Annals of the Institute of Statistical Mathematics (2023) 75:369–392
https://doi.org/10.1007/s10463-022-00848-0

Asymptotic theory in network models with covariates and a growing number of node parameters

Qiuping Wang[1] · Yuan Zhang[2] · Ting Yan[3]

Received: 28 October 2020 / Revised: 22 April 2022 / Accepted: 13 July 2022 /
Published online: 2 September 2022

Abstract
We propose a general model that jointly characterizes degree heterogeneity and homophily in weighted, undirected networks. We present a moment estimation method using node degrees and homophily statistics. We establish consistency and asymptotic normality of our estimator using novel analysis. We apply our general framework to three applications, including both exponential family and non-exponential family models. Comprehensive numerical studies and a data example also demonstrate the usefulness of our method.

Keywords β-Model · Degree heterogeneity · Network homophily · Network method of moments

Qiuping Wang and Yuan Zhang have contributed equally to this work.

✉ Ting Yan
tingyanty@mail.ccnu.edu.cn

Qiuping Wang
qp.wang@mails.ccnu.edu.cn

Yuan Zhang
yzhanghf@stat.osu.edu

1 School of Mathematics and Statistics, Zhaoqing University, Zhaoqing Avenue, Duanzhou District, Zhaoqing 526061, Guangdong, China

2 Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH 43210, USA

3 Department of Statistics, Central China Normal University, 152 Luoyu Road, Wuhan 430079, Hubei, China

1 Introduction

Jointly modeling a network and nodal or edge-wise coviarates has long been an interesting problem. One natural idea is to extend a widely used network model to incorporate covariates. For example, Yang et al. (2013), Zhang et al. (2016) and Binkiewicz et al. (2017) introduced nodal covariates into a stochastic block model (SBM), which captures the clustering structure in networks. In this paper, we will study the extension of another network model, called β-model (Rinaldo et al., 2013; Yan and Xu, 2013; Yan et al., 2016b, 2019; Chen et al., 2021; Zhang et al., 2021) that characterizes a different important aspect of network data, namely, *degree heterogeneity* (Cho et al., 2011). The degree (total number of connections) of a node provides important profiling information about its structural role in the network (Borgatti and Everett, 2000; Zhang and Xia, 2022; Maugis, 2020). A famous example is that (Babai et al. 1980) shows that efficient graph matching can usually succeed with high probability between two shuffled random graphs, using a degree-based algorithm. The β-model, named by Chatterjee et al. (2011), is an undirected, binary network:

$$\begin{aligned} \mathbb{P}(a_{ij} = 1) &= \frac{e^{\beta_i+\beta_j}}{1 + e^{\beta_i+\beta_j}}, \\ a_{ij} &= a_{ji}, 1 \le i < j \le n \end{aligned} \tag{1}$$

where $A = (a_{ij}), 1 \le \{i,j\} \le n$ is a binary adjacency matrix. Later, Yan et al. (2016b) and Fan et al. (2022+) extend this model to weighted networks with edge distributions including Poisson, geometric, exponential and so on. This paper will generalize (Yan et al., 2016b) rather than the original binary edge β-model.

What we incorporate into a weighted β-model are *edge-wise covariates*. We notice that this set up also accommodates nodal covariates (e.g., immutable characteristics such as gender, race and genetic features; and/or mutable ones, including location, occupation and hobbies) since they can be easily transformed into edge-wise similarity/dissimilarity measures. According to Graham (2017), this part of the data encodes the *homophily* effect in network formation. As a quick illustration, consider two node pairs (i_1, i_2) and (j_1, j_2). Even if $\{\beta_{i_1}, \beta_{i_2}\}$ are very different from $\{\beta_{j_1}, \beta_{j_2}\}$, their edge expectations $\mathbb{E}[a_{i_1,i_2}]$ and $\mathbb{E}[a_{j_1,j_2}]$ might not differ too much, if they have similar edge covariates $z_{i_1,i_2} \approx z_{j_1,j_2}$.

Jointly modeling both degree heterogeneity and homophily, as well as developing effective estimation and inference methods along with supporting theory, is an interesting challenge. As aforementioned, covariate-assisted stochastic block models have been comparatively well-studied, whereas few works exist to extend the β-model. Among notable exceptions, Graham (2017) generalizes (1) by extending $\beta_i + \beta_j$ to $\beta_i + \beta_j + z_{ij}^T\gamma$ and devises a likelihood-based method for estimation and inference. Recently, independent works (Stein and Leng, 2020, 2021) further introduce ℓ_1 regularization to the joint model for undirected and directed networks, respectively. Both research groups (Graham, 2017; Stein and Leng, 2020, 2021) focus exclusively on binary edges for cleanness. On the other hand, many networks, such as communications, co-authorship, brain activities and others, have weighted edges.

In this paper, we develop a general joint model for weighted edges. Different from the likelihood-based approaches in Graham (2017) and Stein and Leng (2020, 2021), we propose and analyze a method-of-moment parameter estimation. As discussed later in the paper, moment method has its unique advantage in addressing slightly dependent network edge formation—despite this paper exclusively focuses on independent edge generation, we understand that a comprehensive study here paves the road toward successfully handling the very challenging problem of dependent edges.

We develop a two-stage Newton method that first finds an error bound for $\|\widehat{\beta}-\beta\|_\infty$ for a fixed γ via establishing the convergence rate of the Newton iterative sequence and then derives $\|\widehat{\gamma}-\gamma\|_\infty$ based on a profiled equation under some conditions. When all parameters are bounded, the ℓ_∞ norm error for $\widehat{\beta}$ is in the order of $O_p(n^{-1/2})$ while the ℓ_∞ norm error for $\widehat{\gamma}$ is in the order of $O_p(n^{-1})$, both up to a logarithm factor. When the parameters diverge, the error bounds depend on additional factors involved with the ranges of β and γ. Further, we derive an asymptotic representation of the moment estimator, based on which, we derive their asymptotic normal distributions under classical CLT conditions. To illustrate the unified results, we present three applications, along with comprehensive numerical simulations and a real-data example.

The rest of the paper is organized as follows. In Sect. 2, we present our general model. In Sect. 3, we propose our moment estimation equations. In Sect. 4, we establish consistency and asymptotic normality of our estimator under mild conditions. Section 5 illustrates the application of our general framework to weighted networks with logistic, Poisson and probit edge formation schemes. Section 6 contains summary and discussion. Due to limited space, simulation results and the real data application are relegated to Supplementary Material.

2 Covariate-assisted β-model

We shall jointly analyze data from two sources: network and edge-wise covariates. The network data are represented by an adjacency matrix $A=(a_{ij})_{n\times n}$, $1\le i<j\le n$. We study undirected networks without self-loops, i.e., A is symmetric $a_{ij}=a_{ji}$ and $a_{ii}=0$. In this paper, each entry a_{ij} may be binary or weighted (such as collaboration counts in a co-authorship network and phone call lengths). Let $d_i=\sum_{j\neq i}a_{ij}$ be the degree of node i and $d=(d_1,\dots,d_n)^T$ be the degree sequence. In addition to network data, we also observe a covariate vector $z_{ij}\in\mathbb{R}^p$ on each edge. This setting also covers the scenario when we observe nodal attributes x_i: simply define a similarity/dissimilarity measure $g(\cdot,\cdot)$ that converts these attributes to edge-wise covariates via $g(x_i,x_j)$. Examples including Euclidean distance for continuous x_i's and Hamming distance for binary x_i's.

Our goal is to jointly model *degree heterogeneity* and *homophily*. Degree heterogeneity is captured by a latent parameter $\beta_i\in\mathbb{R}$ on each node. Homophily is driven by edge-wise covariates under our framework. Specifically, it is accounted for by $z_{ij}^T\gamma$, where the exogenous parameter $\gamma \in \mathbb{R}^p$ can be understood analogously like a regression coefficient in a generalized linear model.

Now we present our model. Given $Z = (z_{ij})$, the network entries a_{ij} are generated independently by the following model, which we call "covariate-assisted β-model":

$$a_{ij} | \{z, \beta, \gamma\} \sim f\left(a_{ij} | \beta_i + \beta_j + z_{ij}^T \gamma\right). \tag{2}$$

where f is a known probability density /mass function, β_i is the degree parameter of node i and γ is a p-dimensional regression coefficient for the covariate z_{ij}. Our model (2) generalizes the semi-parametric models in econometrics literature (Fernández-Vál and Weidner, 2016) with binary and exponential responses for undirected networks. We focus on these additive models for computational tractability. It would be an interesting future work to generalize the method of our analysis to address the more general case where β, z and γ enter the model as non-additive effects.

The model (2) extends not only the well-known β-model, but also many of its variants (Yan et al., 2016b). In many examples, such as β-model, $f(\cdot)$ is an increasing function of β_i. Consequently, nodes having relatively large degree parameters will have more links than those nodes with low degree parameters, without considering homophily. To further illustrate the usefulness of (2), we consider two running examples.

