JUSTC

Article

http://justc.ustc.edu.cn

Received: November 02, 2024; Accepted: December 31, 2024

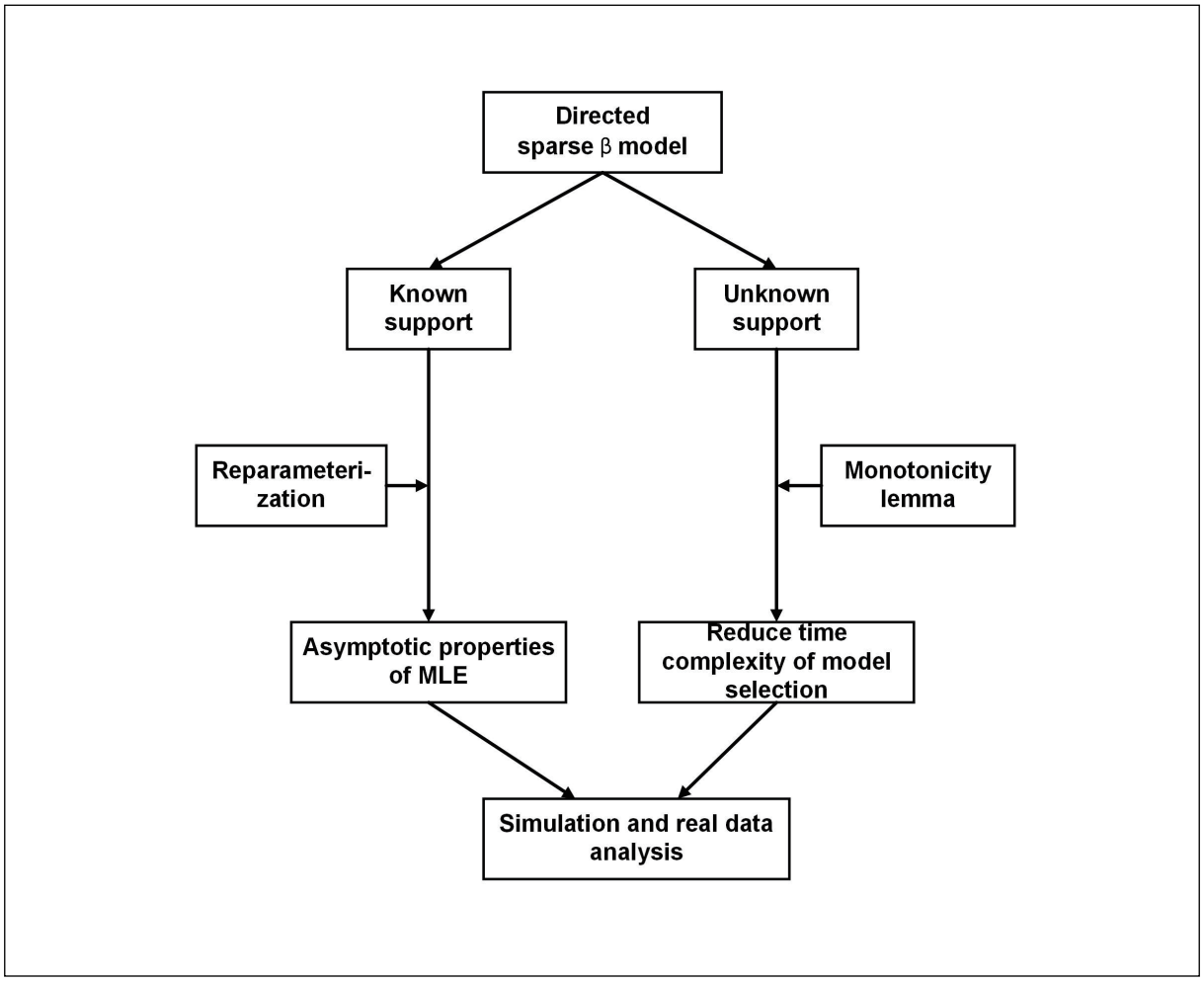
Sparse β model for the directed networks

Zi'ang Xu, and Qunqiang Feng ✉

Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Qunqiang Feng, E-mail: fengqq@ustc.edu.cn

Graphical abstract



Sparse β model for the directed networks.

Public summary

- Construct a sparse β model for directed networks (DSβM).
- Derive consistency and asymptotic normality by reparameterizing the global and local density parameters when the support is known.
- Propose the monotonicity lemma with the l_0 penalized likelihood approach to estimate the parameters.

