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, Guiyang, Guizhou, China*
[2]*Department of Applied Mathematics, Hong Kong Polytechnic University, Hung Hom, Hong Kong, China*
[3]*Department of Statistics, Central China Normal University, Wuhan, Hubei, China*
[4]*Department of Statistics, The Ohio State University, Columbus, Ohio, USA*

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 these 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 generalized β-, Poisson and Rayleigh models. Simulation studies and a realistic example that involves the Poisson model provide concrete evidence that supports our theoretical findings.

Résumé: Les réseaux d'affiliation se composent d'un ensemble d'acteurs et un ensemble d'événements reliés entre eux par des arêtes. Les auteurs de cet article s'intéressent à la modélisation de l'hétérogénéité des degrés de ce type de réseaux et proposent une classe de modèles construits à partir de deux ensembles de paramètres des degrés ayant respectivement trait à l'activité des acteurs et à la popularité des événements. En utilisant la méthode des moments pour estimer ces paramètres, ils réussissent à établir un cadre théorique unifié dans lequel ils prouvent la convergence et la normalité asymptotiques des estimateurs obtenus par la méthode des moments quand les nombres d'acteurs et d'événements tendent tous les deux vers l'infini. Les résultats ainsi obtenus sont ensuite appliqués à plusieurs modèles populaires ayant des arêtes pondérées, tels que les modèles β-, Poisson et Rayleigh généralisés. Enfin, les auteurs appuient leurs résultats théoriques au moyen d'études de simulation et un exemple réaliste basé sur le modèle de Poisson.

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 generic term that may stand for "actress," "author," "member" and the like. Correspondingly, "event" could denote "movie," "paper," and "club," where edges denote a movie in which actresses appear, a paper on which authors

collaborate and a club to which club members belong. In this article, 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 interest in recent years; for example, see Robins & Alexander (2004), Latapy, Magnien & Del Vecchio (2008), Snijders, Lomi & Torló (2013) and Aksoy, Kolda & Pinar (2017). The latter authors propose a data generation procedure that produces new synthetic networks that match the degree distributions and the metamorphosis coefficient of the original network. Compared with the method that we describe below, their approach is model free but focuses solely on the two mentioned structural features. Their approach and ours involve different perspectives, and provide mutually complementary additions to the available toolbox.

Node degrees furnish important structural information (Albert & Barabási, 2002) and play central roles in many network models; for example, see Newman, Strogatz & Watts (2001), Chatterjee, Diaconis & Sly (2011) and Barvinok & Hartigan (2013). A frequently observed phenomenon is that many nodes have small degrees while others have large degrees, which corresponds to a condition known as degree heterogeneity. Random graph models have been proposed to model the degree heterogeneity in both undirected and directed networks, including the p_1-model (Holland & Leinhardt, 1981), the β-model (Chatterjee, Diaconis & Sly, 2011), the null model (Perry & Wolfe, 2012) and the maximum entropy models (Hillar & Wibisono, 2013), 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 has also been derived; for example, see Chatterjee, Diaconis & Sly (2011), Hillar & Wibisono (2013), Yan & Xu (2013), Yan, Leng & Zhu (2016), Yan, Qin & Wang (2016) and Zhang et al. (2021).

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 article, we introduce a class of bipartite graph models for modelling the degree heterogeneity in bipartite graphs and study the theoretical properties of this class. Our main contributions are three-fold. First, we formulate a *general* model framework for the structures of the exponential random graph model, driven by the degree heterogeneity in bipartite graphs. Our framework significantly extends the scope of existing works such as Yan, Leng & Zhu (2016) and Zhang et al. (2017). Second, we develop a computationally feasible moment estimator. Conveniently, our proposed estimator works for several popular models that are special cases of our general framework, such as the β-model, the Poisson model and the Rayleigh model. Third, we outline a theoretical analysis of our proposed estimator and establish its consistency and asymptotic normality under mild conditions. Additionally, these three aspects are reinforced by simulation studies and a realistic example that demonstrate the merits of our method on both synthetic and real datasets.

The rest of this article 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 from 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 reports simulation studies and a realistic case study involving the Poisson model. Section 6 contains discussion. All proofs may be found in the accompanying Supplementary Material.

2. BIPARTITE NETWORK MODELS AND ESTIMATION

Let $\mathcal{G}(m, n)$ be a bipartite graph with m actors and n events. Define $[m] := \{1, \ldots, m\}$ and $[n] := \{1, \ldots, n\}$ by the set of the actors and the set of the events, respectively. Without loss of generality, we assume $n \leq m$ hereafter in order to simplify the presentation of results and proofs. In this article, 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 $X = \left(x_{i,j}\right)_{m \times n}$ be the bi-adjacency matrix of $\mathcal{G}(m, n)$. Define $\mathbf{d} = \left(d_1, \ldots, d_m\right)^\top$ and $\mathbf{b} = \left(b_1, \ldots, b_n\right)^\top$ to be the degrees of actors and events, respectively, where $d_i = \sum_{j=1}^n x_{i,j}$ and $b_j = \sum_{i=1}^m x_{i,j}$.

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 choosing flowers.

