the bounded case [[10], page 289], $v_{i,i}^{-1/2}(d_i - \mathbb{E}(d_i))$ and $v_{m+j,m+j}^{-1/2}(b_j - \mathbb{E}(b_j))$ are asymptotically standard normal if $v_{i,i}$ diverges. If $e^{Q_{m,n}} = o\left(\frac{n^{1/9}}{(\log m)^{1/9} m^{1/18}}\right)$, then

$$\frac{b_{m,n,1}^6 b_{m,n,3}^2 b_{m,n,2} m^{1/2} \log m}{b_{m,n,0}^9 n} = O\left(\frac{e^{9Q_{m,n}} m^{1/2} \log m}{n}\right) = o(1).$$

Now that all required conditions checked, we apply Theorem 2 and obtain

Corollary 2. If $e^{Q_{m,n}} = o\left(\frac{n^{1/9}}{(\log m)^{1/9} m^{1/18}}\right)$, $m/n = O(1)$, then for any fixed $k \geq 1$, as $m \to \infty$, the vector consisting of the first k elements of $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*$ is asymptotically multivariate normal with mean zero and covariance matrix given by the upper left $k \times k$ submatrix of Z.

4.2. Poisson model

In this example, independent weighted edges take value in $\Lambda = \mathbb{N}_0$ and each follows an edge-specific Poisson distribution:

$$P(x_{i,j} = a) = \frac{e^{a(\alpha_i+\beta_j)}}{a!} \exp(-e^{\alpha_i+\beta_j}), a = 0, 1, 2, \cdots . \tag{19}$$

Now we check the conditions of Theorem 1. For Condition (1), we have

$$\mu(\alpha_i + \beta_j) = e^{\alpha_i+\beta_j} \quad \text{and} \quad \mu'(\alpha_i + \beta_j) = e^{\alpha_i+\beta_j}.$$

For Condition (1), we have, for $\boldsymbol{\theta} \in \mathbb{D}$,

$$e^{q_{m,n}-2r} \le |\mu'(\alpha_i + \beta_j)| \le e^{Q_{m,n}+2r}, \quad \text{and} \quad e^{q_{m,n}} \le |\mu'(\alpha_i^* + \beta_j^*)| \le e^{Q_{m,n}}.$$

Thus $F'(\boldsymbol{\theta}^*) \in \mathcal{L}_{m,n}(b^*_{m,n,0}, b^*_{m,n,1})$, where $b^*_{m,n,1} = e^{Q_{m,n}}$ and $b^*_{m,n,0} = e^{q_{m,n}}$. By Lemmas S1.1 and S1.4 in the Supplementary material, we can set

$$r = O\left(e^{4Q_{m,n}-3q_{m,n}}\sqrt{\frac{\log m}{n}}\right).$$

If $e^{4Q_{m,n}-3q_{m,n}} = o\left((n/\log m)^{1/2}\right), m/n = O(1)$, then $r \to 0$ as m goes to infinity. For any $\boldsymbol{\theta} \in \Omega(\boldsymbol{\theta}^*, 2r)$, we have $F'(\boldsymbol{\theta}) \in \mathcal{L}_{m,n}(b_{m,n,0}, b_{m,n,1})$, where

$$b_{m,n,1} = e^{Q_{m,n}}, \quad b_{m,n,0} = e^{q_{m,n}-2r}, \quad \text{and} \quad b_{m,n,2} = e^{Q_{m,n}}. \tag{20}$$

By Lemma S1.4 in the Supplementary material, we can choose

$$b_{m,n,3} = 2c\sqrt{\frac{2e^{4Q_{m,n}}}{\gamma}}, \tag{21}$$

where c, γ are absolute constants. If $e^{7Q_{m,n}-6q_{m,n}} = o\left((n/\log m)^{1/2}\right)$,

$$\frac{b_{m,n,1}^2 b_{m,n,3}}{b_{m,n,0}^3}\sqrt{\frac{\log m}{n}} = o(1),$$
$$\frac{b_{m,n,1}^4 b_{m,n,2} b_{m,n,3}}{b_{m,n,0}^6}\sqrt{\frac{\log m}{n}} = O\left(e^{7Q_{m,n}-6q_{m,n}}\sqrt{\frac{\log m}{n}}\right) = o(1),$$

where $b_{m,n,1}$ and $b_{m,n,0}$ are given in (20). Now Condition (1) is verified. Also, Condition (2) is satisfied since Poisson distribution is sub-exponential. By Theorem 1, we have

Corollary 3. If $e^{7Q_{m,n}-6q_{m,n}} = o\left((n/\log m)^{1/2}\right)$, $m/n = O(1)$, then as m goes to infinity, with probability approaching one, the moment estimator $\widehat{\boldsymbol{\theta}}$ exists and is unique in $\{\boldsymbol{\theta} : q_{m,n} \le \alpha_i + \beta_j \le Q_{m,n},\ i \in [m], j \in [n]\}$. Furthermore this unique solution satisfies

$$\| \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \|_\infty = O\left(e^{4Q_{m,n}-3q_{m,n}}\sqrt{\frac{\log m}{n}}\right) = o(1).$$

In this example, the moment equations coincide with the maximum likelihood equations. The covariance matrix of $\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}$ is $V = F'(\boldsymbol{\theta}) \in \mathcal{L}_{m,n}(b_{m,n,0}, b_{m,n,1})$. As the third moment of the Poisson with parameter $e^{\alpha_i+\beta_j}$ is $\mathbb{E}(a_{i,j}^3) = e^{\alpha_i+\beta_j} + 3e^{2(\alpha_i+\beta_j)} + e^{3(\alpha_i+\beta_j)}$. Recall that $0 < q_{m,n} \le \alpha_i^* + \beta_j^* \le Q_{m,n}$, we have

$$\frac{\sum_{j=1}^n \mathbb{E}(a_{i,j}^3)}{v_{i,i}^{3/2}} \le \frac{n(e^{Q_{m,n}} + 3e^{2Q_{m,n}} + e^{3Q_{m,n}})}{n^{3/2}e^{3q_{m,n}/2}} \le \frac{5e^{4Q_{m,n}-2q_{m,n}}}{n^{1/2}}$$

and

$$\frac{\sum_{i=1}^m \mathbb{E}(a_{i,j}^3)}{v_{m+j,m+j}^{3/2}} \le \frac{5e^{4Q_{m,n}-2q_{m,n}}}{n^{1/2}}.$$