Citation: Xu Z A, Feng Q Q. Sparse β model for the directed networks. *JUSTC*, **2025**, 55(0): . CSTR: 32290.14.JUSTC-2024-0139. DOI: 10.52396/JUSTC-2024-0139

JUSTC

Article

http://justc.ustc.edu.cn

Received: November 02, 2024; Accepted: December 31, 2024

Sparse β model for the directed networks

Zi'ang Xu, and Qunqiang Feng ✉

Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, China

✉Correspondence: Qunqiang Feng, E-mail: fengqq@ustc.edu.cn

Cite This: *JUSTC*, **2025**, 55(X): (13pp)

Read Online

Supporting Information

Abstract: Many complex systems involve plenty of units that interact with each other. The study of networks has attracted increasing attention in a wide variety of areas. This paper concerns a sparse β model for directed networks (DSβM) by allowing self-loops, which can capture degree heterogeneity. Specifically, we have derived consistency and asymptotic normality by reparameterizing the global and local density parameters when the support is known. For the other case when the support is unknown, we propose the monotonicity lemma with the l_0 penalized likelihood approach to estimate the parameters. The simulation results help prove our theoretical properties and the real data analyses confirm the usefulness of our model.

Keywords: statistical network model; social network analysis; sparsity; degree heterogeneity; l_0 penalized likelihood

CLC number: O212.1 **Document code:** A

2020 Mathematics Subject Classification: 05C80; 62F12; 62F30

1 Introduction

Many complex networks are specified by a set of nodes and a set of edges that connect certain pairs of nodes[1]. During the past two decades, statistical network analysis has attracted increasing attention in a wide variety of areas, such as science[2], economics[3], sociology[4], and biology[5].

In a real-life society, observed networks are always sparse, which means there are much fewer edges than the maximum possible number of edges. Moreover, we can usually find degree heterogeneity in real networks. That is, there are few high-degree core nodes with many edges and many low-degree nodes with few links (see, e.g., Ref. [6]). Many statistical network models have been developed to directly account for degree heterogeneity. Two famous examples are the stochastic block model (SBM) and the β model. The SBM puts nodes into communities with similar connection patterns to capture degree heterogeneity[7–9], and the β model assigns each node with parameters to show degree heterogeneity[10]. The β model is a generalization of the Erdös-Rényi random graph model, where the probability that a pair of nodes are connected depends on the corresponding nodal parameters. It is also a special case of the p_1 model, which forms an exponential family of distributions to analyze directed networks[1].

Following the simple structure of the β model, many studies have been conducted on it. Until recently, the maximum likelihood estimator (MLE) of the β model has been demonstrated to be consistent and asymptotically normal for undirected dense networks[10, 11]. For undirected sparse networks, Ref. [12] proposed a new network model called the 'sparse β-model' (SβM), which can capture node degree heterogeneity and simultaneously allow parameter estimators with desirable statistical properties. Ref. [13] proposed the l_2 regularized maximum likelihood for the β model in large and sparse networks. Considering that likelihood ratio tests and the Wilks theorems have been pivotal in statistics but have rarely been explored in network models with an increasing dimension, Ref. [14] concerned likelihood ratio tests in the β model for undirected networks. When this problem comes to the directed networks, Ref. [15] found similar results on the MLE of the parameters in the p_1 model. Ref. [16] added covariates to the β model to find more interesting results, which can be applied for relatively dense networks. Ref. [17] proposed a new parameter-sparse random graph model for density-sparse directed networks, with parameters to explicitly account for all these features.

Despite the contributions of the SβM from dense to sparse network analysis, Ref. [12] showed that there is lack of theoretical results to apply their method from undirected to directed networks. To fill this gap, our main concern in this paper is to propose an improved generative model for directed sparse networks, which we call the directed sparse β model (DSβM).

The rest of this paper is structured as follows. In Section 2, we introduce the DSβM and give some notations. In Section 3, we derive the consistency and asymptotic normality of the parameter estimator in terms of excess risk and l_0-norm under known and unknown support. In Sections 4 and 5, we present some simulation results with this model and apply this model to real data, and concluding remarks are presented in Section 6. Full proofs for all of our results are presented in the supplementary materials.