We introduce a general model framework for modelling 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:

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

where $f(\cdot)$ is a probability density or mass function, α_i is the degree parameter of actor i measuring the activity 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., Holland & Leinhardt, 1981; Chatterjee, Diaconis & Sly, 2011; Yan, Leng & Zhu, 2016) to bipartite graphs. For example, a logistic $f(\cdot)$ corresponds to the bipartite version of the p_1-model for directed graphs in Holland & Leinhardt (1981).

We note that the value of $f(\cdot)$ in Equation (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:

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

One can easily verify that $\sum_{i=1}^m d_i = \sum_{j=1}^m b_j$; 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 Equation (2) and shall use them to estimate the $m + n - 1$ free model parameters. We denote our moment estimator as the solution to Equation (2) by $\widehat{\theta} := \left(\widehat{\alpha}_1, \ldots, \widehat{\alpha}_m, \widehat{\beta}_1, \ldots, \widehat{\beta}_{n-1}\right)$. We could also use the Newton–Raphson algorithm to solve for $\widehat{\theta}$.

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

$$F_i(\theta) = d_i - \sum_{k=1}^{n} \mu(\alpha_i + \beta_k), \quad i = 1, \ldots, m,$$

$$F_{m+j}(\theta) = b_j - \sum_{k=1}^{m} \mu(\alpha_k + \beta_j), \quad j = 1, \ldots, n-1. \tag{3}$$

and let $F(\theta) = \left(F_1(\theta), \ldots, F_{m+n-1}(\theta)\right)^\top$. Generally speaking, one cannot always expect the Jacobian matrix $F'(\theta)$ to be invertible, which naturally leads to the existence and uniqueness of $\widehat{\theta}$, but fortunately, in the next section, we will show that $\widehat{\theta}$ 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 $\mathbf{x} = \left(x_1, \ldots, x_n\right)^\top \in \mathbb{R}^n$. 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

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

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

The asymptotic behaviours of the moment estimator crucially depend on the Jacobian matrix of $F(\theta)$. 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:

$$q \le v_{i,i} - \sum_{j=m+1}^{m+n-1} v_{i,j} \le Q, \quad i = 1, \ldots, m,$$

$$v_{i,j} = 0, \quad i, j = 1, \ldots, m, i \ne j,$$

$$v_{i,j} = 0, \quad i, j = m+1, \ldots, m+n-1, i \ne j,$$

$$q \le v_{i,j} = v_{j,i} \le Q, \quad i = 1, \ldots, m, j = m+1, \ldots, m+n-1,$$

$$v_{i,i} = \sum_{k=1}^{m} v_{k,i} = \sum_{k=1}^{m} v_{i,k}, \quad i = m+1, \ldots, m+n-1. \tag{4}$$

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 $F'(\theta) \in \mathcal{L}_{m,n}(q, Q)$.

For narration convenience, we also define $v_{m+n,i} = v_{i,m+n} := v_{i,i} - \sum_{j=1; j \ne i}^{m+n-1} v_{i,j}$, 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:

$$
s_{i,j} = \begin{cases}
\dfrac{\delta_{i,j}}{v_{i,i}} + \dfrac{1}{v_{m+n,m+n}}, & i, j = 1, \ldots, m, \\[2mm]
-\dfrac{1}{v_{m+n,m+n}}, & i = 1, \ldots, m, \ j = m+1, \ldots, m+n-1, \\[2mm]
-\dfrac{1}{v_{m+n,m+n}}, & i = m+1, \ldots, m+n-1, j = 1, \ldots, m, \\[2mm]
\dfrac{\delta_{i,j}}{v_{i,i}} + \dfrac{1}{v_{m+n,m+n}}, & i, j = m+1, \ldots, m+n-1,
\end{cases}
\tag{5}
$$

and $\delta_{i,j} = 1$ if $i = j$ and is otherwise zero. The upper bound of the approximation error is identified in Lemma S1.1; see the 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 neighbourhood

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

of $\boldsymbol{\theta}^*$, where $r = \left\| \left[F'(\boldsymbol{\theta}^*) \right]^{-1} F(\boldsymbol{\theta}^*) \right\|_\infty$, 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 $\mathbb{E}\left[x_{i,j}^p \right]^{1/p} \leq \lambda_{i,j} p$ 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 specified in Equation (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 will use this fact later. A well-known corollary of Condition (2) is the sub-exponential Bernstein's inequality (see below). A proof can be found in Corollary 2.8.3 of Vershynin (2018).

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*

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

Since we earlier assumed $n \leq m$, the right-hand side of Equation (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 $\widehat{\theta}$ exists and is unique in the following neighbourhood:*

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

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

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

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

Remark 1. The main technique that we used to establish the key results in Theorem 1 is the analysis of an oracle Newton iteration initiated at θ^*, which provably converges to $\widehat{\theta}$. Here we briefly explain the roles of the assumptions in this theorem. In Condition (1), Equation (7) ensures that the Jacobian matrix $F'(\theta)$ belongs to the matrix class $\mathcal{L}_{m,n}(b_{m,n,0}, b_{m,n,1})$ or $-F'(\theta) \in \mathcal{L}_{m,n}(b_{m,n,0}, b_{m,n,1})$. For simplicity, in this article, 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. Equation (8) guarantees that the Jacobian matrix $F'(\theta)$ is Lipschitz continuous, which would be a key in bounding the errors of the Newton iterations and eventually the error in $\widehat{\theta}$. 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 and S1.8 in the Supplementary Material. The apparently mild assumptions identified in Equations (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 identified in Equations (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 specified in Equation (3) exists and is unique in $\Omega(\theta^*, \epsilon)$.