If $e^{2Q_{m,n}-q_{m,n}} = o(n^{1/4})$, the above expression goes to zero. This satisfies the conditions of the Lyapunov Central Limit Theorem [[5], page 362]. Therefore, $v_{i,i}^{-1/2}(d_i - \mathbb{E}(d_i))$ is also asymptotically standard normal when $e^{2Q_{m,n}-q_{m,n}} = o(n^{1/4})$. Similarly, $v_{m+j,m+j}^{-1/2}(b_j - \mathbb{E}(b_j))$ is also asymptotically standard normal under the same conditions. If $e^{4Q_{m,n}-3q_{m,n}} = o\left(\frac{n^{1/3}}{m^{1/6}(\log m)^{1/3}}\right)$, then

$$\frac{b_{m,n,1}^6 b_{m,n,3}^2 b_{m,n,2} m^{1/2}\log m}{b_{m,n,0}^9 n} = O\left(\frac{e^{12Q_{m,n}-9q_{m,n}} m^{1/2}\log m}{n}\right) = o(1).$$

Now we can apply Theorem 2 and obtain the following result.

Corollary 4. If $e^{4Q_{m,n}-3q_{m,n}} = o\left(\frac{n^{1/3}}{m^{1/6}(\log m)^{1/3}}\right)$, $m/n = O(1)$, then for any fixed $k \geq 1$, as $m \to \infty$, the vector consisting of the first k elements of $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*$ is asymptotically multivariate normal with mean zero and covariance matrix given by the upper left $k \times k$ submatrix of Z.

4.3. Rayleigh distribution

Our third example assumes that independent edges $x_{i,j} \in \mathbb{R}^+$ are sampled from the following Rayleigh density function

$$f(x_{i,j} = a) = \frac{a}{e^{\alpha_i+\beta_j}} e^{-a^2/(2e^{\alpha_i+\beta_j})}, a > 0.$$

To verify Condition (1), notice that

$$\mu(\alpha_i + \beta_j) = \sqrt{\frac{\pi}{2}} e^{(\alpha_i+\beta_j)/2}.$$

By direct calculations, we have, for $\boldsymbol{\theta} \in \mathbb{D}$,

$$
\begin{aligned}
&\mu'(\alpha_i + \beta_j) = \frac{1}{2}\sqrt{\frac{\pi}{2}} e^{(\alpha_i+\beta_j)/2}, \\
&\frac{1}{2}\sqrt{\frac{\pi}{2}} e^{(q_{m,n}-2r)/2} \le |\mu'(\alpha_i + \beta_j)| \le \frac{1}{2}\sqrt{\frac{\pi}{2}} e^{(Q_{m,n}+2r)/2}, \\
&\frac{1}{2}\sqrt{\frac{\pi}{2}} e^{q_{m,n}/2} \le |\mu'(\alpha_i^* + \beta_j^*)| \le \frac{1}{2}\sqrt{\frac{\pi}{2}} e^{Q_{m,n}/2}.
\end{aligned}
$$

Thus $F'(\boldsymbol{\theta}^*) \in \mathcal{L}_{m,n}(b^*_{m,n,0}, b^*_{m,n,1})$, where $b^*_{m,n,1} = \frac{1}{2}\sqrt{\frac{\pi}{2}} e^{Q_{m,n}/2}, b^*_{m,n,0} = \frac{1}{2}\sqrt{\frac{\pi}{2}} e^{q_{m,n}/2}$. By Lemmas S1.1 and S1.6 in the Supplementary material, we have

$$
r = O\left(e^{\frac{3}{2}(Q_{m,n}-q_{m,n})} \sqrt{\frac{\log m}{n}} \right).
$$

If $e^{Q_{m,n}-q_{m,n}} = o\left((n/\log m)^{1/3}\right)$, then $r \to 0$ as m goes to infinity. For any $\boldsymbol{\theta} \in \Omega(\boldsymbol{\theta}^*, 2r)$, we have $F'(\boldsymbol{\theta}) \in \mathcal{L}_{m,n}(b_{m,n,0}, b_{m,n,1})$, where

$$
b_{m,n,1} = \frac{1}{2}\sqrt{\frac{\pi}{2}} e^{Q_{m,n}/2}, b_{m,n,0} = \frac{1}{2}\sqrt{\frac{\pi}{2}} e^{q_{m,n}/2}. \tag{22}
$$

On the other hand, for any $i \in [m], j \in [n]$,

$$
\mu''(\alpha_i + \beta_j) = \frac{\partial^2 \mu(\alpha_i + \beta_j)}{\partial \beta_j \partial \alpha_i} = \frac{1}{4}\sqrt{\frac{\pi}{2}} e^{(\alpha_i+\beta_j)/2}.
$$

Thus, $b_{m,n,2} = \frac{1}{4}\sqrt{\frac{\pi}{2}} e^{Q_{m,n}/2}$. By Lemma S1.6 in Supplementary material, we can choose

$$
b_{m,n,3} = 2c\sqrt{\frac{2e^{Q_{m,n}}}{\gamma}}, \tag{23}
$$

where c, γ are absolute constants, respectively. If $e^{Q_{m,n}-q_{m,n}} = o\left((n/\log m)^{1/6}\right)$,

$$\frac{b_{m,n,1}^2 b_{m,n,3}}{b_{m,n,0}^3}\sqrt{\frac{\log m}{n}} = O\left(e^{\frac{3}{2}(Q_{m,n}-q_{m,n})}\sqrt{\frac{\log m}{n}}\right) = o(1),$$
$$\frac{b_{m,n,1}^4 b_{m,n,2} b_{m,n,3}}{b_{m,n,0}^6}\sqrt{\frac{\log m}{n}} = O\left(e^{3(Q_{m,n}-q_{m,n})}\sqrt{\frac{\log m}{n}}\right) = o(1),$$

where $b_{m,n,1}$ and $b_{m,n,0}$ are given in (22). Also Condition (2) holds because the distribution is sub-Gaussian, thus also sub-exponential. By Theorem 1, we have

Corollary 5. If $e^{Q_{m,n}-q_{m,n}} = o\left((n/\log m)^{1/6}\right), m/n = O(1)$, then as m goes to infinity, with probability approaching one, the moment estimator $\widehat{\boldsymbol{\theta}}$ exists and is unique in $\{\boldsymbol{\theta} : q_{m,n} \le \alpha_i + \beta_j \le Q_{m,n},\ i \in [m], j \in [n]\}$. Furthermore this unique solution satisfies

$$\| \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \|_\infty = O\left(e^{\frac{3}{2}(Q_{m,n}-q_{m,n})}\sqrt{\frac{\log m}{n}}\right) = o(1).$$