Example 1 (Binary edges) In statistical network analysis, a long-studied problem is to jointly model network data with additional covariates. For example, Yang et al. (2013) and Zhang et al. (2016) incorporate nodal covariates into a stochastic block model; in contrast our model provides a flexible tool for incorporating nodal and/or edge-wise covariates into a β-model. Here, we consider binary edges, i.e. $a_{ij} \in \{0, 1\}$. Let F be some properly chosen transformation: $F : \mathbb{R} \to [0, 1]$. The probability of a_{ij} is

$$\mathbb{P}(a_{ij} = a) = \{F(\beta_i + \beta_j + z_{ij}^T \gamma)\}^a \{1 - F(\beta_i + \beta_j + z_{ij}^T \gamma)\}^{1-a}, \quad a \in \{0, 1\}.$$

Two popular choices of $F(\cdot)$ are sigmoid transformation $F(x) = e^x/(1 + e^x)$ (Graham, 2017) and probit transformation $F(x) = \Phi(x)$, where $\Phi(x)$ is the CDF of $N(0, 1)$.

Example 2 (*Unbounded discrete edges*) This example generalizes the Poisson β-model in Section 3.4 of Yan et al. (2016b). Here, we model $a_{ij} \sim \text{Poisson}\left(\lambda_{ij}\right)$, where $\lambda_{ij} = \exp(\beta_i + \beta_j + z_{ij}^T \gamma)$. That is,

$$\begin{aligned} \log \mathbb{P}(a_{ij} = a) = & a(\beta_i + \beta_j + z_{ij}^T \gamma) \\ & - \exp(\beta_i + \beta_j + z_{ij}^T \gamma) - \log a!. \end{aligned}$$

3 Parameter estimation

To motivate the estimation method for our covariate-assisted β-model (2), it is helpful to very briefly recall that for the classical β-model, i.e., $\gamma \equiv 0$, there are two mainstreams of estimation methods: MLE (Chatterjee et al., 2011; Rinaldo et al., 2013) and method of moments (Yan et al., 2016b). While MLE is a widely recognized method for model parameter estimation, as aforementioned in the introduction, method-of-moments may enjoy an easier extension to dependent-edge scenarios in the future (see also Sect. 6). Therefore, we propose and study a moment estimator in this paper. Notice that $\mathbb{E}[a_{ij}]$ depends on the model parameters only through

$$\pi_{ij} := \beta_i + \beta_j + z_{ij}^T \gamma. \tag{3}$$

There exists a function $\mu(\cdot)$ such that $\mathbb{E}[a_{ij}] = \mu(\pi_{ij})$. In some future texts, we find it more convenient to emphasize that $\mu(\pi_{ij})$ can be viewed as the (i, j) element of an $n \times n$ matrix. Therefore, we slightly abuse notation and might sometimes use a different notation $\mu_{ij}(\beta, \gamma)$ to represent $\mu(\pi_{ij})$.

Now we are ready to present our method of moments parameter estimation. To motivate our formulation, consider a special case when the edge distribution belongs to an exponential family, namely,

$$L(a|\beta, z, \gamma) = C(\beta, z, \gamma) \cdot e^{\left(\sum_{1\le i<j\le n} a_{ij}\cdot(\beta_i+\beta_j+z_{ij}^T\gamma) \right)} \cdot h(a, z). \tag{4}$$

Examples of (4) include logistic model and Poisson model with covariates, as in Sect. 5. Sending the partial derivatives of the log-likelihood function of (4) to zero, we obtain the following moment equations.

$$d_i = \sum_{j:j\neq i} \mu_{ij}(\beta, \gamma), \quad i \in [n] \tag{5}$$

$$\sum_{1\le i<j\le n} z_{ij} a_{ij} = \sum_{1\le i<j\le n} z_{ij} \mu_{ij}(\beta, \gamma) \tag{6}$$

where $[n] = \{1, \dots, n\}$. Denote the solution to (5) and (6) by $(\widehat{\beta}, \widehat{\gamma})$. We shall address the natural questions of existence and uniqueness of $(\widehat{\beta}, \widehat{\gamma})$ in Theorem 1.

Now we discuss some computational issues. When the number of nodes n is small and f is the binomial, Probit, or Poisson probability function or Gamma density function, we can simply use the package "glm" in the R language to solve (5) and (6). For relatively large n, it might not have large enough memory to store the design matrix for β required by the R package "glm". In this case, we recommend the use of a two-step iterative algorithm by alternating between solving the first equation in (5) via the fixed point method (Chatterjee et al., 2011) or the gradient descent algorithm (Bubeck, 2015) and solving the second equation in (6).

4 Asymptotic properties

In this section, we present the consistency and asymptotic normality of the moment estimator. We start with notation. For any $C \subset \mathbb{R}^n$, let C^0 and $\overline{C}$ denote the interior and closure of C, respectively. For a vector $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$, let $\|x\|$ be a generic notation for vector norm. Specifically, inherit the notion of $\|x\|_p$ to denote ℓ_p norm from functional analysis. Let $B_\infty(x, \epsilon) = \{y : \|x - y\|_\infty \le \epsilon\}$ be an ϵ-neighborhood of x under ℓ_∞ metric. For an $n \times n$ matrix $J = (J_{i,j})$, let $\|J\|_\infty$ denote the matrix norm induced by the ℓ_∞-norm on vectors in $\mathbb{R}^n$, i.e.,

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

$\|J\|_{\max} = \max_{i,j} |J_{ij}|$, and let $\|J\|$ be a generic notion for matrix norm. We use a superscript "*" to mark the true parameter. But in early sections of this paper, without causing ambiguity, we might omit it when stating the model.

Next, we set up regularity conditions for our main theorems. Assume $\mu(\cdot)$ has continuous third derivative. Recall $\pi_{ij} = \beta_i + \beta_j + z_{ij}^T \gamma$ as defined in (3). Suppose there exist $b_{n0}, b_{n1}, b_{n2}, b_{n3} > 0$ such that

$$\min_{i,j} \mu'(\pi_{ij}) \cdot \max_{i,j} \mu'(\pi_{ij}) > 0, \tag{7a}$$

$$b_{n0} \le \min_{i,j} |\mu'(\pi_{ij})| \le \max_{i,j} |\mu'(\pi_{ij})| \le b_{n1}, \tag{7b}$$

$$\max_{i,j} |\mu''(\pi_{ij})| \le b_{n2}, \tag{7c}$$

$$\max_{i,j} |\mu'''(\pi_{ij})| \le b_{n3}. \tag{7d}$$

hold for all $\beta \in B_\infty(\beta^*, \epsilon_{n1}), \gamma \in B_\infty(\gamma^*, \epsilon_{n2})$, where $\epsilon_{n1}, \epsilon_{n2} > 0$ are two diminishing numbers as $n \to \infty$.

Condition (7a) is mild, requiring that the derivative of the expectation function $\mu(x)$ is positive for all $x \in \mathbb{R}$ or negative. The conditions (7b)–(7d) may seem quite technical and abstract for readers. To help with intuitive understanding, let us illustrate them using Example 1 with a logistic link function. In this case, $\mu(x) = e^x/(1 + e^x)$. Straight calculations show

$$\begin{aligned}
\mu'(x) &= \frac{e^x}{(1+e^x)^2}, \\
\mu''(x) &= \frac{e^x(1-e^x)}{(1+e^x)^3}, \\
\mu'''(x) &= \frac{e^x(1-4e^x+e^{2x})}{(1+e^x)^4}.
\end{aligned}$$

It is easy to verify that $\max\left\{|\mu'(x)|, |\mu''(x)|, |\mu'''(x)|\right\} \le 1/4$, where we used

$$|\mu''(x)| \le \frac{e^x}{(1+e^x)^2}\left|\frac{(1-e^x)}{(1+e^x)}\right| \quad \text{and}$$
$$|\mu'''(x)| = \frac{e^x}{(1+e^x)^2}\left|\frac{(1-e^x)^2}{(1+e^x)^2} - \frac{2e^x}{(1+e^x)^2}\right|.$$

Therefore, in this example, we can set $b_{n1} = b_{n2} = b_{n3} = 1/4$ and

$$b_{n0} = \min_{i,j} \frac{e^{\pi_{ij}}}{(1+e^{\pi_{ij}})^2} \ge \frac{e^{2\|\beta^*\|_\infty + \|\gamma^*\|_1 z_* + 2\epsilon_{n1} + p\epsilon_{n2}}}{(1+e^{2\|\beta^*\|_\infty + \|\gamma^*\|_1 z_* + 2\epsilon_{n1} + p\epsilon_{n2}})^2}, \tag{8}$$

where $z_* := \max_{i,j} \|z_{ij}\|_\infty$.

In addition to the regularity conditions (7a)–(7d) this section, we shall also need the following assumptions:

Assumption 1 (*Bounded covariates*) Suppose $\max_{i,j} \|z_{ij}\|_\infty \le C_z$ holds for some universal constant C_z.