Moreover, if the initial point $\theta^{(0)}$ is close enough to the true value, that is, $\theta^{(0)} \in \Omega(\theta^*, \epsilon)$, the iterative sequence provably converges to $\widehat{\theta}$. For more details, see conclusion (iii) of Lemma S1.2 in the Supplementary Material. For simplicity, in this article, we only present theoretical guarantees for the estimator obtained by the Newton iterations initiated at $\theta^{(0)} = \theta^*$; the analysis for $\theta^{(0)} \in \Omega(\theta^*, \epsilon)$ is similar but requires a much more involved formulation.

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

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

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

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

Lemma 2. *Under Condition (3) and*

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

$$b_{m,n,1}^2 \eta_{m,n,5} / b_{m,n,0}^3 \eta_{m,n,4}^{1/2} = o(n^{1/2}), \tag{15}$$

then we have

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

where S^* is the particular approximation specified in Equation (5) *to the inverse matrix of $V^* = F'(\theta^*)$, V^* denotes the Jacobian matrix of $F(\theta)$ at the true point θ^* and we borrow the $o_p(\cdot)$ notation from probability theory.*

Notice that S^* in Equation (16) is a nonrandom 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 the assumption specified in Equation (14), for $1 \leq i \neq i' \leq m$ and $m + 1 \leq j \neq j' \leq m + n - 1$, we have $u_{i,i'} = \text{Cov}\left(\sum_{j'} x_{i,j'}, \sum_{j'} x_{i',j'}\right) = 0$, and similarly $u_{j,j'} = 0$, and further, $u_{i,j} = \text{Cov}\left(\sum_{j'} x_{i,j'}, \sum_{i'} x_{i',j}\right) = \text{Var}(x_{i,j})$. 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

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

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 the asymptotic behaviour of g:

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

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

where g_i is the ith element of **g**. *Also, for any fixed $k \geq 1$, 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 coincide with the moment equations, as is the case in the maximum entropy models, Poisson model and β-model, then Equation (14) implies Equation (7), and it can be shown by a first-order 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 $\lambda_{i,j}$, $\mathbb{E}|x_{i,j}|^3 \leq (3\lambda_{i,j})^3$. For $i \in [m], u_{i,i}^{-3/2} \sum_{j=1}^n \mathbb{E}|x_{i,j}|^3 \leq C_0 m^{-1/2} \eta_{m,n,4}^{-3/2} \to 0$, as $m \to \infty$, where C_0 is some absolute constant. By the Lyapunov central limit theorem (Vershynin, 2018; 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 Equations* (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 $(\eta_{m,n,4}/mb_{m,n,1}^2)^{1/2} \lesssim u_{i,i}^{1/2}/v_{i,i} \lesssim (\eta_{m,n,5}/nb_{m,n,0}^2)^{1/2}$. Using Theorem 2, we can conveniently construct approximate marginal and joint confidence intervals for estimating θ^*. 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 $v_{j,j}$ that are obtained by replacing all θ_i with their moment estimates. Here $\widehat{u}$ and $\widehat{v}$ are the estimated covariance matrix 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 (Chatterjee, Diaconis & Sly 2011) is an exponential random graph model with the degree sequence as the exclusively sufficient statistic. Here we generalize the β-model 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 $x_{i,j}$ 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 let $\mathbb{D}$ be as defined in Equation (6) and set $q_{m,n} = -Q_{m,n}$. By definition, we have

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

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

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

This yields

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

Thus, $F'(\theta^*) \in \mathcal{L}_{m,n}(b_{m,n,0}^*, b_{m,n,1}^*)$, 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 so 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 $\theta \in \Omega(\theta^*, 2r)$, we have $F'(\theta) \in \mathcal{L}_{m,n}(b_{m,n,0}, b_{m,n,1})$, where

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

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

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

Since $\sum_{k \ne l, a} e^{(k+l+a)(\alpha_i + \beta_j)} \le \left(\sum_{a=0}^{q-1} e^{a(\alpha_i + \beta_j)}\right)^3$, 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 Equation (18); hence Equation (13) is satisfied. By Theorem 1, the uniform consistency of $\widehat{\theta}$ 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 1, the moment estimator $\widehat{\theta}$ exists and is unique in $\{\theta : q_{m,n} \leq \alpha_i + \beta_j \leq Q_{m,n},\ i \in [m], j \in [n]\}$. Furthermore, we have*

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

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 $V = F'(\theta) \in \mathcal{L}_{m,n}\left(b_{m,n,0}, b_{m,n,1}\right)$, such that

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

By the central limit theorem for the bounded case (Loève, 1977; 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 the required conditions have been checked, we apply Theorem 2 and obtain the following result.

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 $\widehat{\theta} - \theta^*$ 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 values in $\Lambda = \mathbb{N}_0$ and each follows an edge-specific Poisson distribution

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

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

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

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

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

Thus $F'(\theta^*) \in \mathcal{L}_{m,n}\left(b_{m,n,0}^*, b_{m,n,1}^*\right)$, 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 $\theta \in \Omega(\theta^*, 2r)$, we have $F'(\theta) \in \mathcal{L}_{m,n}(b_{m,n,0}, b_{m,n,1})$, 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 and 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 Equation (20). Now Condition (1) is verified. Also, Condition (2) is satisfied since the Poisson distribution is sub-exponential. By Theorem 1 we have the following result.