Again, note that both $d_i = \sum_{k=1}^n a_{i,k}$ and $b_j = \sum_{k=1}^m a_{k,j}$ are sums of n and m independent random variables, respectively. It can be shown that $U = \mathrm{Cov}\{\mathbf{g} - \mathbb{E}(\mathbf{g})\} \in \mathcal{L}_{m,n}(\eta_{m,n,4}, \eta_{m,n,5})$, where

$$\eta_{m,n,4} = \frac{4-\pi}{2}e^{q_{m,n}}, \eta_{m,n,5} = \frac{4-\pi}{2}e^{Q_{m,n}}.$$

Hence we have

$$\frac{4-\pi}{2}ne^{q_{m,n}} \le u_{i,i} \le \frac{4-\pi}{2}me^{Q_{m,n}},\ i = 1, \cdots, m+n.$$

As the third moment of the Rayleigh distribution with parameter $e^{\frac{\alpha_i+\beta_j}{2}}$ is $\mathbb{E}(a_{i,j}^3) = 3\sqrt{\frac{\pi}{2}}e^{\frac{3(\alpha_i+\beta_j)}{2}}$. Recall that $0 < q_{m,n} \leq \alpha_i^* + \beta_j^* \leq Q_{m,n}$, we have

$$\frac{\sum_{j=1}^n \mathbb{E}(a_{i,j}^3)}{v_{i,i}^{3/2}} \leq \frac{3n * 4\sqrt{\frac{\pi}{2}}e^{3Q_{m,n}}}{n^{3/2}e^{3q_{m,n}/2}} \leq \frac{24e^{3Q_{m,n}-q_{m,n}}}{n^{1/2}},$$

and

$$\frac{\sum_{i=1}^m \mathbb{E}(a_{i,j}^3)}{v_{m+j,m+j}^{3/2}} \leq \frac{24e^{3Q_{m,n}-q_{m,n}}}{n^{1/2}}.$$

If $e^{3Q_{m,n}-q_{m,n}} = o(n^{1/4})$, the above expression goes to zero. This satisfies the conditions of the Lyapunov Central Limit Theorem [[5], page 362]. Therefore, $v_{i,i}^{-1/2}(d_i - \mathbb{E}(d_i))$ is also asymptotically standard normal when $e^{3Q_{m,n}-q_{m,n}} = o(n^{1/4})$. Similarly, $v_{m+j,m+j}^{-1/2}(b_j - \mathbb{E}(b_j))$ is also asymptotically standard normal under the same conditions. If $e^{Q_{m,n}-q_{m,n}} = o\left(\frac{n^{2/9}}{(\log m)^{2/9}m^{1/9}}\right)$, then

$$\frac{b_{m,n,1}^6 b_{m,n,3}^2 b_{m,n,2} m^{1/2}\log m}{b_{m,n,0}^9 n} = O\left(\frac{e^{\frac{9}{2}(Q_{m,n}-q_{m,n})}m^{1/2}\log m}{n}\right) = o(1).$$

Applying Theorem 2, we have the folloiwng asymptotic normality result for $\widehat{\boldsymbol{\theta}}$.

Corollary 6. If $e^{Q_{m,n}-q_{m,n}} = o\left(\frac{n^{2/9}}{(\log m)^{2/9}m^{1/9}}\right)$, $m/n = O(1)$, then for any fixed $k \geq 1$, as $m \to \infty$, the vector consisting of the first k elements of $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*$ is asymptotically multivariate normal with mean zero and covariance matrix given by the upper left $k \times k$ submatrix of Z.

4.4. Maximum entropy distributions with continuous weights

We consider maximum entropy distributions with continuous weights, that is, $\Lambda = \mathbb{R}_0$, where the moment equations are equal to the maximum likelihood equations. Assume that each $x_{i,j}, i \in [m], j \in [n]$, are mutually independent exponential random variables with the density function

$$f(x_{i,j} = a) = (\alpha_i + \beta_j)e^{-a(\alpha_i+\beta_j)}, a > 0.$$

We first apply Theorem 1 to obtain the existence and consistency of $\widehat{\boldsymbol{\theta}}$. In this case, we have

$$\mu(\alpha_i + \beta_j) = (\alpha_i + \beta_j)^{-1}.$$

By direct calculations, we have, for $\boldsymbol{\theta} \in \mathbb{D}$,

$$\mu'(\alpha_i + \beta_j) = \frac{-1}{(\alpha_i + \beta_j)^2},$$

$$\frac{1}{(Q_{m,n} + 2r)^2} \le \left|\frac{1}{(\alpha_i + \beta_j)^2}\right| \le \frac{1}{(q_{m,n} - 2r)^2}, \quad \frac{1}{Q_{m,n}^2} \le \left|\frac{1}{(\alpha_i^* + \beta_j^*})^2\right| \le \frac{1}{q_{m,n}^2}.$$

Thus $-F'(\boldsymbol{\theta}^*) \in \mathcal{L}_{m,n}(b^*_{m,n,0}, b^*_{m,n,1})$, where $b^*_{m,n,1} = q_{m,n}^{-2}, b^*_{m,n,0} = Q_{m,n}^{-2}$. By Lemmas S1.1 and S1.8 in the Supplementary material, we have

$$r = O\left(\frac{Q_{m,n}^6}{q_{m,n}^5}\sqrt{\frac{\log m}{n}}\right).$$

If $Q_{m,n}/q_{m,n} = o\left((n/\log m)^{1/12}\right)$, then $r \to 0$ as m goes to infinity. For any $\boldsymbol{\theta} \in \Omega(\boldsymbol{\theta}^*, 2r)$, we have $-F'(\boldsymbol{\theta}) \in \mathcal{L}_{m,n}(b_{m,n,0}, b_{m,n,1})$, where

$$b_{m,n,1} = \frac{1}{q_{m,n}^2}, b_{m,n,0} = \frac{1}{Q_{m,n}^2}. \tag{24}$$

On the other hand, for any $i \in [m], j \in [n]$,

$$\mu''(\alpha_i + \beta_j) = \frac{\partial^2 \mu(\alpha_i + \beta_j)}{\partial \beta_j \partial \alpha_i} = \frac{2}{(\alpha_i + \beta_j)^3}.$$