Assumption 2 (Sub-exponential edge distribution) The distribution of $a_{ij} - \mathbb{E}a_{ij}$ is sub-exponential, with parameter h_{ij}. Denote $h_n := \max_{i,j} h_{ij}$.

Assumption 1 is naturally satisfied by some popular dissimilarity measures between nodal covariates, such as Hamming distance. If the observed z_{ij}'s or nodal covariates seem to vary wildly, we can simply apply a transformation such as sigmoid or probit functions to tame them into universally bounded edge covariates, see Sect. 6 of Zhang and Xia (2022). Assumption 2 is satisfied by many popular edge distributions, such as those in Yan et al. (2016b). We make this assumption mostly to make our narration succinct – it can be replaced by any other conditions that guarantee $|d_i - \mathbb{E}d_i| = O_p(n^{1/2})$ and $|\sum_{i<j}(a_{ij} - \mathbb{E}a_{ij})| = O_p(n)$, respectively.

4.1 Consistency

The asymptotic behavior of the estimator $(\widehat{\beta}, \widehat{\gamma})$ critically depends on the curvature of $\mu(\beta, \gamma)$. To study this curvature, we start with setting up some notation. Define

$$F_i(\beta, \gamma) = \sum_{j=1, j\neq i}^{n} \mu_{ij}(\beta, \gamma) - d_i, \quad i = 1, \ldots, n. \tag{9}$$

For simplicity, we write $F(\beta, \gamma) = (F_1(\beta, \gamma), \ldots, F_n(\beta, \gamma))^T$. Also denote $F_{\gamma,i}(\beta)$ to be $F_i(\beta, \gamma)$ for an arbitrary given γ, denote $F_\gamma(\beta) = (F_{\gamma,1}(\beta), \ldots, F_{\gamma,n}(\beta))^T$ and define $\widehat{\beta}_\gamma$ to be the solution to $F_\gamma(\beta) = 0$. Set

$$Q(\beta,\gamma)=\sum_{i<j} z_{ij}(\mu_{ij}(\beta,\gamma)-a_{ij}), \tag{10}$$

$$Q_c(\gamma)= \sum_{i<j} z_{ij}\big(\mu_{ij}(\hat{\beta}_\gamma,\gamma)-a_{ij}\big). \tag{11}$$

By definition, we have the following relationships:

$$\begin{aligned} F(\widehat{\beta},\widehat{\gamma}) =&0, \quad F_\gamma(\widehat{\beta}_\gamma)=0,\\ Q(\widehat{\beta},\widehat{\gamma}) =&0, \quad Q_c(\widehat{\gamma})=0. \end{aligned}$$

Similar to Chatterjee et al. (2011) and Yan et al. (2016a), we define a notion called "$\mathcal{L}$ class matrices" for narration convenience. Given some $M \ge m > 0$, we say an $n\times n$ matrix $V=(v_{ij})$ belongs to the matrix class $\mathcal{L}_n(m,M)$ if V is a diagonally balanced matrix with positive elements bounded by m and M, i.e.,

$$v_{ii}=\sum_{j=1,j\neq i}^{n} v_{ij}, \quad i=1,\dots,n, \quad m\le v_{ij}\le M, \quad i,j=1,\dots,n;i\neq j. \tag{12}$$

Since π is linear in β, for any $1 \le \{i\neq j\} \le n$, we have

$$\begin{aligned} \frac{\partial F_i(\beta,\gamma)}{\partial \beta_i} =&\sum_{j\neq i}\mu'(\pi_{ij}),\\ \frac{\partial F_i(\beta,\gamma)}{\partial \beta_j} =&\mu'(\pi_{ij}). \end{aligned} \tag{13}$$

It is easy to verify that (13) yields that when $\mu'(x)>0$, $\beta\in B_\infty(\beta^*,\epsilon_{n1})$ and $\gamma\in B_\infty(\gamma^*,\epsilon_{n2})$, we have $F'_\gamma(\beta)\in\mathcal{L}_n(b_{n0},b_{n1})$. For simplicity, we assume $F'_\gamma(\beta)\in\mathcal{L}(b_{n0},b_{n1})$ hereafter (if $-F'_\gamma(\beta)\in\mathcal{L}(b_{n0},b_{n1})$, we could rewrite $\widetilde{F}_\gamma(\beta) := -F_\gamma(\beta)$). Define a convenient shorthand

$$V(\beta,\gamma) := F'_\gamma(\beta),$$

and define an abbreviation $V=V(\beta^*,\gamma^*)$. We will establish the consistency of the estimator $\widehat{\beta}_\gamma$ using the theorems of Newton method, for which we shall need an explicity formulation of $F'_\gamma(\beta)$. This inverse does not have a closed form, but fortunately, by mimicking (Simons and Yao, 1999; Yan et al., 2015) proposed a convenient approximate inversion formula $V\in\mathcal{L}_n(m,M)$ by

$$S=\mathrm{diag}(1/v_{11},\dots,1/v_{nn}) \tag{14}$$

at approximate error of $O\big(M^2/(n^2m^3)\big)$ under the matrix maximum norm (i.e., the maximum of all absolute elements of a matrix).

With the above notation preparations, now we commence the asymptotic analysis of the estimator. First, we have

$$\frac{\partial F_\gamma(\widehat{\beta}_\gamma)}{\partial \gamma^T} = \frac{\partial F(\widehat{\beta}_\gamma, \gamma)}{\partial \beta^T}\frac{\partial \widehat{\beta}_\gamma}{\gamma^T} + \frac{\partial F(\widehat{\beta}_\gamma, \gamma)}{\partial \gamma^T} = 0, \tag{15}$$

$$\frac{\partial Q_c(\gamma)}{\partial \gamma^T} = \frac{\partial Q(\widehat{\beta}_\gamma, \gamma)}{\partial \beta^T}\frac{\partial \widehat{\beta}_\gamma}{\gamma^T} + \frac{\partial Q(\widehat{\beta}_\gamma, \gamma)}{\partial \gamma^T}. \tag{16}$$

Combining (15) and (16), the Jacobian matrix $Q_c'(\gamma) = \partial Q_c'(\gamma)/\partial\gamma$ has the following formulation:

$$\frac{\partial Q_c(\gamma)}{\partial \gamma^T} = \frac{\partial Q(\widehat{\beta}_\gamma, \gamma)}{\partial \gamma^T} - \frac{\partial Q(\widehat{\beta}_\gamma, \gamma)}{\partial \beta^T}\left[\frac{\partial F(\widehat{\beta}_\gamma, \gamma)}{\partial \beta^T}\right]^{-1}\frac{\partial F(\widehat{\beta}_\gamma, \gamma)}{\partial \gamma^T}. \tag{17}$$

The asymptotic behavior of $\widehat{\gamma}$ crucially depends on $Q_c'(\gamma)$. But the $\widehat{\beta}_\gamma$ that appears in the definition of $Q_c'(\gamma)$ does not have a closed form. To facilitate the quantitative study of the curvature of $Q_c(\gamma)$, define

$$\begin{aligned} H(\beta, \gamma) =& \frac{\partial Q(\beta, \gamma)}{\partial \gamma^T} \\ &- \frac{\partial Q(\beta, \gamma)}{\partial \beta^T}\left[\frac{\partial F(\beta, \gamma)}{\partial \beta^T}\right]^{-1}\frac{\partial F(\beta, \gamma)}{\partial \gamma^T}, \end{aligned} \tag{18}$$

which can be viewed as a relaxed version of $\partial Q_c(\gamma)/\partial\gamma$. When $\beta \in B_\infty(\beta^*, \epsilon_{n1})$, by Section 10 of Supplemental Material, we have

$$\frac{1}{n^2}\big(H(\beta, \gamma^*)\big)_{ij} = \frac{1}{n^2}\big(H(\beta^*, \gamma^*)\big)_{ij} + o(1), \tag{19}$$

for each given $(i,j) : 1 \le i < j \le n$, where recall that the entries of H are sums $n(n-1)/2$ of terms, thus n^{-2} would be a proper rescaling factor.

We assume $H(\beta, \gamma)$ is positively definite. When a_{ij} belongs to exponential family of distributions, $H(\beta, \gamma)$ is the Fisher information matrix of the concentrated likelihood function on γ (e.g. page 126 of Amemiya, 1985) and is thus positive definite. See also Sect. 5. In fact, the asymptotic variance of $\widehat{\gamma}$ is $H^{-1}(\beta, \gamma)$ when $f(\cdot)$ in (2) is an exponential-family distribution; see the applications in Sect. 5. Thus, the asymptotic behavior of $\widehat{\gamma}$ will be ill-posed without this assumption. Define

$$\kappa_n := \sup_{\beta \in B_\infty(\beta^*, \epsilon_{n1})} \|n^2 \cdot H^{-1}(\beta, \gamma^*)\|_\infty \tag{20}$$

Now we formally state the consistency result.