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 $\widehat{\theta}$ exists and is unique in $\{\theta : q_{m,n} \le \alpha_i + \beta_j \le Q_{m,n}, i \in [m], j \in [n]\}$. Furthermore, this unique solution satisfies*

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

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

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

and

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

If $e^{2Q_{m,n}-q_{m,n}} = o\left(n^{1/4}\right)$, the above expression goes to zero. This satisfies the conditions of the Lyapunov central limit theorem (Billingsley, 1995; page 362). Therefore, $v_{i,i}^{-1/2}\left(d_i - \mathbb{E}(d_i)\right)$ is also asymptotically standard normal when $e^{2Q_{m,n}-q_{m,n}} = o\left(n^{1/4}\right)$. Similarly,

$v_{m+j,m+j}^{-1/2}\left(b_j - \mathbb{E}(b_j)\right)$ 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 Rayleigh density function

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

To verify Condition (1), notice that

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

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

$$\mu'\left(\alpha_i + \beta_j\right) = \frac{1}{2}\sqrt{\frac{\pi}{2}} e^{(\alpha_i+\beta_j)/2},$$

$$\frac{1}{2}\sqrt{\frac{\pi}{2}} e^{(q_{m,n}-2r)/2} \leq \left|\mu'\left(\alpha_i + \beta_j\right)\right| \leq \frac{1}{2}\sqrt{\frac{\pi}{2}} e^{(Q_{m,n}+2r)/2},$$

$$\frac{1}{2}\sqrt{\frac{\pi}{2}} e^{q_{m,n}/2} \leq \left|\mu'\left(\alpha_i^* + \beta_j^*\right)\right| \leq \frac{1}{2}\sqrt{\frac{\pi}{2}} e^{Q_{m,n}/2}.$$

Thus $F'(\boldsymbol{\theta}^*) \in \mathcal{L}_{m,n}\left(b_{m,n,0}^*, b_{m,n,1}^*\right)$, 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 the Supplementary Material, we can choose

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

where c and 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 Equation (22). Also Condition (2) holds because the distribution is sub-Gaussian, thus also sub-exponential. By Theorem 1, we have the following result.

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 1, the moment estimator $\widehat{\theta}$ exists and is unique in* $\{\theta : q_{m,n} \le \alpha_i + \beta_j \le Q_{m,n}, \ i \in [m], j \in [n]\}.$ *Furthermore, this unique solution satisfies*

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

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 $U = \text{Cov}\{\mathbf{g} - \mathbb{E}(\mathbf{g})\} \in \mathcal{L}_{m,n}(\eta_{m,n,4}, \eta_{m,n,5})$, 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

$$\frac{4-\pi}{2} n e^{q_{m,n}} \le u_{i,i} \le \frac{4-\pi}{2} m e^{Q_{m,n}}, \quad i = 1, \ldots, m+n.$$

The third moment of the Rayleigh distribution with parameter $e^{\frac{\alpha_i+\beta_j}{2}}$ is $\mathbb{E}\left(a_{i,j}^3\right) = 3\sqrt{\frac{\pi}{2}} e^{\frac{3(\alpha_i+\beta_j)}{2}}$. Recall that $0 < q_{m,n} \le \alpha_i^* + \beta_j^* \le Q_{m,n}$; it follows that

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

and

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

If $e^{3Q_{m,n}-q_{m,n}} = o\left(n^{1/4}\right)$, the above expression goes to zero. This satisfies the conditions of the Lyapunov central limit theorem (Billingsley, 1995; page 362). Therefore $v_{i,i}^{-1/2}\left(d_i - \mathbb{E}(d_i)\right)$ is also asymptotically standard normal when $e^{3Q_{m,n}-q_{m,n}} = o\left(n^{1/4}\right)$. Similarly, $v_{m+j,m+j}^{-1/2}\left(b_j - \mathbb{E}(b_j)\right)$ 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 following asymptotic normality result for $\widehat{\theta}$.

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 $\widehat{\theta} - \theta^*$ 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 now 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 $x_{i,j}, i \in [m], j \in [n]$, is a mutually independent exponential random variable with the density function

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

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

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

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

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

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

Thus $-F'(\theta^*) \in \mathcal{L}_{m,n}\left(b_{m,n,0}^*, b_{m,n,1}^*\right)$, 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, it follows that

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

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 $\theta \in \Omega(\theta^*, 2r)$, we have $-F'(\theta) \in \mathcal{L}_{m,n}\left(b_{m,n,0}, b_{m,n,1}\right)$, 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 Equation (24). Condition (2) is checked since this distribution is sub-exponential. By Theorem 1, the uniform consistency of $\widehat{\theta}$ 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 1, the moment estimator $\widehat{\theta}$ exists and is unique in $\{\theta : q_{m,n} \leq \alpha_i + \beta_j \leq Q_{m,n}, \ i \in [m], j \in [n]\}$. Furthermore, this unique solution satisfies*