Thus, $b_{m,n,2} = 2/q_{m,n}^3$. By Lemma S1.8 in the Supplementary material, we can choose

$$b_{m,n,3} = \sqrt{\frac{8}{\gamma q_{m,n}^2}}, \tag{25}$$

where γ is an absolute constant. If $Q_{m,n}/q_{m,n} = o\left((n/\log m)^{1/24}\right)$,

$$\frac{b_{m,n,1}^2 b_{m,n,3}}{b_{m,n,0}^3}\sqrt{\frac{\log m}{n}} = O\left(\frac{Q_{m,n}^6}{q_{m,n}^5}\sqrt{\frac{\log m}{n}}\right) = o(1),$$

$$\frac{b_{m,n,1}^4 b_{m,n,2} b_{m,n,3}}{b_{m,n,0}^6}\sqrt{\frac{\log m}{n}} = O\left(\frac{Q_{m,n}^{12}}{q_{m,n}^{12}}\sqrt{\frac{\log m}{n}}\right) = o(1),$$

where $b_{m,n,1}$ and $b_{m,n,0}$ are given in (24). Condition (2) is checked since this distribution is sub-exponential. By Theorem 1, the uniform consistency of $\widehat{\boldsymbol{\theta}}$ is as follows.

Corollary 7. If $Q_{m,n}/q_{m,n} = o\left((n/\log m)^{1/24}\right), m/n = O(1)$, then as m goes to infinity, with probability approaching one, the moment estimator $\widehat{\boldsymbol{\theta}}$ exists and is unique in $\{\boldsymbol{\theta} : q_{m,n} \le \alpha_i + \beta_j \le Q_{m,n},\ i \in [m], j \in [n]\}$. Furthermore this unique solution satisfies

$$\| \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \|_\infty = O\left(\frac{Q_{m,n}^6}{q_{m,n}^5}\sqrt{\frac{\log m}{n}}\right) = o(1).$$

Again, note that both $d_i = \sum_{k=1}^n a_{i,k}$ and $b_j = \sum_{k=1}^m a_{k,j}$ are sums of n and m independent exponential random variables, respectively. Note that the moment equations of the maximum entropy distributions are equal to the maximum likelihood equations, then the covariance matrix of $\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}$ is $V = -F'(\boldsymbol{\theta}) \in \mathcal{L}_{m,n}(b_{m,n,0}, b_{m,n,1})$. As the third moment of the exponential random variable with parameter $1/(\alpha_i + \beta_j)$ is $\mathbb{E}(a_{i,j}^3) = 6/(\alpha_i + \beta_j)^3$. Under the assumption that $0 < q_{m,n} \le \alpha_i^* + \beta_j^* \le Q_{m,n}$, we have

$$\frac{\sum_{j=1}^n \mathbb{E}(a_{i,j}^3)}{v_{i,i}^{3/2}} = \frac{6\sum_{j=1}^n (\alpha_i+\beta_j)^{-3}}{v_{i,i}^{3/2}} \le \frac{6Q_{m,n}^3/q_{m,n}^3}{n^{1/2}}$$

and

$$\frac{\sum_{j=1}^n \mathbb{E}(a_{i,j}^3)}{v_{m+j,m+j}^{3/2}} = \frac{6\sum_{j=1}^n (\alpha_i+\beta_j)^{-3}}{v_{m+j,m+j}^{3/2}} \le \frac{6Q_{m,n}^3/q_{m,n}^3}{n^{1/2}}$$

If $Q_{m,n}/q_{m,n} = o(n^{1/6})$, then the above expression goes to zero. This shows that the condition for the Lyapunov Central Limit Theorem [[5], page 362] holds. Therefore, $v_{i,i}^{-1/2}(d_i - \mathbb{E}(d_i))$ is also asymptotically standard normal if $Q_{m,n}/q_{m,n} = o(n^{1/6})$. Similarly, $v_{m+j,m+j}^{-1/2}(b_j - \mathbb{E}(b_j))$ is also asymptotically standard normal under the same conditions. If $Q_{m,n}/q_{m,n} = o(\frac{n^{1/18}}{m^{1/36}(\log m)^{1/18}})$, then

$$\frac{b_{m,n,1}^6 b_{m,n,3}^2 b_{m,n,2} m^{1/2}\log m}{b_{m,n,0}^9 n} = o(1),$$

By Theorem 2, the asymptotic normality of $\widehat{\boldsymbol{\theta}}$ is as below.

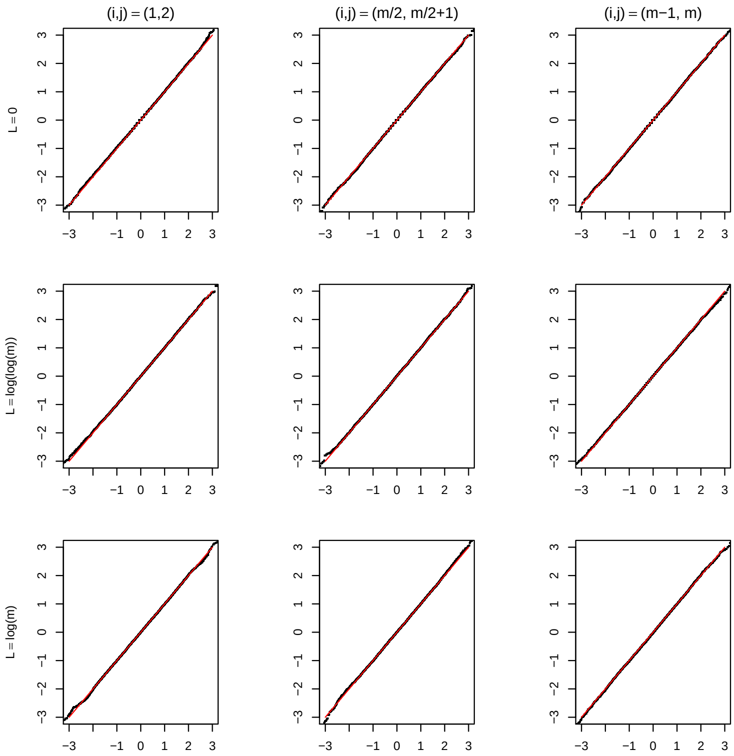
Corollary 8. If $Q_{m,n}/q_{m,n} = o(\frac{n^{1/18}}{m^{1/36}(\log m)^{1/18}})$, $m/n = O(1)$ then for any fixed $k \ge 1$, as $m \to \infty$, the vector consisting of the first k elements of $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*$ is asymptotically multivariate normal with mean zero and covariance matrix given by the upper left $k \times k$ submatrix of Z.