Theorem 1 *Let* $\sigma_n^2 = n^2\|(V^{-1} - S)\mathrm{Cov}(F(\beta^*, \gamma^*))(V^{-1} - S)\|_{\max}$. *Suppose Assumptions* 1 *and* 2 *and conditions* (7a)–(7d) *hold, and*

$$\frac{\kappa_n^2 b_{n1}^4 b_{n2}}{b_{n0}^3}\left(\frac{b_{n2}h_n^2}{b_{n0}^3}+\sigma_n\right)=o\left(\frac{n}{\log n}\right). \tag{21}$$

Then the moment estimator $\hat{\gamma}$ *exists with high probability, and we further have*

$$\|\hat{\gamma}-\gamma^*\|_\infty = O_p\left(\frac{\kappa_n b_{n1}\log n}{n}\left(\frac{h_n^2 b_{n2}}{b_{n0}^3}+\sigma_n\right)\right)=o_p(1)$$
$$\|\hat{\beta}-\beta^*\|_\infty = O_p\left(\frac{h_n}{b_{n0}}\sqrt{\frac{\log n}{n}}\right)=o_p(1).$$

Our proof of Theorem 1 analyzes a two-stage Newton method and is thus different from Graham (2017) that uses a convergence rate analysis of the fixed point method in Chatterjee et al. (2011).

When $f(\cdot)$ in model (2) is an exponential family distribution, then $V=\mathrm{Cov}(F(\beta^*,\gamma^*))$. In this case, the expression inside the norm of σ_n^2 simplifies into

$$(V^{-1}-S)V(V^{-1}-S)=V^{-1}-S+\frac{v_{ij}(1-\delta_{ij})}{v_{ii}v_{jj}},$$

By Lemma 1, $\|V^{-1}-S\|_{\max}=O(b_{n1}^2 b_{n0}^{-3}n^{-2})$. Thus, $\sigma_n^2=O(b_{n1}^2/b_{n0}^3)$. We have the following corollary.

Corollary 1 *Assume* $V=\mathrm{Cov}(F(\beta^*,\gamma^*))$ *and the conditions of Theorem 1 hold. If*

$$\frac{\kappa_n^2 h_n^2 b_{n1}^5 b_{n2}}{b_{n0}^6}=o\left(\frac{n}{\log n}\right),$$

then

$$\|\hat{\gamma}-\gamma^*\|_\infty = O_p\left(\frac{\kappa_n b_{n1}^2 h_n^2 b_{n2}\log n}{n b_{n0}^3}\right)=o_p(1)$$
$$\|\hat{\beta}-\beta^*\|_\infty = O_p\left(\frac{h_n}{b_{n0}}\sqrt{\frac{\log n}{n}}\right)=o_p(1).$$

When $f(\cdot)$ in model (2) belongs to exponential-family distributions and $\|\beta^*\|_\infty$ and $\|\gamma^*\|_\infty$ are universally bounded, then b_{n0}, b_{n1}, b_{n2} and σ_n are constants. Further, if all covariates are bounded, $H(\beta^*,\gamma^*)/n^2$ is approximately a constant matrix such that κ_n is also a constant. In this case, the conditions in Theorem 1 easily hold. Further, if $b_{n0}, b_{n1}, b_{n2}, \kappa_n, h_n$ are constants, then the convergence rates of $\hat{\beta}$ and $\hat{\gamma}$ are $O_p((\log n/n)^{1/2})$ and $O_p(\log n/n)$, respectively. This reproduces the error bound in Chatterjee et al. (2011). This convergence rate matches the minimax optimal upper bound $\|\hat{\beta}-\beta\|_\infty=O_p((\log p/n)^{1/2})$ for the Lasso estimator in the linear model with a p-dimensional parameter vector β and the sample size n (Lounici, 2008). The convergence rate $O_p(\log n/n)$ for $\widehat{\gamma}$ is very close to the square root rate $N^{-1/2}$ in the classical large sample theory, where $N = n(n-1)/2$.

4.2 Asymptotic normality of $\widehat{\beta}$

We derive the asymptotic expansion format of $\widehat{\beta}$ by applying a second-order Taylor expansion to $F(\widehat{\beta}, \widehat{\gamma})$ and showing that various remainder terms are asymptotically negligible.

Theorem 2 *Assume the conditions of Theorem 1 hold. If*

$$\kappa_n^2 b_{n1}^2 \left(\frac{h_n^2 b_{n2}}{b_{n0}^3} + \sigma_n^2 \right)^2 = o\left(\frac{n}{\log n} \right),$$

then for any fixed i,

$$\widehat{\beta}_i - \beta_i = v_{ii}^{-1}(d_i - \mathbb{E}d_i) + O_p\left(\frac{\kappa_n b_{n1} \log n}{n b_{n0}} \left(\frac{b_{n2} h_n^2}{b_{n0}^3} + \sigma_n \right) \right).$$

Let $u_{ii} = \sum_{j \neq i} \mathrm{Var}(a_{ij})$. If $\sum_{j \neq i} \mathbb{E}(a_{ij} - \mathbb{E}a_{ij})^3 / v_{ii}^{3/2} \to 0$, then, by the Lyapunov's central limit theorem, $u_{ii}^{-1/2}\{d_i - \mathbb{E}(d_i)\}$ converges in distribution to the standard normal distribution. When considering the asymptotic behaviors of the vector $(d_1, \ldots, d_r)$ with a fixed r, one could replace the degrees $d_1, \ldots, d_r$ by the independent random variables $\widetilde{d}_i = a_{i,r+1} + \cdots + a_{in}$, $i = 1, \ldots, r$. Therefore, we have the following lemma.

Proposition 1 *Under the conditions of Theorem 2, if $u_{ii}^{-3/2} \sum_{j: j \neq i} \mathbb{E}(a_{ij} - \mathbb{E}a_{ij})^3 \to 0$, then we have:*

(1) *For any fixed $r \geq 1$, $(d_1 - \mathbb{E}(d_1), \ldots, d_r - \mathbb{E}(d_r))$ are asymptotically independent and normally distributed with mean zero and marginal variances $u_{11}, \ldots, u_{rr}$, respectively.*

(2) *More generally, $\sum_{i=1}^n c_i(d_i - \mathbb{E}(d_i))/\sqrt{u_{ii}}$ is asymptotically normally distributed with mean zero and variance $\sum_{i=1}^\infty c_i^2$ whenever $c_1, c_2, \ldots$ are fixed constants, and $\sum_{i=1}^\infty c_i^2 < \infty$.*

Part (2) follows from part (1) and the fact that

$$\lim_{r \to \infty} \limsup_{t \to \infty} \mathrm{Var}\left(\sum_{k=r+1}^{n} c_i \frac{d_i - \mathbb{E}(d_i)}{\sqrt{u_{ii}}} \right) = 0 \tag{22}$$

by Theorem 4.2 of Billingsley (1995). To see (22), it suffices to show that the eigenvalues of the covariance matrix of $(d_i - \mathbb{E}(d_i))/u_{ii}^{1/2}$, $i = r+1, \ldots, n$ are bounded by

2 for all $r < n$, which is implied by the well-known Perron-Frobenius theorem: if A is a symmetric positive definite matrix with diagonal elements equaling to 1, with nonnegative off-diagonal elements, then its largest eigenvalue is less than 2. In view of Proposition 1, we immediately have the following corollary.

Corollary 2 *Assume that conditions in Theorem 2 hold. If* $u_{ii}^{-3/2}\sum_{j\neq i}\mathbb{E}(a_{ij}-\mathbb{E}a_{ij})^3 \to 0$, *then for fixed k the vector* $(u_{11}^{-1/2}v_{11}(\widehat{\beta}_1-\beta^*),\ldots,u_{kk}^{-1/2}v_{kk}(\widehat{\beta}_k-\beta_k^*))$ *converges in distribution to the k-dimensional multivariate standard normal distribution.*

4.3 Asymptotic normality of $\widehat{\gamma}$

Let $T_{ij}=e_i+e_j$, where $e_i\in\mathbb{R}^n$ is all zero except its ith element equals 1. Define

$$\begin{aligned} V(\beta,\gamma) &= \frac{\partial F(\beta,\gamma)}{\partial\beta^T}, \quad V_{Q\beta}(\beta,\gamma)=\frac{\partial Q(\beta,\gamma)}{\partial\beta^T},\\ s_{ij}(\beta,\gamma) &= (a_{ij}-\mathbb{E}a_{ij})(z_{ij}-V_{Q\beta}(\beta,\gamma)[V(\beta,\gamma)]^{-1}T_{ij}). \end{aligned}$$