2 Directed sparse β models

To describe our proposed network model, we first introduce

CSTR: 32290.14.JUSTC-2024-0139
DOI: 10.52396/JUSTC-2024-0139
JUSTC, **2025**, 55(X):

the following notation. A network with size n is represented as a directed graph $G_n=(V,E)$, where the node set $V=\{1,2,\cdots,n\}$, and the edge set E is a subset of all the two-element pair sets of V, i.e., $E\subset V\times V$. If an ordered node pair $(i,j)\in E$, we say that there is an edge from node i to node j in G_n. Alternatively, to represent G_n one can use a binary adjacency matrix $\boldsymbol{A}\in\{0,1\}^{n\times n}$ with the entry A_{ij} indicating the presence or absence of an edge from node i to node j, assigned as

$$A_{ij}=\begin{cases}1, & (i,j)\in E;\\ 0, & (i,j)\notin E.\end{cases}$$

In this paper, we may assume that self-loops in G_n are allowed, which means that the diagonal elements of $\boldsymbol{A}$ can be equal to 1. In addition, the out-degree d_{i+} of a node i and the in-degree d_{+j} of a node j are specified as $d_{i+}=\sum_{k=1}^{n}A_{ik}$ and $d_{+j}=\sum_{k=1}^{n}A_{kj}$, respectively. As basic knowledge in graph theory, one can obtain that the total number of edges in G_n is

$$d_+=\sum_{i=1}^{n}d_{i+}=\sum_{j=1}^{n}d_{+j}=\sum_{1\le i,j\le n}A_{ij}.$$

Before the formal description of our DSβM, we provide a brief review of the classical directed β model. This model, also known as the p_0 model, is a simple version of the well-known p_1 model without reciprocity (see, e.g., Ref. [18]). For any node pair (i,j) with $1\le i\ne j\le n$, the directed β model assumes that A_{ij}'s are generated as independent Bernoulli random variables with

$$\mathbb{P}(A_{ij}=1)=1-\mathbb{P}(A_{ij}=0)=\frac{\mathrm{e}^{\alpha_i+\beta_j}}{1+\mathrm{e}^{\alpha_i+\beta_j}},$$

where $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_n)^\top$ is called the *productivity parameter* vector, and $\boldsymbol{\beta}=(\beta_1,\cdots,\beta_n)^\top$ is called the *attractiveness parameter* vector. As explained in Ref. [1], α_i quantifies the effect of an outgoing edge from node i and β_j quantifies the effect of an incoming edge connecting to node j. Constraints on $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ are required for parameter identification, and two popular options are as follows: (i) fix a single parameter, say β_n, to be 0; (ii) set the sum $\sum_{i=1}^{n}\alpha_i$ or $\sum_{j=1}^{n}\beta_j$ to be 0. In any case all the parameters α_i and β_j are usually assumed to be around 0 and of order at most $\log n$[15]. As a result, the β model is more suitable to fit the relatively dense networks; see also Refs. [10, 11, 12]. Ref. [19] studied sharp thresholds for detecting the presence of differentially attractive nodes in the β model on sparse graphs from a hypothesis testing perspective.

In order to incorporate the sparsity in the DSβM, we first introduce an additional parameter $\mu\in\mathbb{R}$, which characterizes the sparsity of the entire network. More precisely, A_{ij}'s are independent Bernoulli random variables with success rates

$$p_{ij}=\mathbb{P}(A_{ij}=1)=\frac{\mathrm{e}^{\mu+\alpha_i+\beta_j}}{1+\mathrm{e}^{\mu+\alpha_i+\beta_j}},\quad 1\le i,j\le n,\tag{1}$$

where the parameters μ, α_i and β_j may depend on the network size n. At first glance the parameter μ seems meaningless in the sense that it can be absorbed into α_i and β_j as $\mu/2+\alpha_i$ and $\mu/2+\beta_j$, respectively. However, we can consider μ as a baseline term, allowing us to fit the real networks at different levels of sparsity, based on the simple fact that letting $\mu\to-\infty$ leads to each of $p_{ij}\to 0$ as $n\to\infty$. Therefore, we can assume that all α_i and β_j are non-negative in (1). This indicates that we require $\min_{1\le i\le n}\alpha_i=\min_{1\le j\le n}\beta_j=0$ to ensure the parameter identification in our model. Then α_i and β_j can be interpreted as the local productivity parameter of node i and the local attractiveness parameter of node j, respectively. Specially, if $\alpha_i=0$, the node i has no individual effect in the presence of an edge from it to other nodes, which may lead to a lower out-degree of i. Similarly, the node j is preferred to have a lower in-degree if $\beta_j=0$. In other words, α_i and β_i are both heterogeneity parameters characterizing the possibility of connections between the given node i and any other node in the network.