5. NUMERICAL STUDIES

In this section, we present numerical experiments on synthetic data generated by the Poisson model (19) to assess the performance of our moment estimator. This model is widely used to describe the likelihood of discrete events occurring in a continuous manner, such as website visits, user-ratings, crime and disease incident reports. We also present an data example of the US Law Firms and World Cities network.

5.1. Simulations

FIGURE 1: The QQ plots of $\widehat{v}_{i,i}^{1/2}(\widehat{\alpha}_i - \alpha_i)$



(a) Poisson model (m=100, n=50)

We set $\alpha_i^* = (m-i)L/(m-1)$, $\beta_i^* = (n-i)L/(n-1)$ and $\beta_n^* = 0$. Here we fix the bipartite network size at $(m, n) = (100, 50)$, $(200, 100)$ and test L on log scale, that is, $L \in \{0, \ \log(\log m), \ \log m\}$. We consider the asymptotical distributions of $\widehat{\xi}_{i,j} = [\widehat{\alpha}_i - \widehat{\alpha}_j - (\alpha_i^* - \alpha_j^*)]/(1/\widehat{v}_{i,i} + 1/\widehat{v}_{j,j})^{1/2}$ and $\widehat{\eta}_{i,j} = [\widehat{\beta}_i - \widehat{\beta}_j - (\beta_i^* -$

$\beta_j^*)]/(1/\widehat{v}_{n+i,n+i} + 1/\widehat{v}_{m+j,m+j})^{1/2}$, where $\widehat{v}_{i,i}$ is the estimate of $v_{i,i}$ by replacing $\boldsymbol{\theta}_i^*$ with $\widehat{\boldsymbol{\theta}}_i$, and empirically verify Corollary 4. We assess the asymptotic normality of $\widehat{\xi}_{i,j}, \widehat{\eta}_{i,j}$ by Q-Q plot under various L. We also evaluate the coverage probabilities and the lengths of the 95% confidence intervals. We also record the frequency that the estimate does not exist. Each simulation is repeated $10,000$ times.

We test the bipartite network size of $(m, n) = (100, 50)$ and $(200, 100)$, respectively, and find the Q-Q plots of $\alpha_i^* - \alpha_j^*$ and $\beta_i^* - \beta_j^*$ to be similar under $(m, n) = (100, 50)$ and $(200, 100)$. Therefore, we only show the Q-Q plot of $\alpha_i^* - \alpha_j^*$ under $(m, n) = (100, 50)$ in Figure 1 due to page limit. In Figure 1, the horizontal and vertical axes are the theoretical and empirical quantiles, respectively, and the red lines correspond to the reference lines $y = x$. From Figure 1, we see that under all three $\alpha_i^* - \alpha_j^*$ configurations, the empirical distribution shows clear normality when $L \leq \log(m)$.

Table 1 contains the results for estimating $\theta_i^* - \theta_j^*$, including coverage probabilities and lengths of the 95% confidence intervals and the frequency that a such confidence interval does not exist. Here, a large L regulates the confidence interval length downward, while the length decreases as the bipartite network size of (m, n) increases. As we read from Table 1, the coverage frequencies are all close to the nominal level 95% for all L. Furthermore, we use a Lilliefors test to verity whether our data sample is from a normally distributed population, as shown in Table 2. In Table 2, we see that when $L = 0$, $p = 0.00$ for all pairs (i, j) implies that it is unlikely that this sample came from a normal population; when $L = \log(\log m)$ and $L = \log m$, the p-values for most pairs (i, j) are not less than 0.05, which implies that the empirical distribution

TABLE 1: Estimated coverage probabilities ($\times 100\%$) of $\theta_i^* - \theta_j^*$ for a pair (i,j) as well as the length of confidence intervals (in square brackets), and the probabilities($\times 100\%$) that the estimate does not exist (in parentheses).

node	(i,j)	$L=0$	$L=\log(\log(m))$	$L=\log m$
m=100	(1,2)	95.03[0.40](0)	95.14[0.12](0)	94.88[0.01](0)
	(50,51)	94.92[0.40](0)	94.89[0.18](0)	95.11[0.03](0)
	(99,100)	94.86[0.40](0)	95.23[0.26](0)	94.70[0.08](0)
n=50	(1,2)	94.63[0.28](0)	95.16[0.09](0)	95.00[0.01](0)
	(25,26)	94.85[0.28](0)	94.82[0.13](0)	94.85[0.02](0)
	(49,50)	95.21[0.28](0)	94.87[0.18](0)	94.97[0.06](0)
m=200	(1,2)	95.10[0.28](0)	95.06[0.08](0)	94.79[0.00](0)
	(100,101)	95.24[0.28](0)	94.83[0.12](0)	95.13[0.01](0)
	(199,200)	95.28[0.28](0)	95.12[0.18](0)	95.08[0.05](0)
n=100	(1,2)	95.16[0.20](0)	95.03[0.05](0)	95.13[0.00](0)
	(50,51)	95.38[0.20](0)	95.00[0.08](0)	95.27[0.01](0)
	(99,100)	94.71[0.20](0)	94.85[0.12](0)	95.18[0.03](0)

TABLE 2: Lilliefors test of $\theta_i^* - \theta_j^*$ for a pair (i,j): statistic (in left) and p.value (in square brackets).

node	(i,j)	$L=0$	$L=\log(\log(m))$	$L=\log m$
m=100	(1,2)	0.02[0.00]	0.01[0.02]	0.00[0.93]
	(50,51)	0.02[0.00]	0.01[0.00]	0.01[0.61]
	(99,100)	0.02[0.00]	0.01[0.00]	0.01[0.20]
n=50	(1,2)	0.02[0.00]	0.01[0.32]	0.00[0.92]
	(25,26)	0.01[0.00]	0.01[0.11]	0.01[0.14]
	(49,50)	0.02[0.00]	0.01[0.08]	0.01[0.66]
m=200	(1,2)	0.02[0.00]	0.01[0.03]	0.01[0.01]
	(100,101)	0.02[0.00]	0.01[0.16]	0.01[0.65]
	(199,200)	0.02[0.00]	0.01[0.02]	0.01[0.26]
n=100	(1,2)	0.01[0.00]	0.01[0.69]	0.01[0.40]
	(49,50)	0.01[0.00]	0.01[0.72]	0.01[0.51]
	(99,100)	0.01[0.00]	0.01[0.37]	0.01[0.80]

TABLE 3: The running times of $\theta_i^* - \theta_j^*$ for a pair (i, j).