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

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; the covariance matrix of $\{\mathbf{g} - \mathbb{E}(\mathbf{g})\}$ is $V = -F'(\theta) \in \mathcal{L}_{m,n}\left(b_{m,n,0}, b_{m,n,1}\right)$. The third moment of the exponential random variable with parameter $1/(\alpha_i + \beta_j)$ is $\mathbb{E}\left(a_{i,j}^3\right) = 6/\left(\alpha_i + \beta_j\right)^3$. Under the assumption that $0 < q_{m,n} \leq \alpha_i^* + \beta_j^* \leq Q_{m,n}$, we have

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

and

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

If $Q_{m,n}/q_{m,n} = o\left(n^{1/6}\right)$, then the above expression goes to zero. This shows that the condition for the Lyapunov central limit theorem (Billingsley, 1995; page 362) holds. Therefore, $v_{i,i}^{-1/2}$ $\left(d_i - \mathbb{E}(d_i)\right)$ is also asymptotically standard normal if $Q_{m,n}/q_{m,n} = o\left(n^{1/6}\right)$. Similarly $v_{m+j,m+j}^{-1/2}$ $\left(b_j - \mathbb{E}(b_j)\right)$ is also asymptotically standard normal under the same conditions. If $Q_{m,n}/q_{m,n} = o\left(\frac{n^{1/18}}{m^{1/36}(\log m)^{1/18}}\right)$, then

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

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

Corollary 8. *If* $Q_{m,n}/q_{m,n} = o\left(\frac{n^{1/18}}{m^{1/36}(\log m)^{1/18}}\right), m/n = O(1)$, *then for any fixed* $k \ge 1$, *as* $m \to \infty$, *the vector consisting of the first* k *elements of* $\widehat{\theta} - \theta^*$ *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 report the results of numerical studies involving synthetic data generated by the Poisson model specified in Equation (19), in order 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 describe the analysis of an example from the US Law Firms and World Cities network.

5.1. Simulations

We set $\alpha_i^* = (m - i)L/(m - 1)$, $\beta_i^* = (n - i)L/(n - 1)$ and $\beta_n^* = 0$. We chose to fix the bipartite network size at $(m, n) = (100, 50)$, $(200,100)$ and test L on the logarithmic scale, that is, $L \in \{0, \log(\log m), \log m\}$. We focussed on the asymptotic distributions of $\widehat{\xi}_{i,j} = \left[\widehat{\alpha}_i - \widehat{\alpha}_j - (\alpha_i^* - \alpha_j^*)\right]/\left(1/\widehat{v}_{i,i} + 1/\widehat{v}_{j,j}\right)^{1/2}$ and $\widehat{\eta}_{i,j} = \left[\widehat{\beta}_i - \widehat{\beta}_j - (\beta_i^* - \beta_j^*)\right]/\left(1/\widehat{v}_{n+i,n+i} + 1/\widehat{v}_{m+j,m+j}\right)^{1/2}$, where $\widehat{v}_{i,i}$ is the estimate of $v_{i,i}$ which was obtained by replacing θ_i^* with $\widehat{\theta}_i$, and then verifying Corollary 4 empirically. We assessed the asymptotic normality of $\widehat{\xi}_{i,j}, \widehat{\eta}_{i,j}$ by Q–Q plot under various L. We also evaluated the coverage probabilities and the lengths of the empirical 95% confidence intervals. In addition, we recorded the observed frequency that the estimate does not exist. Each simulation involved 10,000 repetitions.

We tested the bipartite network size of $(m, n) = (100, 50)$ and $(200,100)$, respectively, and found the Q–Q plots of $\alpha_i^* - \alpha_j^*$ and $\beta_i^* - \beta_j^*$ to be similar for each setting. Therefore, in Figure 1, we only show the Q–Q plot of $\alpha_i^* - \alpha_j^*$ for $(m, n) = (100, 50)$. 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 no evidence of departure from normality when $L \le \log(m)$.