Following the idea in Refs. [20, 21], we then employ $\log n$ shifts to reparameterize μ, α_i and β_j as follows:

$$\mu=-\gamma\log n+\mu^*,\quad \alpha_i=\alpha\log n+\alpha_i^*\quad\text{and}\quad \beta_j=\beta\log n+\beta_j^*\tag{2}$$

for some $\gamma\in[0,2)$ and $\alpha,\beta\in[0,1)$ such that $0\le\gamma-\alpha-\beta<1$. Denote $\boldsymbol{\alpha}^*=(\alpha_1^*,\ldots,\alpha_n^*)^\top$ and $\boldsymbol{\beta}^*=(\beta_1^*,\cdots,\beta_n^*)^\top$. It is announced that this method is important in the proof of consistency and asymptotic normality of MLE. A similar reparameterization for an undirected sparse β model can also be found in Ref. [12]. It is worth noting that if the self-loops are ignored, our model may be considered as a special case of that proposed in Ref. [17], where the nodal covariates are added in networks, and a penalized likelihood with l_1-regularization is applied for parameter estimation. However, the consistency and asymptotic normality are established only for the global sparsity parameter μ and the covariate parameter $\boldsymbol{\gamma}$ in Ref. [17], while such asymptotic properties of the local parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are still unknown. Here we shall show these asymptotic properties for all the parameters $\mu,\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in our model (see Section 3.1 below).

As mentioned above, another slight difference from the β model is the allowance of self-loops in DSβM. As seen in the later, this is mainly due to technical reasons. On the other hand, since the sparsity of DSβM, the proportion of nodes with self-loops is always small and thus has little influence in the entire network. Indeed, we can also define the DSβM without self-loops similarly by replacing all diagonal entries in $\boldsymbol{A}$ with 0. Denote by $\boldsymbol{D}$ the resulting matrix. That is,

$$\boldsymbol{A}=\boldsymbol{D}+\boldsymbol{R},$$

where $\boldsymbol{R}$ is a diagonal matrix with the entry $R_{ii}=A_{ii}$ for $1\le i\le n$. We will derive more details about the relationships among these three matrices in the next section.

Finally, we define the sparsity of the parameters. For a vector $\boldsymbol{v}=(v_1,v_2,\cdots,v_n)\in\mathbb{R}^n$, let us denote by $S_{\boldsymbol{v}}=\{i:v_i\ne 0\}$ its support and $\|\boldsymbol{v}\|_0=|S_{\boldsymbol{v}}|$ the cardinality of $S_{\boldsymbol{v}}$. For convenience, we use $\boldsymbol{v}_S$ to denote the subvector of $\boldsymbol{v}$ with indices in $S_{\boldsymbol{v}}$. Since the DSβM shows much concern for the word 'sparse', we are interested in the case where $\|\boldsymbol{\alpha}\|_0\ll n$ and $\|\boldsymbol{\beta}\|_0\ll n$.

3 Parameter estimation in DSβM

In this section, we investigate the parameter estimation in the DSβM. In what follows, we denote by $(\mu_0,\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)^\top$ the true

CSTR: 32290.14.JUSTC-2024-0139
DOI: 10.52396/JUSTC-2024-0139
JUSTC, **2025**, 55(X):

value of the unknown parameter vector $(\mu,\boldsymbol{\alpha},\boldsymbol{\beta})^\top$. Since the sample is just one realization of the random graph, the likelihood function is the joint probability mass function that is the products of p_{ij} given by Eq. (1) over all the pairs $\{(i,j),1\le i,j\le n\}$. That is, the likelihood function is given by

$$L(\mu,\boldsymbol{\alpha},\boldsymbol{\beta})=\prod_{i=1}^{n}\prod_{j=1}^{n}\left(\frac{\mathrm{e}^{\mu+\alpha_i+\beta_j}}{1+\mathrm{e}^{\mu+\alpha_i+\beta_j}}\right)^{A_{ij}}\left(\frac{1}{1+\mathrm{e}^{\mu+\alpha_i+\beta_j}}\right)^{1-A_{ij}},$$

which implies the negative log-likelihood

$$\begin{aligned}l(\mu,\boldsymbol{\alpha},\boldsymbol{\beta})=&-\mu\left(\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}\right)-\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_iA_{ij}-\\&\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_jA_{ij}+\sum_{i=1}^{n}\sum_{j=1}^{n}\log(1+\mathrm{e}^{\mu+\alpha_i+\beta_j})=\\&-\mu d_+-\sum_{i=1}^{n}\alpha_id_{i+}-\sum_{j=1}^{n}\beta_jd_{+j}+\\&\sum_{i=1}^{n}\sum_{j=1}^{n}\log(1+\mathrm{e}^{\mu+\alpha_i+\beta_j}).\end{aligned}\tag{3}$$