When evaluating $H(\beta,\gamma)$, $Q(\beta,\gamma)$, $V(\beta,\gamma)$ and $V_{Q\beta}(\beta,\gamma)$ at their true values (β^*,γ^*), we omit the arguments β^*,γ^*, i.e., $V=V(\beta^*,\gamma^*)$. Recall we earlier defined $N=n(n-1)$. Also define

$$\bar{H}=\lim_{n\to\infty}\frac{1}{N}H(\beta^*,\gamma^*),$$

where we recall the definition of $H(\beta,\gamma)$ from (18). We have

Theorem 3 *Let* $U=\mathrm{Var}(d)$. *Assume the conditions in Theorem 1 hold. If* $b_{n3}h_n^3b_{n0}^{-3}=o(n^{1/2}/(\log n)^{3/2})$, *then we have*

$$\sqrt{N}(\widehat{\gamma}-\gamma^*)=\bar{H}^{-1}B_*+\bar{H}^{-1}\times\frac{1}{\sqrt{N}}\sum_{i<j}s_{ij}(\beta^*,\gamma^*)+o_p(1),$$

where

$$B_*=\lim_{n\to\infty}\frac{1}{2\sqrt{N}}\sum_{k=1}^{n}\left[\frac{\partial^2 Q(\beta^*,\gamma^*)}{\partial\beta_k\partial\beta^T}V^{-1}UV^{-1}e_k\right]. \tag{23}$$

Note that $s_{ij}(\beta,\gamma)$, $i<j$, are independent vectors. By Lyapunov's central limit theorem, we have

Proposition 2 *Let* $\lambda_{ij} = \mathrm{Var}(a_{ij})$ *and* $\widetilde{z}_{ij} = z_{ij} - V_{Q\beta}V^{-1}T_{ij}$. *For any nonzero vector* $c = (c_1, \dots c_p)^T$, *if*

$$\frac{\sum_{i<j}(c^T\widetilde{z}_{ij})^3\lambda_{ij}^3}{[\sum_{i<j}(c^T\widetilde{z}_{ij})^2\lambda_{ij}]^{3/2}} = o(1), \tag{24}$$

then $(c^T\Sigma c)^{-1/2}\sum_{i<j}\widetilde{s}_{\gamma_{ij}}(\beta^*, \gamma^*)$ *converges in distribution to the standard normal distribution, where* $\Sigma = \mathrm{Cov}(Q - V_{Q\beta}V^{-1}H)$.

In view of Proposition 2 and Theorem 3, we immediately have

Corollary 3 *Assume the conditions in Theorem 3 and (24) hold. Then, for any nonzero vector* $c = (c_1, \dots c_p)^T$, *we have*

$$\sqrt{N}c^T(\hat{\gamma} - \gamma)\xrightarrow{d}N\left(\bar{H}^{-1}B_*, c^T\bar{H}^T\Sigma\bar{H}c\right) \tag{25}$$

When the edge distribution (2) belongs to exponential family, we have $V = U$. Consequently, $\partial F(\beta^*, \gamma^*)/\partial\beta = \mathrm{Var}(d)$, B_* and Σ can be simplified as follows:

$$B_* = \frac{1}{\sqrt{N}}\sum_{k=1}^{n}\frac{\sum_{j\neq k} z_{kj}\mu''_{kj}(\pi^*_{ij})}{\sum_{j\neq k}\mu'_{kj}(\pi^*_{ij})}, \tag{26}$$

and

$$\begin{aligned}\Sigma = &\sum_{i<j} z_{ij}z_{ij}^T\mu'_{ij}\\ &-\sum_{i=1}^{n}\frac{\left(\sum_{j\neq i} z_{ij}\mu'_{ij}\right)\left(\sum_{j\neq i} z_{ij}^T\mu'_{ij}\right)}{v_{ii}}.\end{aligned}$$

Note that asymptotic normality of $\hat{\gamma}$ contains a bias term and needs to be corrected when constructing confidence interval and hypothesis testing. Here, we employ the analytical bias correction formula in Dzemski (2019): $\hat{\gamma}_{bc} = \hat{\gamma} - N^{-1/2}H^{-1}(\hat{\beta}, \hat{\gamma})\hat{B}$, where $\hat{B}$ is a plug-in estimator for B_* using $\hat{\beta}$ and $\hat{\gamma}$. Other bias-corrections include (Graham, 2017; Fernández-Vál and Weidner, 2016).

5 Applications

In this section, we illustrate the theoretical result by three applications: the logistic distribution, Poisson distribution and probit distribution for $f(\cdot)$. Moreover, any other distributions such as the geometric distribution that lead to the well-defined moment estimator could also be used, besides the logistic distribution and the Poisson distribution.

5.1 The logistic model

We consider the generalized β-model in Graham (2017) with the logistic distribution:

$$\mathbb{P}(a_{ij} = 1) = \frac{e^{\beta_i+\beta_j+z_{ij}^T\gamma}}{1 + e^{\beta_i+\beta_j+z_{ij}^T\gamma}}.$$

Graham (2017) derived the consistency and asymptotic normality of the restricted MLE. The aim of this application is to show that these properties of the unrestricted MLE continue to hold. In this model, the MLE is the same as the moment estimator.

The numbers involved with the conditions in theorems are as follows. Because a_{ij}'s are Bernoulli random variables, they are sub-exponential with $h_n = 1$. The numbers b_{n0}, b_{n1}, b_{n2} and b_{n3} are as defined in (8) and the paragraph right above it. The condition (21) in Theorem 1 becomes that

$$\kappa_n^2\omega_n^3 = o\left(\sqrt{\frac{n}{\log n}}\right), \tag{27}$$

where $\omega_n = e^{2\|\beta^*\|_\infty+\|\gamma^*\|_\infty}$.

By Theorem 1, we have the following corollary.

Corollary 4 *If* (27) *holds, then*

$$\begin{aligned}\|\widehat{\gamma} - \gamma^*\|_\infty &= O_p\left(\frac{\kappa_n\omega_n^3\log n}{n}\right),\\ \|\widehat{\beta} - \beta^*\|_\infty &= O_p\left(\omega_n\sqrt{\frac{\log n}{n}}\right).\end{aligned}$$

We discuss the condition and convergence rates related to the graph density. The expectation of the graph density is

$$\rho_n := \frac{1}{N}\sum_{1\le i<j\le n}\mathbb{E}a_{ij} = \frac{1}{N}\sum_{1\le i<j\le n}\frac{e^{\beta_i+\beta_j+z_{ij}^T\gamma}}{1 + e^{\beta_i+\beta_j+z_{ij}^T\gamma}},$$

where $N = n(n-1)/2$. To see what is κ_n, let us consider the case of that z_{ij} is one dimension. By using S in (14) to approximate $V^{-1}(\beta, \gamma)$, one can get

$$\begin{aligned}H(\beta,\gamma) = &\sum_{1\le i<j\le n} z_{ij}^2\mu'(\pi_{ij})\\ &-\sum_{i=1}^{n}\frac{1}{v_{ii}}\left(\sum_{j=1,j\neq i}^{n} z_{ij}\mu'(\pi_{ij})\right)^2.\end{aligned}$$

In this case, κ_n is approximately the inverse of $n^{-2}H(\beta^*,\gamma^*)$, which depends on the covariates, the configuration of parameters, and the derivative of the mean function

$\mu(\cdot)$. Since the relationship between (κ_n, b_{n0}) and ρ_n depends on the configuration of the parameters β and γ, where recall the definition of b_{n0} in (8), it is not possible to express κ_n and b_{n0} as a function of ρ_n for a general β and γ. Therefore, we consider one special case that $\beta_1 = \cdots = \beta_n \le c$ for illustration, where c is a constant, and assume that z_{ij} is independently drawn from the standard normality. In this case, by large sample theory, we have

$$\frac{1}{N}\sum_{1\le i<j\le n} z_{ij}^2 \mu'(\pi_{ij}) \xrightarrow{p.} \frac{e^{2\beta_1}}{(1+e^{2\beta_1})^2}, \quad \frac{1}{n}\sum_{j=1, j\neq i}^{n} z_{ij}\mu'(\pi_{ij}) \xrightarrow{p.} 0,$$

such that $\kappa_n \asymp 1/\rho_n$, where $a_n \asymp b_n$ means $c_1 a_n \le b_n \le c_2 a_n$ with two constants c_1 and c_2 for sufficiently large n. Further, $b_{n0} = O(\rho_n)$.

Then the condition in Corollary 1 becomes

$$\frac{\rho_n}{(\log n/n)^{1/8}} \to \infty,$$

and, the convergence rates are

$$\|\hat{\gamma} - \gamma^*\|_\infty = O_p\left(\frac{\log n}{n\rho_n^4}\right), \quad \|\hat{\beta} - \beta^*\|_\infty = O_p\left(\frac{1}{\rho_n}\sqrt{\frac{\log n}{n}}\right).$$

Here, estimation consistency requires a strong assumption $\rho_n \gg (n/\log n)^{1/8}$. It would be of interest to relax it.

Since a_{ij}'s $(j < i)$ are independent, it is easy to show the central limit theorem for d_i and $N^{-1/2}\sum_{j<i} \tilde{s}_{ij}(\beta, \gamma)$ as given in Su et al. (2018) and Graham (2017) respectively. So by Theorems 2 and 3, the central limit theorem holds for $\hat{\beta}$ and $\hat{\gamma}$. See Su et al. (2018) and Graham (2017) for details.

5.2 The Poisson model

We now consider the Poisson model in Example 2.