Table 1 summarizes the observed results for estimating $\theta_i^* - \theta_j^*$, including estimated coverage probabilities, the length of the empirical 95% confidence intervals and the observed frequency that such a confidence interval does not exist. Here, a large L regulates the confidence interval
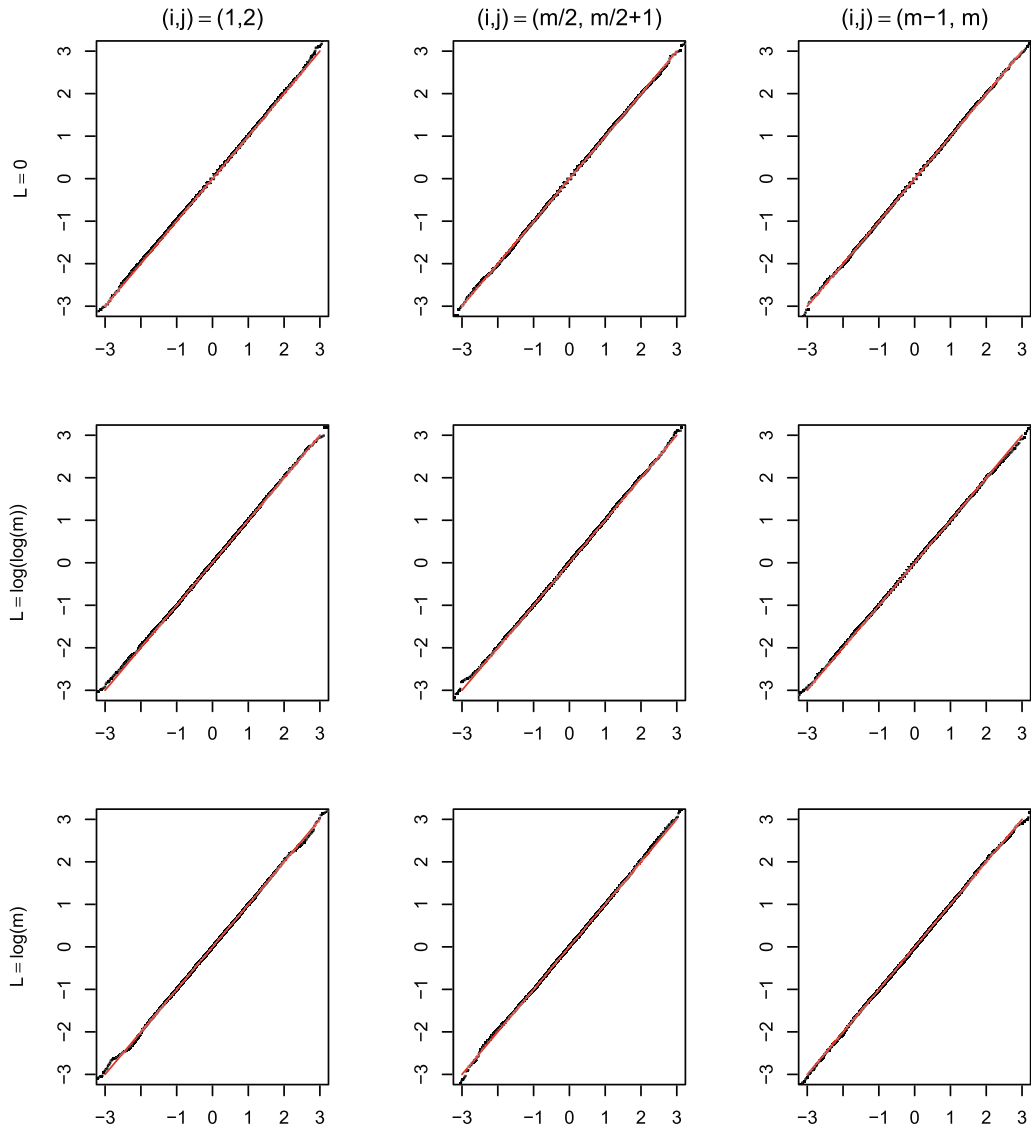
FIGURE 1: The Q–Q plots of $\hat{v}_{i,i}^{1/2}(\hat{\alpha}_i - \alpha_i)$. Poisson model ($m = 100$, $n = 50$).

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 of 95% for all L. Furthermore, we use a Lilliefors test to verify whether our data sample is drawn from a normally distributed population; see Table 2. When $L = 0$, $p = 0.00$ for all pairs (i, j), thereby implying 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, implying that the empirical distribution may reasonably be regarded to follow the normal distribution.

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. Using a Newton method, the memory cost of our approach is $O((m + n)^2)$.

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 with 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)

5.2. The US Law Firms and World Cities Example

We used our method to analyze data concerning the US Law Firms and World Cities network discussed in Beaverstock, Smith & Taylor (2000); the dataset may be found at https://www.lboro.ac.uk/gawc/datasets/da5_1.html. The data concern numbers of lawyers of 100 American law firms with foreign offices in 72 cities outside United States. We preprocessed the data by removing isolated nodes and obtained a bipartite graph with 98 firms and 68 cities. We used the GLM package with a Poisson link function to estimate the Poisson model parameters. Table 4 reports 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 $\hat{\alpha}$ and $\mathbf{d}$, and between $\hat{\beta}$ and $\mathbf{b}$ as well. The full version of this table may be found in the Supplementary Material. Table 5 summarizes the estimated quantiles for firm and city degrees and shows that $\mathbf{d}$ varies markedly from 1 to 1802, whereas the range of $\mathbf{b}$, from 1 to 761, is somewhat more confined. This is also reflected in the spread of $\hat{\alpha}_i$, ranging from -5.01 to 2.48, and $\hat{\beta}_j$ from -3.33 to 3.30, as is shown in Figure 2. The histograms of both $\hat{\alpha}_i$'s and $\hat{\beta}_j$'s indicate that they follow

TABLE 4: US Law Firms and World Cities network dataset: The estimators of $\hat{\alpha}_i$ and $\hat{\beta}_j$ and their standard errors (in parentheses), and selected top 10 and bottom 10 firms and cities according to the degree sequences, respectively.

Firm ID	Degree	$\hat{\alpha}_i$	City ID	Degree	$\hat{\beta}_j$
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)
⋮	⋮	⋮	⋮	⋮	⋮
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.