Here we may consider two cases separately: (i) the supports of both $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are known; (ii) the supports of both $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are unknown. To describe our main results, we shall use the following notation. For two sequences of positive numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n=\Theta(b_n)$, if $0<\liminf_{n\to\infty}a_n/b_n\le\limsup_{n\to\infty}a_n/b_n<\infty$, and $a_n=\overline{o}(b_n)$, if there exists a constant $c>0$ such that $a_n=O(n^{-c}b_n)$. Similar definitions are also used for random sequences.

3.1 MLE with known supports

In this case, we assume that one has the prior knowledge of the supports of $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$. Let us denote by S_α and S_β the true support of $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$, respectively. Define $s_1=\|\boldsymbol{\alpha}\|_0$ and $s_2=\|\boldsymbol{\beta}\|_0$. Recall the reparameterization in (2). When the supports of $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are known, we only need to consider the estimation of μ^*, $\boldsymbol{\alpha}^*_S$ and $\boldsymbol{\beta}^*_S$. In what follows, we may also use similar notation, which is clear from the context.

Let $(\mu^*_0,\boldsymbol{\alpha}^*_{0S},\boldsymbol{\beta}^*_{0S})^\top$ and $(\hat{\mu}^*,\hat{\boldsymbol{\alpha}}^*_S,\hat{\boldsymbol{\beta}}^*_S)^\top$ be the true value and the MLE of $(\mu^*,\boldsymbol{\alpha}^*_S,\ \boldsymbol{\beta}^*_S)^\top$, respectively. Then the consistency and asymptotic normality of the MLE of these parameters as $n\to\infty$ is stated in the following.

Theorem 1. Assume that $|\mu^*|\le C_1,0\le\alpha^*_i\le C_2,0\le\beta^*_j\le C_3$, where $C_1,C_2,C_3>0$ are constants depending only on n such that $\max\{C_1,C_2,C_3\}=o(\log n)$.

(i) If $s_1=\overline{o}(n^{1-\alpha})$ and $s_2=\overline{o}(n^{1-\beta})$, then

$$\hat{\mu}^*=\mu^*_0+\overline{o}_p(1),\quad \max_{i\in S_\alpha}|\hat{\alpha}^*_i-\alpha^*_{0i}|=\overline{o}_p(1),\quad \max_{j\in S_\beta}|\hat{\beta}^*_j-\beta^*_{0j}|=\overline{o}_p(1).$$

(ii) If the true parameter value $(\mu^*_0,\boldsymbol{\alpha}^*_{0S},\boldsymbol{\beta}^*_{0S})$ lies in the interior of the parameter space $[-C_1,C_1]\times[0,C_2]^{s_1}\times[0,C_3]^{s_2}$, and $s_1=\overline{o}(n^{(1-\alpha)/2}),s_2=\overline{o}(n^{(1-\beta)/2})$, for any fixed subsets $A\subset S_\alpha$ and $B\subset S_\beta$, the following assertions hold.

(a) If $\gamma-\alpha-\beta\ge0$ and $\alpha,\beta>0$, then we have

$$\begin{pmatrix}n^{1-\frac{\gamma}{2}}e^{\frac{\mu^*_0}{2}}(\hat{\mu}^*-\mu^*_0)\\ n^{\frac{1}{2}-\frac{1}{2}(\gamma-\alpha)}\boldsymbol{\Lambda}_A^{-\frac{1}{2}}(\hat{\boldsymbol{\alpha}}^*_A-\boldsymbol{\alpha}^*_{0A})\\ n^{\frac{1}{2}-\frac{1}{2}(\gamma-\beta)}\boldsymbol{\Lambda}_B^{-\frac{1}{2}}(\hat{\boldsymbol{\beta}}^*_B-\boldsymbol{\beta}^*_{0B})\end{pmatrix}\xrightarrow{d}N(0,\boldsymbol{I}_{1+|A|+|B|}),$$

where $\boldsymbol{\Lambda}_A=\mathrm{diag}\{\mathrm{e}^{-\mu^*_0-\alpha^*_{0i}},i\in A\}$ and $\boldsymbol{\Lambda}_B=\mathrm{diag}\{\mathrm{e}^{-\mu^*_0-\beta^*_{0j}},j\in B\}$.