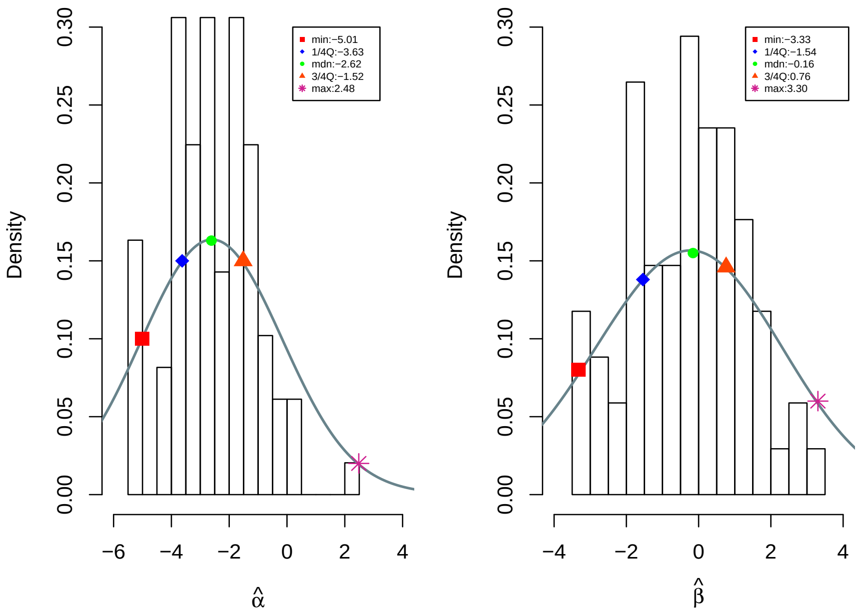
(m, n)	$(100, 50)$	$(200, 100)$	$(400, 200)$
Time in seconds: mean (std.)	0.50(0.02)	7.35(0.06)	113.43(0.20)

may be normally distributed.

In Table 3, we report the running times of our algorithm under different configurations. Our method scales comfortably to networks of roughly 500 senders and receivers, respectively. The memory cost of our method is $O((m+n)^2)$ using a Newton method.

5.2. Data Example

FIGURE 2: The histograms of $\widehat{\alpha}_i$'s and $\widehat{\beta}_j$'s for the US Law Firms and World Cities data with 98 firms and 68 cities.



(a) The histogram of the estimates of the 98 firms (left) and 68 cities (right) parameters in the US Law Firms and World Cities data.

TABLE 4: US Law Firms and World Cities network dataset: the estimators of $\widehat{\alpha}_i$ and $\widehat{\beta}_j$ and their standard errors (in parentheses), and selected top ten and bottom ten Firms and Cities according to the degree sequences, respectively.

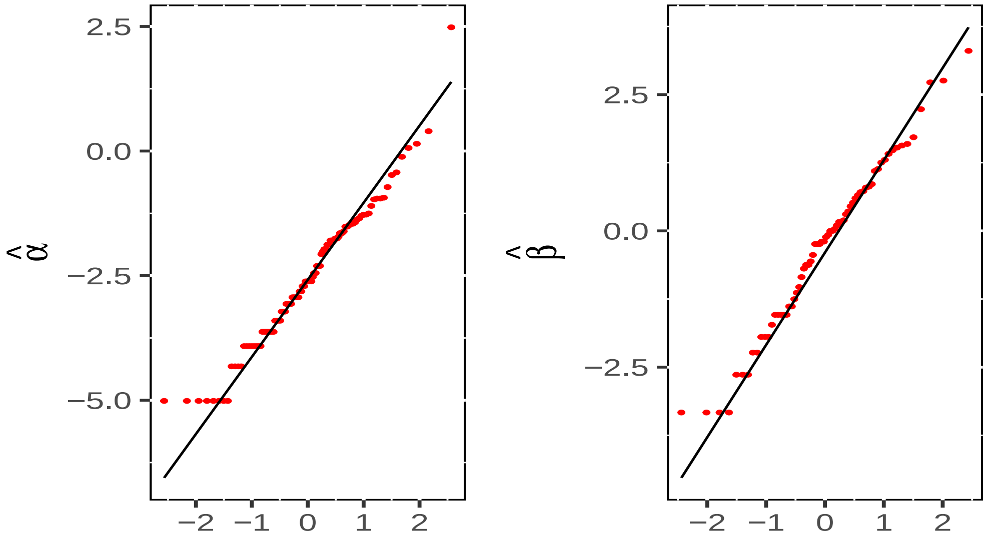
Firm ID	Degree	$\widehat{\alpha}_i$	City ID	Degree	$\widehat{\beta}_j$
45	1802	2.48(0.02)	36	761	3.30(0.04)
2	224	0.40(0.07)	46	441	2.76(0.05)
6	174	0.15(0.08)	27	427	2.72(0.05)
18	160	0.06 (0.08)	11	261	2.23(0.06)
4	134	-0.12 (0.09)	68	156	1.72(0.08)
83	98	-0.43(0.10)	62	138	1.60(0.09)
1	93	-0.48(0.10)	20	134	1.57(0.09)
3	73	-0.72(0.12)	43	129	1.53(0.09)
11	59	-0.94(0.13)	55	123	1.48(0.09)
7	58	-0.95(0.13)	58	115	1.41(0.09)
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
85	2	-4.32(0.71)	67	4	-1.95(0.50)
87	2	-4.32(0.71)	17	3	-2.23(0.58)
43	1	-5.01(1.00)	45	3	-2.23(0.58)
52	1	-5.01(1.00)	12	2	-2.64(0.71)
61	1	-5.01(1.00)	44	2	-2.64(0.71)
71	1	-5.01(1.00)	66	2	-2.64(0.71)
75	1	-5.01(1.00)	21	1	-3.33(1.00)
89	1	-5.01(1.00)	33	1	-3.33(1.00)
95	1	-5.01(1.00)	51	1	-3.33(1.00)
99	1	-5.01(1.00)	65	1	-3.33(1.00)

TABLE 5: The minimum, quartiles and maximum values of degrees from 98 firms and 68 cities.