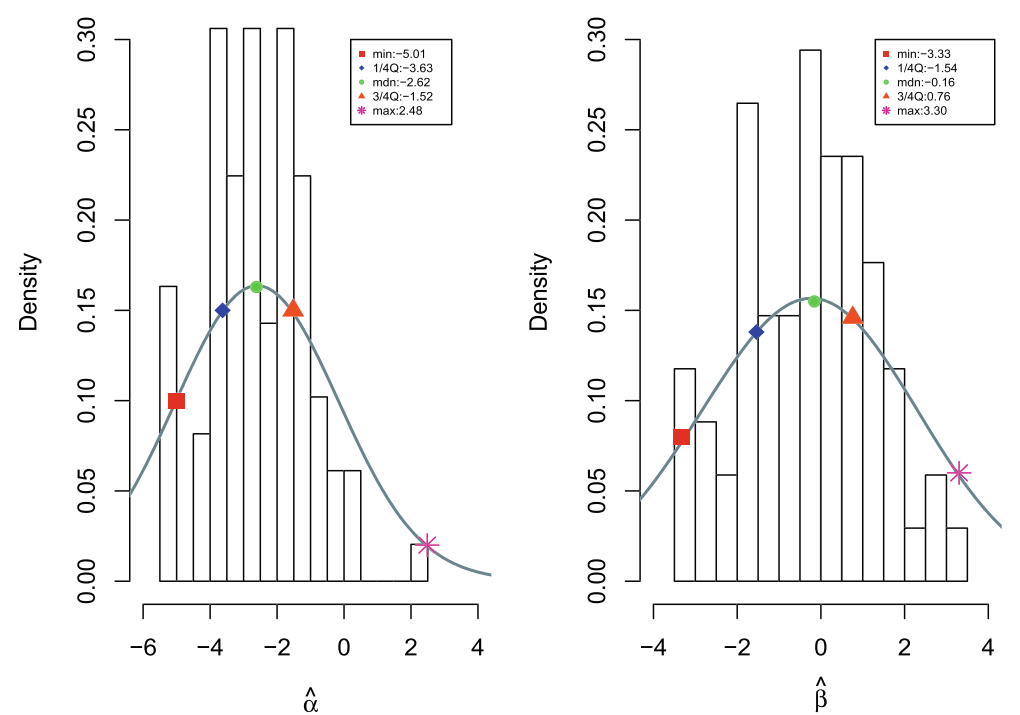| Degree	Minimum	1/4 quantile	Median	3/4 quantile	Maximum
d	1	4	11	33	1802
b	1	6	24	59	761

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

approximate normal distributions. We also give the Q–Q plots of both $\widehat{\alpha}_i$'s and $\widehat{\beta}_j$'s 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 lie close to the reference lines $y = x$, especially in the middle, but the points in the tails are somewhat more variable. We see that both $\widehat{\alpha}_i$ and $\widehat{\beta}_j$ appear to follow normal distributions.

As an application of our method, we used the estimated parameters to generate 100 bootstrapped adjacency matrices. The resulting bootstrapped adjacency matrices were then 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 data concerning degrees from the observed network dataset. Moreover, these original degrees belong to the 95% bootstrap confidence intervals. On the other hand, we also calculated the mean skewness of 100 bootstrap degree distributions and the skewness of each bootstrap degree distribution, respectively; see Table 7. Compared with the skewness of the original degree distributions, the skewness values for the bootstrap degree distributions seem quite similar to each other.
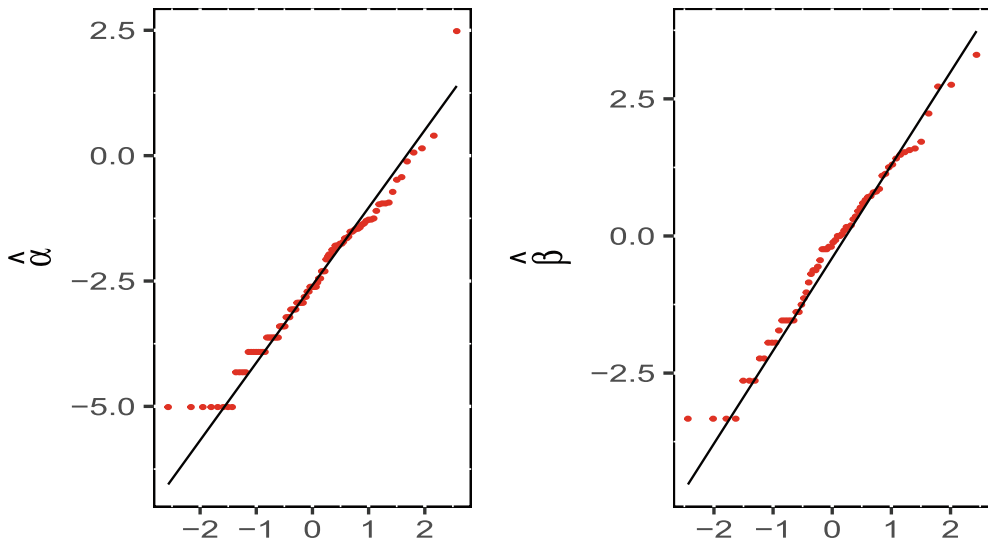
FIGURE 3: The Q–Q plots of $\hat{\alpha}_i$'s and $\hat{\beta}_j$'s for the US Law Firms and World Cities data with 98 firms and 68 cities. The Q–Q plots of the estimates of the 98 firms (left) and 68 cities (right) parameters in the US Law Firms and World Cities data.