(b) If $0<\alpha=\gamma<1$ and $\beta=0$, then we have

$$\begin{pmatrix}n^{1-\frac{\gamma}{2}}e^{\frac{\mu^*_0}{2}}(\hat{\mu}^*-\mu^*_0)\\ n^{\frac{1}{2}-\frac{1}{2}(\gamma-\alpha)}\boldsymbol{\Lambda}_A^{-\frac{1}{2}}(\hat{\boldsymbol{\alpha}}^*_A-\boldsymbol{\alpha}^*_{0A})\\ n^{\frac{1}{2}-\frac{1}{2}(\gamma-\beta)}\boldsymbol{\Lambda}_B^{-\frac{1}{2}}(\hat{\boldsymbol{\beta}}^*_B-\boldsymbol{\beta}^*_{0B})\end{pmatrix}\xrightarrow{d}N(0,\boldsymbol{I}_{1+|A|+|B|}),$$

where $\boldsymbol{\Lambda}_A=\mathrm{diag}\{2+e^{-\mu^*_0-\alpha^*_{0i}}+e^{\mu^*_0+\alpha^*_{0i}},i\in A\}$ and $\boldsymbol{\Lambda}_B=\mathrm{diag}\{e^{-\mu^*_0-\beta^*_{0j}},j\in B\}$.

(c) If $0<\beta=\gamma<1$ and $\alpha=0$, then we have

$$\begin{pmatrix}n^{1-\frac{\gamma}{2}}e^{\frac{\mu^*_0}{2}}(\hat{\mu}^*-\mu^*_0)\\ n^{\frac{1}{2}-\frac{1}{2}(\gamma-\alpha)}\boldsymbol{\Lambda}_A^{-\frac{1}{2}}(\hat{\boldsymbol{\alpha}}^*_A-\boldsymbol{\alpha}^*_{0A})\\ n^{\frac{1}{2}-\frac{1}{2}(\gamma-\beta)}\boldsymbol{\Lambda}_B^{-\frac{1}{2}}(\hat{\boldsymbol{\beta}}^*_B-\boldsymbol{\beta}^*_{0B})\end{pmatrix}\xrightarrow{d}N(0,\boldsymbol{I}_{1+|A|+|B|}),$$

where $\boldsymbol{\Lambda}_A=\mathrm{diag}\{e^{-\mu^*_0-\alpha^*_{0i}},i\in A\}$ and $\boldsymbol{\Lambda}_B=\mathrm{diag}\{2+e^{-\mu^*_0-\beta^*_{0j}}+e^{\mu^*_0+\beta^*_{0j}},j\in B\}$.

(d) If $\alpha=\beta=\gamma=0$, then we have

$$\begin{pmatrix}n^{1-\frac{\gamma}{2}}(2+e^{-\mu^*_0}+e^{\mu^*_0})^{-\frac{1}{2}}(\hat{\mu}^*-\mu^*_0)\\ n^{\frac{1}{2}-\frac{1}{2}(\gamma-\alpha)}\boldsymbol{\Lambda}_A^{-\frac{1}{2}}(\hat{\boldsymbol{\alpha}}^*_A-\boldsymbol{\alpha}^*_{0A})\\ n^{\frac{1}{2}-\frac{1}{2}(\gamma-\beta)}\boldsymbol{\Lambda}_B^{-\frac{1}{2}}(\hat{\boldsymbol{\beta}}^*_B-\boldsymbol{\beta}^*_{0B})\end{pmatrix}\xrightarrow{d}N(0,\boldsymbol{I}_{1+|A|+|B|}),$$

where $\boldsymbol{\Lambda}_A=\mathrm{diag}\{2+e^{-\mu^*_0-\alpha^*_{0i}}+e^{\mu^*_0+\alpha^*_{0i}},i\in A\}$ and $\boldsymbol{\Lambda}_B=\mathrm{diag}\{2+e^{-\mu^*_0-\beta^*_{0j}}+e^{\mu^*_0+\beta^*_{0j}},j\in B\}$.

If we just look at the common case where $\gamma-\alpha-\beta>0$ for convenience of description, several immediate consequences of the proof of Theorem 1 are given below.