Degree	minimum	$1/4$ quantile	median	$3/4$ quantile	maximum
d	1	4	11	33	1802
b	1	6	24	59	761

We apply our method to the US Law Firms and World Cities network dataset in [4], extracted from https://www.lboro.ac.uk/gawc/datasets/da5_1.html. The dataset contains the numbers of lawyers of 100 American law firms with foreign offices in 72 cities outside

FIGURE 3: The QQ plots of $\widehat{\alpha}_i$'s and $\widehat{\beta}_j$'s for the US Law Firms and World Cities data with 98 firms and 68 cities.



(a) The QQ plots of the estimates of the 98 firms (left) and 68 cities (right) parameters in the US Law Firms and World Cities data.

US. We pre-processed the data by removing isolated nodes and obtained a bipartite graph with 98 firms and 68 cities. We used the GLM package with a Poission link function in parameter estimation. Table 4 shows the estimated α_i and β_j values with standard errors and observed degrees. Recall that we set $\beta_{69} = 0$. In Table 4, we observe a clear positive relationship between $\widehat{\boldsymbol{\alpha}}$ and $\mathbf{d}$, and between $\widehat{\boldsymbol{\beta}}$ and $\mathbf{b}$ alike. The full version of this table is presented in the Supplementary Material. Table 5 contains quantiles of firm- and city-degrees and shows that $\mathbf{d}$ spreads a wild range from 1 to 1802, so is $\mathbf{b}$, varying 1 to 761. This is also reflected in the spread of $\widehat{\alpha}_i$, ranging from -5.01 to 2.48, and $\widehat{\beta}_j$ from -3.33 to 3.30, as is shown in Figure 2. The histograms of both $\widehat{\alpha}_i$'s and $\widehat{\beta}_j$'s indicate that they approximately follow normal distributions. We

TABLE 6: US Law Firms and World Cities network dataset: the mean degrees of 100 bootstrap degree sequences and their 95% bootstrap confidence intervals(i.e, CI)(in square brackets), and selected top ten and bottom ten Firms and Cities in order, respectively.

Firm ID	d	$\bar{d}$	95% bootstrap CI	City ID	b	$\bar{b}$	95%bootstrap CI
1	93	93.78	[76.31, 111.25]	1	40	39.05	[26.81, 51.29]
2	224	225.57	[194.43, 256.71]	2	103	102.51	[85.00, 120.02]
3	73	73.03	[57.73, 88.33]	3	84	82.57	[64.59, 100.55]
4	134	133.29	[113.32, 153.26]	4	22	21.66	[13.25, 30.07]
5	50	50.76	[37.71 , 63.81]	5	33	33.63	[22.42, 44.84]
7	58	57.47	[42.18 , 72.76]	7	33	33.74	[22.93, 44.55]
8	6	5.94	[1.34 , 10.54]	8	9	9.25	[3.89, 14.61]
9	29	28.4	[18.04 , 38.76]	9	6	6.02	[1.77 , 10.27]
10	43	42.61	[28.00 , 57.22]	10	261	261.36	[225.65, 297.07]
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	3	2.91	[-0.32 , 6.14]	60	5	5.39	[1.11, 9.67]
91	3	3.19	[-0.06 , 6.44]	61	15	14.71	[7.98, 21.44]
92	39	38.29	[26.41 , 50.17]	62	138	138.25	[114.03, 162.47]
93	8	8.28	[2.44 , 14.12]	63	63	64.81	[49.19, 80.43]
94	9	9.08	[3.33 , 14.83]	64	4	3.72	[0.03, 7.41]
95	1	1.1	[-1.23 , 3.43]	65	1	1.03	[-0.88 , 2.94]
96	34	33.51	[21.40 , 45.62]	66	2	2.01	[-0.80 , 4.82]
97	25	25.51	[15.45 , 35.57]	67	4	4.13	[-0.05 , 8.31]
99	1	1.08	[-0.75 , 2.91]	68	156	156.34	[133.49, 179.19]
100	11	10.99	[4.53, 17.45]	69	28	28.05	[18.78, 37.32]

also give the Q-Q plots of both $\widehat{\alpha}_i$'s and $\widehat{\beta}_j$'s in Figure 3. In Figure 3, the horizontal and vertical axes are the theoretical and empirical quantiles, respectively, and the red lines correspond to the reference lines $y = x$. From Figure 3, we see that many points are close to the reference lines $y = x$, especially in the middle, but the points in the tails are a little far off. It appears that both

$\widehat{\alpha}_i$ and $\widehat{\beta}_j$ are approximately follow normal distributions.

As an application of our method, we use the estimated parameters to generate 100 bootstrapped adjacency matrices. The outputted bootstrapped adjacency matrices can then be used to build 95% bootstrap confidence intervals for the degree sequences, as reported in Table 6. We can see that the mean degrees of 100 bootstrap degree sequences are close to the original degrees from the observed network dataset. Moreover, these original degrees are included in the 95% bootstrap confidence intervals. On the other hand, we also calculate the mean skewness of 100 bootstrap degree distributions and the skewness of each bootstrap degree distribution, respectively, as shown in Table 7. Compared with the skewness of original degree distributions, these skewness values are found to be quite similar to each other in Table 7.

We make the following observations from Table 4. First, as one naturally expects, the estimated standard errors on high-degree nodes (also known as "hub nodes" or core nodes) are significantly smaller than low-degree nodes ("leaf nodes", or peripheral nodes), since nodes with higher popularity provide more data about their connection patterns. Second, in light of Remark 4 in Section 3.2, we can construct a marginal approximate 95% t-confidence interval for each α_i and β_j parameter, using $\big(\widehat{\alpha}_i \pm 2\widehat{\mathrm{s.e.}}(\widehat{\alpha}_i)\big)$. Let us inspect the top-10 highest degree firms in Table 4. A high $\widehat{\alpha}_i$ suggests that the company is highly "internationalized". Similarly, a high $\widehat{\beta}_j$ suggests the city's attraction to US law firms. Tracing the top-ranked cities in Table 4, we find the top 3 to be London, Paris and Hong Kong, all of which ranked at the top tiers in the recent *GaWC Global City Index* (London is Alpha++, the other two are Alpha+). Those ranked 4–10 (Brussels, Warsaw, Tokyo, Frankfurt, Moscow and Singapore, in order) suggest that while there is a clear positive relationship between a city's global centrality and its attraction to US law firms, there may be other factors that lead to their discrepancy. For instance, two potentially very relevant other factors are law systems and geographical distance. UK, Hong Kong and Singapore share the same law system (common law) as US, which might facilitate their connection in law businesses. This might also partially explain for Tokyo since Japanese law system is a mixture of civil and common law systems. On the other hand, the observation that European continent cities that adopt civil laws are significantly elevated in the ranking of Table 4 compared to their global city indexes, possibly due to the geographical and cultural closeness to US, compared to some other Asian-Pacific cities with high global centrality. On the other hand, the estimated $\widehat{\alpha}_i$ values in the same table may be a good reminder of the potential limitation of our study. The most internationalized companies (Firm 45) has almost twice as many as the total connections of the rest of top 10 firms. Consequently, this firm alone has a high impact on the estimation of β_j parameters, and one should keep this in mind when interpreting the $\widehat{\beta}_j$ values.