Recall that the expectation of a_{ij} is $\lambda_{ij} = e^{z_{ij}^T\gamma+\beta_i+\beta_j}$. In this case, $\mu(x) = e^x$. The likelihood function is

$$\mathbb{P}(A) \propto \exp\left(\sum_{i=1}^{n} \beta_i d_i + \sum_{1\le i<j\le n} a_{ij}(z_{ij}^T\gamma)\right).$$

It is a special case of the general exponential random graph model, where $(d^T, \sum_{i<j} a_{ij} z_{ij}^T)^T$ is the sufficient statistic for the parameter vector $(\beta^T, \gamma^T)^T$. Therefore, the maximum likelihood equations are identical to the moment equations defined in (5) and (6).

Define

$$q_n := \sup_{\beta\in B_\infty(\beta^*,\epsilon_{n1}),\gamma\in B_\infty(\gamma^*,\epsilon_{n2})} \max_{i,j} |\beta_i + \beta_j + z_{ij}^T\gamma|.$$

So b_{ni}'s ($i = 0, \ldots, 3$) in inequalities (7b), (7c) and (7d) are

$$b_{n0} = e^{-q_n}, \quad b_{n1} = e^{q_n}, \quad b_{n2} = e^{q_n}, \quad b_{n3} = e^{q_n}.$$

Clearly, Poisson(λ) is sub-exponential with parameter $c\lambda$, where c is a constant; see Example 4.6 in Zhang and Chen (2021). Thus, h_n in Assumption 1 is ce^{2q_n}. By Theorem 1, we have the following corollary.

Corollary 5 *If* $\kappa_n e^{7q_n} = o((n/\log n)^{1/2})$, *then then*

$$\|\hat{\gamma} - \gamma^*\|_\infty = O_p\left(\frac{\kappa_n e^{8q_n}\log n}{n}\right) = o_p(1),$$
$$\|\hat{\beta} - \beta^*\|_\infty = O_p\left(\frac{e^{2q_n}(\log n)^{1/2}}{n^{1/2}}\right) = o_p(1).$$

We discuss the condition and convergence rates related to the average weight. The expectation of the average weight is

$$\lambda_n := \frac{1}{N}\sum_i \mathbb{E}d_i = \frac{1}{n}\sum_i\sum_{j\neq i} e^{\beta_i+\beta_j+z_{ij}^T\gamma}.$$

As in the first application, b_{n0}, b_{n1} and b_{n2} can not be represented as functions on λ_n for general parameters β and γ. To get some intuitive understandings, let us consider a simple special case where $\beta_1 = \cdots = \beta_n < c$ with a constant c, γ is a constant and z_{ij} independently follows from a symmetric continuous distribution with a bounded support and the unit variance. In this case,

$$\kappa_n \asymp \lambda_n^{-1}, \quad \lambda_n = e^{2\beta_1+O(1)}, \quad c_1\lambda_n \le b_{n0}, b_{n1}, b_{n2} \le c_2\lambda_n, \quad h_n = c_3\lambda_n^2,$$

where c_1, c_2 and c_3 are positive constants. Then, the condition in Theorem 1 becomes

$$\lambda_n = o\left(\left(\frac{n}{\log n}\right)^{1/4}\right),$$

and the convergence rates are

$$\|\hat{\gamma} - \gamma^*\|_\infty = O_p\left(\frac{\lambda_n^2\log n}{n}\right),$$
$$\|\hat{\beta} - \beta^*\|_\infty = O_p\left(\lambda_n\sqrt{\frac{\log n}{n}}\right).$$

Note that $d_i = \sum_{j\neq i} a_{ij}$ is a sum of $n-1$ independent Poisson random variables. Since $v_{ij} = \mathbb{E}a_{ij} = \lambda_{ij}$, we have

$$e^{-q_n} \le v_{ij} = e^{\beta_i+\beta_j+z_{ij}^T\gamma} \le e^{q_n}, \quad 1 \le i < j \le n.$$

Byusing the Stein-Chen identity (Stein, 1972; Chen, 1975) for the Poisson distribution, it is easy to verify that

$$\mathbb{E}(a_{ij}^3) = \lambda_{ij}^3 + 3\lambda_{ij}^2 + \lambda_{ij}. \tag{28}$$

It follows

$$\frac{\sum_{j\neq i}\mathbb{E}(a_{ij}^3)}{v_{ii}^{3/2}} \le \frac{(n-1)e^{q_n}}{(n-1)^{3/2}e^{-q_n}} = O(\frac{e^{4q_n}}{n^{1/2}}).$$

If $e^{4q_n} = o(n^{1/2})$, then the above expression goes to zero. For any nonzero vector $c = (c_1, \dots c_p)^T$, if

$$\frac{\sum_{j<i}(c^T\widetilde{z}_{ij})^3\lambda_{ij}^3}{\left[\sum_{j<i}(c^T\widetilde{z}_{ij})^2\lambda_{ij}\right]^{3/2}} = o(1), \tag{29}$$

This verifies the condition (24). Consequently, by Corollaries 2 and 3, we have the following result.

Corollary 6 *If* (29) *holds and* $\lambda_n^2\kappa_n^6 e^{28q_n} = o(n^{1/2}/(\log n)^{3/2})$, *then*:

(1) $N^{1/2}\overline{\Sigma}^{-1/2}(\widehat{\gamma} - \gamma^*)$ *converges in distribution to multivariate normal distribution with mean* $\overline{\Sigma}^{-1/2}\bar{H}^{-1}B_*$ *and covariance* I_p, *where* I_p *is the identity matrix, where* $\bar{\Sigma} = N^{-1}\bar{H}^{-1}\widetilde{\Sigma}\bar{H}^{-1}$;
(2) *for a fixed r, the vector* $(v_{11}^{1/2}(\widehat{\beta}_1 - \beta_1^*), \dots, v_{rr}^{1/2}(\widehat{\beta}_r - \beta_r^*)$ *converges in distribution to the r-dimensional standard normal distribution.*

5.3 The probit model

The two examples above are exponential family of distributions. Here, we pay attention to the probit distribution, which is not exponential. Let $\phi(x) = (2\pi)^{1/2}e^{-x^2/2}$ be the standard normal density function and $\Phi(x) = \int_{-\infty}^x \phi(x)dx$ be its the distribution function. The probit model assumes

$$\mathbb{P}(a_{ij} = 1) = \Phi\Big(\frac{1}{\sigma}(\beta_i + \beta_j + z_{ij}^\top\gamma)\Big),$$

where σ is the standard derivation. Since the parameters are scale invariable, we simply set $\sigma = 1$. Then,

$$\mu'(x) = \phi(x), \quad \mu''(x) = \frac{x}{\sqrt{2\pi}} e^{-x^2/2}.$$

Since $\phi(x) = (2\pi)^{1/2} e^{-x^2/2}$ is an decreasing function on |x|, we have when $|x| \le Q_n$,

$$\frac{1}{2\pi} e^{-Q_n^2/2} \le \phi(x) \le \frac{1}{2\pi}.$$

Let $h(x) = xe^{-x^2/2}$. Then $h'(x) = (1 - x^2)e^{-x^2/2}$. Therefore, when $x \in (0, 1)$, $h(x)$ is an increasing function on its argument x; when $x \in (1, \infty)$, $h(x)$ is an decreasing function on x. As a result, $h(x)$ attains its maximum value at $x = 1$ when $x > 0$. Since $h(x)$ is a symmetric function, we have $|h(x)| \le e^{-1/2} \approx 0.6$. So

$$b_{n0} \asymp \frac{1}{2\pi} e^{-(\max_{i,j} \pi_{ij}^*)^2/2}, \quad b_{n1} = \frac{1}{2\pi}, \quad b_{n2} = (2\pi e)^{-1/2}.$$

We only consider conditions for consistency here and those for central limit theorem are similar and omitted. By (17), it is not difficult to verify

$$\sigma_n^2 = O\big(n^4 \|V^{-1} - S\|_{\max}^2\big) = O\left(\frac{b_{n1}^4}{b_{n0}^6}\right).$$

The parameter h_n in Assumption 2 for a bounded random variable is a constant. In view of Theorem 1, we have the following corollary.

Corollary 7 *If*

$$\kappa_n e^{3(\max_{i,j} \pi_{ij}^*)^2} = o\left(\frac{n}{\log n}\right),$$

then the moment estimator $(\hat{\beta}, \hat{\gamma})$ exists with high probability, and we further have

$$\|\hat{\gamma} - \gamma^*\|_\infty = O_p\left(\kappa_n e^{3(\max_{i,j} \pi_{ij}^*)^2/2} \frac{\log n}{n}\right), \quad \|\hat{\beta} - \beta^*\|_\infty = O_p\left(e^{(\max_{i,j} \pi_{ij}^*)^2/2} \sqrt{\frac{\log n}{n}}\right).$$

6 Discussion

In this paper, we present a moment estimation for inferring the degree parameter β and homophily parameter γ in model (2). We establish consistency of the moment estimator $(\hat{\beta}, \hat{\gamma})$ under several conditions and also derive its asymptotic normality. The convergence rates of $\hat{\beta}$ and $\hat{\gamma}$ are nearly optimal when all parameters are bounded by a constant; but may not be optimal when the numbers b_{n0}, b_{n1}, b_{n2} and κ_n diverge. Theorems 2 and 3 require stronger assumptions than consistency, but this is a widely observed phenomenon in existing literature (Yan et al., 2016a, 2019; Zhang et al., 2021). Whether it is possible to establish consistency and asymptotic normality under even weaker conditions will be an interesting future work.