Corollary 1. Under the condition of Theorem 1, the expected number of edges in the DSβM is given by

$$\mathbb{E}[d_+]=n^{2-\gamma}\mathrm{e}^{\mu^*_0}+\overline{o}(n^{2-\gamma}).$$

Corollary 2. Under the condition of Theorem 1, the expected out-degree of any node i is given by

$$\mathbb{E}[d_{i+}]=\begin{cases}n^{1-(\gamma-\alpha)}e^{\mu^*_0+\alpha^*_{0i}}+\overline{o}(n^{1-(\gamma-\alpha)}), & i\in S_\alpha;\\ n^{1-\gamma}e^{\mu^*_0}+\overline{o}(n^{1-\gamma}), & i\notin S_\alpha,\end{cases}$$

whereas the expected in-degree of any node j is given by

$$\mathbb{E}[d_{+j}]=\begin{cases}n^{1-(\gamma-\beta)}e^{\mu^*_0+\beta^*_{0j}}+\overline{o}(n^{1-(\gamma-\beta)}), & j\in S_\beta;\\ n^{1-\gamma}e^{\mu^*_0}+\overline{o}(n^{1-\gamma}), & j\notin S_\beta.\end{cases}$$

The above corollaries show that the parameter γ (and thus μ) controls not only the density of edges in the entire network but also the degree of any single node, while the parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can distinguish nodes from one another. These results just match the interpretations we mentioned in Section 2.

3.2 MLE with unknown supports

In practice, the supports of $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are usually unknown, as are the quantities $\|\boldsymbol{\alpha}\|_0$ and $\|\boldsymbol{\beta}\|_0$. A natural method for estim-

CSTR: 32290.14.JUSTC-2024-0139
DOI: 10.52396/JUSTC-2024-0139
JUSTC, **2025**, 55(X):

ating of the parameters in such cases is to apply the l_0-norm constrained maximum likelihood estimation. Recall the negative log-likelihood in (3). Then we should consider the estimator as follows:

$$\begin{aligned}(\hat{\mu}(\boldsymbol{s}),\hat{\boldsymbol{\alpha}}(\boldsymbol{s}),\hat{\boldsymbol{\beta}}(\boldsymbol{s})) &= \mathop{\arg\min}_{\mu\in\mathbb{R},\,\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathbb{R}_+^n} l(\mu,\boldsymbol{\alpha},\boldsymbol{\beta})\\ \text{subject to}\quad &\|\boldsymbol{\alpha}\|_0 \le s_1, \|\boldsymbol{\beta}\|_0 \le s_2,\end{aligned} \tag{4}$$

where $0 \le s_1, s_2 \le n-1$ are integer-valued tuning parameters, and $\boldsymbol{s} = (s_1, s_2)^\top$.

However, a direct implementation of the optimization problem (4) seems computationally inefficient due to the combinatorial complexity. For any given pair (s_1, s_2), assuming that s_1 elements of $\boldsymbol{\alpha}$ and s_2 elements of $\boldsymbol{\beta}$ are not equal to 0, one may solve the solution of (4) to fit $\binom{n}{s_1}\cdot\binom{n}{s_2}$ possible models, and then choose the model that gives the smallest negative log-likelihood. Without any restriction on s_1 and s_2, it would lead to compare among $O(4^n)$ models in this approach, and thus the computational burden is quite expensive when n is large. Therefore, in practice we need a suitable method to significantly reduce the computational complexity.

We now introduce a more efficient approach to handle the optimization problem (4). Let $\boldsymbol{d}^{\text{out}} = (d_{1+},\cdots,d_{n+})^\top$ and $\boldsymbol{d}^{\text{in}} = (d_{+1},\cdots,d_{+n})^\top$ be out-degree and in-degree sequences in the network, respectively. Since nodes may share the common degrees, we first sort all the values of $\boldsymbol{d}^{\text{out}}$ and $\boldsymbol{d}^{\text{in}}$ as

$$d_{(1)+} > d_{(2)+} > \cdots > d_{(m_1)+}, \quad d_{+(1)} > d_{+(2)} > \cdots > d_{+(m_2)},$$

where m_1 and m_2 are some integers not larger than n. We also define two node sets $\mathcal{A}_k = \{i : d_{i+} = d_{(k)+}, 1 \le i \le n\}$ for $0 \le k \le m_1$ and $\mathcal{B}_l = \{j : d_{+j} = d_{+(l)}, 1 \le j \le n\}$ for $0 \le l \le m_2$. Denote their cardinalities by $a_k = |\mathcal{A}_k|$ and $b_l = |\mathcal{B}_l|$. Then it follows by definition that $\sum_{k=1}^{m_1} a_k = \sum_{l=1}^{m_2} b_l = n$. The following result plays an important role in our approach.