TABLE 7: The skewness of degree distributions from 98 firms and 68 cities: β_S indicates the skewness of original degree distributions, $\widehat{\beta_{S_m}}$ indicates the mean skewness of 100 bootstrap degree distributions, while the skewness of the ith bootstrap degree distributions is denoted by $\widehat{\beta_{S_i}}$.

Type	β_S	$\widehat{\beta_{S_m}}$	$\widehat{\beta_{S_1}}$	$\widehat{\beta_{S_{25}}}$	$\widehat{\beta_{S_{50}}}$	$\widehat{\beta_{S_{75}}}$	$\widehat{\beta_{S_{100}}}$
Firms	9.03	9.02	9.01	8.99	9.05	9.05	9.00
Cities	3.92	3.93	3.82	3.89	4.10	3.83	3.97

6. DISCUSSION

In this paper, we proposed a general model for degree-based modeling and analysis of bipartite graphs. Our model generalizes several popular models along this line of approach in existing literature. In contrast to the common likelihood-based methods, we proposed a moment method for parameter estimation that is computationally efficient and enjoys nice theoretical properties under mild conditions. Our proof makes an original use of the theory on Newton iterations to this setting.

Our model applies to a rich family of network models, and as demonstrated in Section 4, it provides a general framework for systematically studying the asymptotic properties of many existing models as its special cases. The simulations under Poisson model and the data application demonstrated the effectiveness of our method on concrete examples.

There are several interesting future works. One interesting direction is to seek finite sample error bound guarantees rather than asymptotics. The analysis method is not very challenging, but the formulation will become much more involved, therefore for cleanness of results, we did not pursue this goal in this paper.

Another interesting question is to investigate the case where m and n are at very different scales. Very different m and n would significantly complicate the approximation to V^{-1}, which is a key technical ingredient in our analysis, and this complication would propagate to all consequent analysis and results. While we envision this as doable, it amounts to be a separate work and exceeds the scope of this paper.

The asymptotic approximation and concentration results that our analysis relies on can in fact accommodate slight dependency across network edges. Considering the recently raising research interest in networks with dependent edges, we are also interested in exploring the relaxation of the independent edge assumption that is almost universally assumed in the current bipartite graph literature.

Finally, a natural research interest is to consider richer network features beyond degrees. In fact, degrees can be viewed as a rescaled version of the simplest *network moment*, namely, edges [21]. Other motifs such as *stars* and *cycles* are also useful and very meaningful quantities to study. On the other hand, in view of the reality that methods and theory for degree-based ERGM's are still in active development and not yet complete, in this paper, we aim at a more comprehensive understanding of the relatively simpler degree-based models as a solid step forward. Also, while including more statistics into the ERGM, one must also be very careful with model identifiability, which is generally a challenging task for many network models.

Supplementary Materials

Supplement to "Asymptotic theory in bipartite graph models with a growing number of parameters." The supplement contains all the proofs for the main results and the full table of *the real data numerical results* in Section 5.2.

REFERENCES

1 Sinan G Aksoy, Tamara G Kolda, and Ali Pinar. Measuring and modeling bipartite graphs with community structure. *Journal of Complex Networks*, 5(4):581–603, 2017.

2 R Albert and A Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002.

3 Alexander Barvinok and John A Hartigan. The number of graphs and a random graph with a given degree sequence. *Random Structures & Algorithms*, 42(3):301–348, 2013.

4 JV Beaverstock, RG Smith, and PJ Taylor. Geographies and globalization: Us law firms in world cities. *Urban Geography*, 21(2):105–117, 2000.

5 P. Billingsley. *Probability and measure. 3rd ed.* Wiley, New York, 1995.

6 Sourav Chatterjee, Persi Diaconis, Allan Sly, et al. Random graphs with a given degree sequence. *The Annals of Applied Probability*, 21(4):1400–1435, 2011.

7 Christopher Hillar and Andre Wibisono. Maximum entropy distributions on graphs. Available at:http://arxiv.org/abs/1301.3321., 2013.

8 Paul W Holland and Samuel Leinhardt. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–50, 1981.

9 Matthieu Latapy, Clémence Magnien, and Nathalie Del Vecchio. Basic notions for the analysis of large two-mode networks. *Social networks*, 30(1):31–48, 2008.

10 M. Loéve. *Probability theory I. 4th ed.* Springer, New York, 1977.

11 Mark EJ Newman, Steven H Strogatz, and Duncan J Watts. Random graphs with arbitrary degree distributions and their applications. *Physical review E*, 64(2):026118, 2001.

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

13 Garry Robins and Malcolm Alexander. Small worlds among interlocking directors: Network structure and distance in bipartite graphs. *Computational & Mathematical Organization Theory*, 10(1):69–94, 2004.

14 Tom AB Snijders, Alessandro Lomi, and Vanina Jasmine Torló. A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social networks*, 35(2):265–276, 2013.

15 Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

16 Ting Yan, Chenlei Leng, and Ji Zhu. Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *The Annals of Statistics*, 44:31–57, 2016.

17 Ting Yan, Hong Qin, and Hansheng Wang. Asymptotics in undirected random graph models parameterized by the strengths of vertices. *Statistica Sinica*, 26(1):273–293, 2016.

18 Ting Yan and Jinfeng Xu. A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika*, 100(2):519–524, 2013.

19 Yong Zhang, Xiaodi Qian, Hong Qin, and Ting Yan. Affiliation networks with an increasing degree sequence. *Communications in Statistics-theory and Methods*, 46(22):11163–11180, 2017.

20 Yu Zhang, Qiuping Wang, Yuan Zhang, Ting Yan, and Jing Luo. L-2 regularized maximum likelihood for β-model in large and sparse networks. *arXiv preprint arXiv:2110.11856*, 2021.

21 Yuan Zhang and Dong Xia. Edgeworth expansions for network moments. *The Annals of Statistics, (In press)*, 2022+.