For cleanness, in this work, we assume that $\max_{i,j} \|z_{ij}\|_\infty < c$ is universally bounded. In fact, our theory can be extended to allow it to slowly diverge. It is another interesting future research to investigate how fast it can diverge while preserving consistency.

The independent edge assumption leads to convenient characterization of the orders of $\|d - \mathbb{E}d\|_\infty$ and $\| \sum_{i<j} z_{ij}(a_{ij} - \mathbb{E}a_{ij})\|_\infty$, based on which, we establish the central limit theorems of d and $\sum_{i<j} z_{ij} a_{ij}$. For sub-exponential a_{ij}, the orders of $\|d - \mathbb{E}d\|_\infty$ and $\| \sum_{i<j} z_{ij}(a_{ij} - \mathbb{E}a_{ij})\|_\infty$ are $O((n \log n)^{1/2})$ and $O(n \log n)$, respectively, up to a factor determined by the sub-exponential parameter h_n. Going forward, we can introduce slight dependency between edges. Under such setting, we can still use some Hoeffding-type inequalities for dependent random variables to establish tail bounds similar to those in this paper (Delyon, 2009), as long as edge dependency is sufficiently light. Remarkably, our method-of-moments estimation remains effective, since it only requires specification of the marginal distributions of a_{ij}'s, not the joint distribution of A. Certainly, quantitative study along this direction would require highly nontrivial future efforts.

Computation for covariate-assisted β-models is challenging in general. The GLM package we use, which was also employed by Chen et al. (2021), Stein and Leng (2020) and Stein and Leng (2021), do not scale well. Directly programming the Newton method seems more promising, but still might encounter difficulty when the network is $\Omega(10^5)$. Unfortunately, the reduction method invented by Zhang et al. (2021) only works for the classical and some generalized β-models without covariates, not for covariate-assisted β-models. Exploring efficient computational methods is an interesting open challenge for future research.

Acknowledgements We are grateful to Editor, Associate Editor and two anonymous referees for their insightful comments and suggestions. TY was partially supported by the National Natural Science Foundation of China (Nos. 11771171, 12171188) and the Fundamental Research Funds for the Central Universities.

Appendix

Preliminaries

In this section, we present three results that will be used in the proofs. The first is on the approximation error of using S to approximate the inverse of V belonging to the matrix class $\mathcal{L}_n(b_{n0}, b_{n1})$, where $V = (v_{ij})_{n\times n}$ and $S = \mathrm{diag}(1/v_{11}, \ldots, 1/v_{nn})$. Yan et al. (2015) obtained the upper bound of the approximation error, which has an order n^{-2}. Hillar et al. (2012) gave a tight bound of $\|V^{-1}\|_\infty$. These results are stated below as lemmas.

Lemma 1 (*Proposition 1 in Yan et al.,* 2015) *If* $V \in \mathcal{L}_n(b_{n0}, b_{n1})$, *then the following holds:*

$$\|V^{-1} - S\|_{\max} = O\left(\frac{b_{n1}^2}{n^2 b_{n0}^3}\right). \tag{30}$$

Lemma 2 (*Hillar et al.*, 2012) *For* $V \in \mathcal{L}_n(b_{n0}, b_{n1})$, *we have*

$$\frac{1}{2b_{n1}(n-1)} \leq \|V^{-1}\|_\infty \leq \frac{3n-4}{2b_{n0}(n-1)(n-2)}.$$

Let $F(x) : \mathbb{R}^n \to \mathbb{R}^n$ be a function vector on $x \in \mathbb{R}^n$. We say that a Jacobian matrix $F'(x)$ with $x \in \mathbb{R}^n$ is Lipschitz continuous on a convex set $D \subset \mathbb{R}^n$ if for any $x, y \in D$, there exists a constant $\lambda > 0$ such that for any vector $v \in \mathbb{R}^n$ the inequality

$$\|[F'(x)]v - [F'(y)]v\|_\infty \leq \lambda \|x - y\|_\infty \|v\|_\infty$$

holds. We will use the Newton iterative sequence to establish the existence and consistency of the moment estimator. Gragg and Tapia (1974) gave the optimal error bound for the Newton method under the Kantovorich conditions (Kantorovich, 1948).

Lemma 3 (*Gragg and Tapia,* 1974) *Let D be an open convex set of* $\mathbb{R}^n$ *and* $F : D \to \mathbb{R}^n$ *a differential function with a Jacobian* $F'(x)$ *that is Lipschitz continuous on D with Lipschitz coefficient* λ. *Assume that* $x_0 \in D$ *is such that* $[F'(x_0)]^{-1}$ *exists,*

$$\|[F'(x_0)]^{-1}\|_\infty \leq \aleph, \quad \|\left[F'(x_0)\right]^{-1} F(x_0)\|_\infty \leq \delta, \quad \rho = 2\aleph\lambda\delta \leq 1,$$
$$B_\infty(x_0, t^*) \subset D, \quad t^* = \frac{2}{\rho}\left(1 - \sqrt{1-\rho}\right)\delta = \frac{2\delta}{1 + \sqrt{1-\rho}} \leq 2\delta.$$

Then: (1) The Newton iterations $x_{k+1} = x_k - [F'(x_k)]^{-1}F(x_k)$ *exist and* $x_k \in B_\infty(x_0, t^*) \subset D$ *for* $k \geq 0$. (2) $x^* = \lim x_k$ *exists,* $x^* \in \overline{B_\infty(x_0, t^*)} \subset D$ *and* $F(x^*) = 0$.

Error bound between $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\beta}^*$

The lemma below shows that $F_\gamma(\beta)$ is Lipschitz continuous. The proofs of all the lemmas in this section are given in the supplementary material.

Lemma 4 *Let* $D = B_\infty(\beta^*, \epsilon_{n1})(\subset \mathbb{R}^n)$ *be an open convex set containing the true point* β^*. *For* $\gamma \in B_\infty(\gamma^*, \epsilon_{n2})$, *if inequality* (7d) *holds, then the Jacobian matrix* $F'_\gamma(x)$ *of* $F_\gamma(x)$ *on x is Lipschitz continuous on D with the Lipschitz coefficient* $4b_{n2}(n-1)$.

Since a_{ij}, $1 \le i < j \le n$, are independent and sub-exponential with parameters $h_{ij}(\le h_n)$, by the concentration inequality for sub-exponential random variables (e.g., Corollary 5.17 in Vershynin, 2012), we have the following lemma.

Lemma 5 *With probability at least* $1 - O(n^{-1})$, *we have*

$$\|F(\beta^*, \gamma^*)\|_\infty = O(h_n\sqrt{n\log n}), \quad \|Q(\beta^*, \gamma^*)\|_\infty = O(h_n n \log n). \tag{31}$$

In view of Lemmas 4 and 5, we obtain the upper bound of the error between $\hat{\beta}_\gamma$ and β^* by using the Newton method.

Lemma 6 *Let* ϵ_{n1} *be a positive number and* $\epsilon_{n2} = o(b_{n0}^{-1}(\log n)^{1/2}n^{-1/2})$. *Assume that* (7b), (7c) *and* (7d) *hold. If*

$$\frac{b_{n2}h_n}{b_{n0}^2} = o\left(\sqrt{\frac{n}{\log n}}\right), \tag{32}$$

then with probability at least $1 - O(n^{-1})$, *for* $\gamma \in B_\infty(\gamma^*, \epsilon_{n2})$, $\hat{\beta}_\gamma$ *exists and satisfies*

$$\|\hat{\beta}_\gamma - \beta^*\|_\infty = O_p\left(\frac{h_n}{b_{n0}}\sqrt{\frac{\log n}{n}}\right) = o_p(1).$$

Proofs for Theorem 1

To show Theorem 1, we need three lemmas below.

Lemma 7 *Let* $D = B_\infty(\gamma^*, \epsilon_{n2}) (\subset \mathbb{R}^p)$ *be an open convex set containing the true point* γ^*. *Assume that* (7b), (7c), (7d) *and* (32) *hold. If* $\|F(\beta^*, \gamma^*)\|_\infty = O(h_n(n\log n)^{1/2})$, *then* $Q_c(\gamma)$ *is Lipschitz continuous onD with the Lipschitz coefficient* $n^2 b_{n2} b_{n1}^3 b_{n0}^{-3}$.