Referring to the observed values reported in Table 4, we make the following observations concerning this example. 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 the corresponding estimated standard errors for 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 $\left(\hat{\alpha}_i \pm 2\widehat{s.e.}(\hat{\alpha}_i)\right)$. Let us inspect the top-10 highest degree firms listed in Table 4. A large value of $\hat{\alpha}_i$ suggests that the company is highly "internationalized." Similarly, a large value of $\hat{\beta}_j$ suggests the city's attraction for US law firms. Tracing the top-ranked cities in Table 4, we find the top three to be London, Paris and Hong Kong, all of which ranked in the top tiers of 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 for US law firms, there may be other factors that also play a role in the choices that law firms finally make. For instance, two potentially very relevant other factors are law systems and geographical distance. The United Kingdom, Hong Kong and Singapore share the same law system (common law) as the United States, which might facilitate their connection in legal practice and business. This might also partially explain the situation of Tokyo, since the Japanese legal 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 their ranking in Table 4 compared with their global city indices, possibly due to their geographical and cultural closeness to the United States, compared with some other Asian-Pacific cities with high global centrality. On the other hand, the estimated $\hat{\alpha}_i$ values reported in the same table may serve as a useful reminder of the possible limitations of our study. The most internationalized company (Firm 45) has almost twice as many network connections as the total connections of the remaining top 10 law firms combined. Consequently, this firm alone strongly influences the estimated values of the β_j parameters, and one should keep this fact in mind when interpreting the $\hat{\beta}_j$ values.

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 10 and bottom 10 firms and cities in order, respectively.

Firm ID	d	$\overline{d}$	95% bootstrap CI	City ID	b	$\overline{b}$	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]
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
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]

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 article, we proposed a general model for degree-based modelling and analysis of bipartite graphs. Our model generalizes several popular models in the existing literature that adopt this particular perspective for bipartite graphs. 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 original use of the theory on Newton's iterations in 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 simulation studies that we conducted for the Poisson model and the practical example that we analyzed concerning US law firms and world cities amply demonstrated the effectiveness of our method in concrete problem settings.

There are several possible directions for future research. One of these would involve seeking finite-sample error bound guarantees rather than asymptotic results. The analysis method is not very challenging, but the formulation will become much more involved; therefore, we chose not to pursue such a challenging goal in this article.

Another interesting question would be 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 analyses and results. While we believe such an extension is feasible, it represents a separate line of investigation and exceeds the scope of this article.

The asymptotic approximation and concentration results that our analysis relies on can in fact accommodate slight dependency across network edges. In light of recent research interest in the topic of 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 topic of inquiry concerns richer network features beyond degrees. In fact, degrees can be viewed as a rescaled version of the simplest *network moment*, namely, edges in Zhang & Xia (2022). Other motifs such as *stars* and *cycles* are also useful and very meaningful quantities to study. However, methods and theory for degree-based exponential random graph models are still in active development and not yet complete. In this article, our aim has been to provide a more comprehensive understanding of relatively simple, degree-based models as a solid step forward. Also, while including more statistics into the exponential random graph model, one must also be very careful with model identifiability, which is generally a challenging task for many network models.

ACKNOWLEDGEMENTS

REFERENCES

Aksoy, S. G., Kolda, T. G., & Pinar, A. (2017). Measuring and modeling bipartite graphs with community structure. *Journal of Complex Networks*, 5(4), 581–603.

Albert, R. & Barabási, A. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 47–97.

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

Beaverstock, J. V., Smith, R. G. J., & Taylor, P. J. (2000). Geographies and globalization: U. S. law firms in world cities. *Urban Geography*, 21(2), 105–117.

Billingsley, P. (1995). *Probability and Measure*, 3rd ed., Wiley, New York.

Chatterjee, S., Diaconis, P., & Sly, A. (2011). Random graphs with a given degree sequence. *The Annals of Applied Probability*, 21(4), 1400–1435.

Hillar, C. & Wibisono, A. (2013). Maximum entropy distributions on graphs. arXiv preprint, arXiv:1301.3321. Available at: http://arxiv.org/abs/1301.3321

Holland, P. W. & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373), 33–50.

Latapy, M., Magnien, C., & Del Vecchio, N. (2008). Basic notions for the analysis of large two-mode networks. *Social Networks*, 30(1), 31–48.

Loève, M. (1977). *Probability Theory I*, 4th ed., Springer, New York.

Newman, M. E., Strogatz, S. H., & Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2), 026118.

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

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

Snijders, T. A., Lomi, A., & Torló, V. J. (2013). 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.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*, Vol. 47, Cambridge University Press, Cambridge.

Yan, T., Leng, C., & Zhu, J. (2016). Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *The Annals of Statistics*, 44, 31–57.

Yan, T., Qin, H., & Wang, H. (2016). Asymptotics in undirected random graph models parameterized by the strengths of vertices. *Statistica Sinica*, 26(1), 273–293.

Yan, T. & Xu, J. (2013). A central limit theorem in the β−model for undirected random graphs with a diverging number of vertices. *Biometrika*, 100(2), 519–524.

Zhang, Y., Qian, X., Qin, H., & Yan, T. (2017). Affiliation networks with an increasing degree sequence. *Communications in Statistics-Theory and Methods*, 46(22), 11163–11180.

Zhang, Y., Wang, Q., Zhang, Y., Yan, T., & Luo, J. (2021). $L − 2$ regularized maximum likelihood for β−model in large and sparse networks. arXiv preprint, arXiv:2110.11856.

Zhang, Y. & Xia, D. (2022). Edgeworth expansions for network moments. *The Annals of Statistics*, 50(2), 726–753.