Lemma 8 *Write* $\hat{\beta}^*$ *as* $\hat{\beta}_{\gamma^*}$ *and* $V = \partial F(\beta^*, \gamma^*)/\partial \beta^T$. $\hat{\beta}^*$ *has the following expansion*:

$$\hat{\beta}^* - \beta^* = V^{-1}F(\beta^*, \gamma^*) + V^{-1}R, \tag{33}$$

where $R = (R_1, \ldots, R_n)^T$ *is the remainder term and*

$$\left\|V^{-1}R\right\|_\infty = O_p\left(\frac{b_{n2}h_n^2\log n}{nb_{n0}^3}\right).$$

Lemma 9 *Let* $\Omega = \mathrm{Cov}(F(\beta^*, \gamma^*))$. *Let* $\sigma_n^2 = n^2\|(V^{-1} - S)\Omega(V^{-1} - S)\|_{\max}$. *For any* $\beta \in B_\infty(\beta^*, \epsilon_{n1})$ *and* $\gamma \in B_\infty(\gamma^*, \epsilon_{n2})$, *we have*

$$\left\|\frac{\partial Q(\beta, \gamma)}{\partial \beta^T}(\hat{\beta}^* - \beta^*)\right\|_\infty = O_p\left(nb_{n1}\log n\left(\frac{b_{n2}h_n^2}{b_{n0}^3} + \sigma_n\right)\right).$$

Further, when $\Omega = -\partial F(\beta^*, \gamma^*)/\partial \beta^T$, *we have*

$$\|\frac{\partial Q(\beta, \gamma)}{\partial \beta^T}(\widehat{\beta}^* - \beta^*)\|_\infty = O_p\left(\frac{h_n^2 b_{n1} b_{n2} n \log n}{b_{n0}^3}\right).$$

Now we are ready to prove Theorem 1.

Proof of Theorem 1 We construct the Newton iterative sequence to show the consistency. In view of Lemma 3, it is sufficient to demonstrate the Newton-Kantovorich conditions. We set γ^* as the initial point $\gamma^{(0)}$ and $\gamma^{(k+1)} = \gamma^{(k)} - [Q_c'(\gamma^{(k)})]^{-1} Q_c(\gamma^{(k)})$.

By Lemma 6, with probability at least $1 - O(n^{-1})$, we have we have

$$\|\widehat{\beta}_\gamma - \beta^*\|_\infty = O_p\left(\frac{h_n}{b_{n0}}\sqrt{\frac{\log n}{n}}\right).$$

This shows that $\widehat{\beta}_{\gamma^{(0)}}$ exists such that $Q_c(\gamma^{(0)})$ and $Q_c'(\gamma^{(0)})$ are well defined.

Recall the definition of $Q_c(\gamma)$ and $Q(\beta, \gamma)$ in (10) and (11). By Lemmas 5 and 9, we have

$$\begin{aligned}\|Q_c(\gamma^*)\|_\infty \leq & \|Q(\beta^*, \gamma^*)\|_\infty + \|Q(\widehat{\beta}_{\gamma^*}, \gamma^*) - Q(\beta^*, \gamma^*)\|_\infty \\ = & O_p\left(nb_{n1} \log n \left(\frac{h_n^2 b_{n2}}{b_{n0}^3} + \sigma_n\right)\right).\end{aligned}$$

By Lemma 7, $\lambda = n^2 b_{n1}^3 b_{n2} b_{n0}^{-3}$. Note that $\aleph = \|[Q_c'(\gamma^*)]^{-1}\|_\infty = O(\kappa_n n^{-2})$. Thus,

$$\delta = \|[Q_c'(\gamma^*)]^{-1} Q_c(\gamma^*)\|_\infty = O_p\left(\frac{\kappa_n b_{n1} \log n}{n}\left(\frac{h_n^2 b_{n2}}{b_{n0}^3} + \sigma_n\right)\right).$$

As a result, if Eq. (21) holds, then

$$\rho = 2\aleph\lambda\delta = O_p\left(\frac{\kappa_n^2 b_{n1}^4 b_{n2} \log n}{n b_{n0}^3}\left(\frac{h_n^2 b_{n2}}{b_{n0}^3} + \sigma_n\right)\right) = o_p(1).$$

By Theorem 3, the limiting point of the sequence $\{\gamma^{(k)}\}_{k=1}^\infty$ exists, denoted by $\widehat{\gamma}$, and satisfies

$$\|\widehat{\gamma} - \gamma^*\|_\infty = O_p(\delta).$$

By Lemma 6, $\widehat{\beta}_{\widehat{\gamma}}$ exists, denoted by $\widehat{\beta}$, and $(\widehat{\beta}, \widehat{\gamma})$ is the moment estimator. It completes the proof. □

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-022-00848-0.

References

Amemiya, T. (1985). *Advanced econometrics*. Cambridge: Harvard University Press.

Babai, L., Erdos, P., & Selkow, S. M. (1980). Random graph isomorphism. *SIAM Journal on computing, 9*(3), 628–635.

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

Binkiewicz, N., Vogelstein, J. T., & Rohe, K. (2017). Covariate-assisted spectral clustering. *Biometrika, 104*(2), 361–377.

Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social Networks, 21*(4), 375–395.

Bubeck, S. (2015). Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning, 8*(3–4), 231–357.

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

Chen, L. (1975). Poisson approximation for dependent trials. *The Annals of Probability, 3*(3), 534–545.

Chen, M., Kato, K., & Leng, C. (2021). Analysis of networks via the sparse β-model. *Journal of the Royal Statistical Society*: *Series B* (*Statistical Methodology*), *83,* 887–910.

Cho, E., Myers, S. A., & Leskovec, J. (2011). Friendship and mobility: User movement in location-based social networks. *Proceedings of the international conference on knowledge discovery and data mining*, 1082–1090.

Delyon, B. (2009). Exponential inequalities for sums of weakly dependent variables. *Electronic Journal of Probability, 14,* 752–779.

Dzemski, A. (2019). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Review of Economics and Statistics, 101,* 763–776.

Fan, Y., Jiang, B, Yan, T., & Zhang, Y. (2022). Asymptotic theory in bipartite graph models with a growing number of parameters. To appear in Canadian Journal of Statistics.

Fernández-Vál, I., & Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics, 192*(1), 291–312.

Gragg, W. B., & Tapia, R. A. (1974). Optimal error bounds for the newtonckantorovich theorem. *SIAM Journal on Numerical Analysis, 11*(1), 10–13.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica, 85*(4), 1033–1063.

Hillar, C. J., Lin, S., & Wibisono, A. (2012). Inverses of symmetric, diagonally dominant positive matrices and applications. arXiv:1203.6812.

Kantorovich, L. V. (1948). Functional analysis and applied mathematics. Uspekhi Mat Nauk z, 89–185.

Lounici, K. (2008). Sup-norm convergence rate and sign concentration property of lasso and dantzig estimators. *Electronic Journal of Statistics, 2,* 90–102.

Maugis, P. (2020). Central limit theorems for local network statistics. arXiv:2006.15738.

Rinaldo, A., Petrović, S., & Fienberg, S. E. (2013). Maximum lilkelihood estimation in the β-model. *The Annals of Statistics, 41*(3), 1085–1110.

Simons, G., Yao, Y. C., et al. (1999). Asymptotics when the number of parameters tends to infinity in the Bradley–Terry model for paired comparisons. *The Annals of Statistics, 27*(3), 1041–1060.

Stein, C. M. (1972). A bound for the error in normal approximation to the distribution of a sum of dependent random variables. *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability*, Vol. 3, 583–602.

Stein, S., & Leng, C. (2020). A sparse β-model with covariates for networks. arXiv:2010.13604.

Stein, S., & Leng, C. (2021). A sparse random graph model for sparse directed networks. arXiv:2108.09504.

Su, L., Qian, X., & Yan, T. (2018). A note on a network model with degree heterogeneity and homophily. *Statistics and Probability Letters, 138,* 27–30.

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In Y. Eldar & G. Kutyniok (Eds.), *Compress Sensing: Theory and Applications* (pp. 210–268). Cambridge: Cambridge University Press.

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,* 519–524.

Yan, T., Zhao, Y., & Qin, H. (2015). Asymptotic normality in the maximum entropy models on graphs with an increasing number of parameters. *Journal of Multivariate Analysis, 133,* 61–76.
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., Jiang, B., Fienberg, S. E., & Leng, C. (2019). Statistical inference in a directed network model with covariates. *Journal of the American Statistical Association, 114,* 857–868.
Yang, J., McAuley, J., & Leskovec, J. (2013). Community detection in networks with node attributes. *2013 IEEE 13th international conference on data mining*, 1151–1156. IEEE.
Zhang, H., & Chen, S. X. (2021). Concentration inequalities for statistical inference. *Communications in Mathematical Research, 37*(1), 1–85.
Zhang, Y., & Xia, D. (2022). Edgeworth expansions for network moments. *The Annals of Statistics, 50*(2), 726–753.
Zhang, Y., Levina, E., & Zhu, J. (2016). Community detection in networks with node features. *Electronic Journal of Statistics, 10*(2), 3153–3178.
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:2110.11856.