Lemma 1. The MLE estimate $(\hat{\boldsymbol{\alpha}}(\boldsymbol{s}),\hat{\boldsymbol{\beta}}(\boldsymbol{s}))$ in the optimization problem (4) has the following properties.

(i) If $d_{i+} < d_{j+}$, then we have $\hat{\alpha}_i(\boldsymbol{s}) \le \hat{\alpha}_j(\boldsymbol{s})$,

(ii) If $d_{i+} = d_{j+}$, then we have $\hat{\alpha}_i(\boldsymbol{s}) = \hat{\alpha}_j(\boldsymbol{s})$,

(iii) If $d_{+i} < d_{+j}$, then we have $\hat{\beta}_i(\boldsymbol{s}) \le \hat{\beta}_j(\boldsymbol{s})$,

(iv) If $d_{+i} = d_{+j}$, then we have $\hat{\beta}_i(\boldsymbol{s}) = \hat{\beta}_j(\boldsymbol{s})$.

The proof of Lemma 1 can be found in Section 2 in the supplementary materials. Using Lemma 1, we can find that for each s_1, we only have to sort the out-degree of nodes, and then assign the s_1 largest out-degree nodes with nonzero parameters. A similar approach can be applied for the in-degree situation. Compared with the method in Ref. 17], using the additional information of nodal covariates their proposed model also has the selection consistency, in the sense that the true support of $(\boldsymbol{\alpha},\boldsymbol{\beta})$ can be found successfully.

Fortunately, due to this lemma, we can reduce the time complexity to the order $O(n^2)$, where the time complexity of the exhaustive attack method is $O(4^n)$. The efficiency of this method is significant. More precisely, considering the ties that might exist in the degree sequence, we can change the optimization problem (4) to another optimization problem below to ensure that the MLE is unique:

$$\begin{aligned}(\hat{\mu}(\boldsymbol{s}),\hat{\boldsymbol{\alpha}}(\boldsymbol{s}),\hat{\boldsymbol{\beta}}(\boldsymbol{s})) &= \mathop{\arg\min}_{\mu\in R,\boldsymbol{\alpha},\boldsymbol{\beta}\in R_+^n} l(\mu,\boldsymbol{\alpha},\boldsymbol{\beta})\\ \text{subject to}\quad &s_1 = \sum_{i=1}^{K} a_i, s_2 = \sum_{j=1}^{L} b_j,\end{aligned} \tag{5}$$

where $K \in \{1,2,\cdots,m-1\}$ and $L \in \{1,2,\cdots,l-1\}$.

It is remarkable that from the proof presented in Section 2 in the supplementary materials, Lemma 1 heavily relies on the assumption that self-loops are allowed in the model. Once this assumption is invalid, Lemma 1 is no longer applicable[12]. On the other hand, only a small proportion of nodes in a large sparse network have self-loops, and the contribution of a self-loop to the degree of such a node with high degree (in-degree or out-degree) is only 1, so self-loops have very limited influence on the ordering of degrees, especially for the nodes whose assigned parameters are not equal to 0. We shall implement extensive simulations to evaluate the performance of our proposed method in the networks without self-loops in the next section.

In the next lemma, we show that the elements of $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ could not be too small, i.e., the smallest one among them should be above a certain threshold.

Lemma 2. Let $\tau \in (0,1)$ be a given small constant. For any $i \in S_\alpha$ and $j \in S_\alpha^c$, if

$$\alpha_{0i} > \log\left[1 + c_{n,\tau}(1+\mathrm{e}^{\mu^-})(1+\mathrm{e}^{\mu^+ + \overline{\alpha} + \overline{\beta}})\right], \tag{6}$$

where $c_{n,\tau} = \sqrt{(2/n)\log(2/\tau)}$, $\overline{\alpha} = \max_{1\le k\le n}\alpha_{0k}$, $\overline{\beta} = \max_{1\le k\le n}\beta_{0k}$, $\mu^+ = \max\{\mu_0, 0\}$, and $\mu^- = \max\{-\mu_0, 0\}$, then with probability $1-\tau$, we have $d_{i+} > d_{j+}$. Similarly, for any $i \in S_\beta$ and $j \in S_\beta^c$, if β_{0i} is greater than the term on the right-hand side of (6), then $d_{+i} > d_{+j}$ also holds with probability $1-\tau$.

We present the proof of Lemma 2 in Section 3 in the supplementary materials. Due to arbitrariness of τ, using the union bound one can immediately obtain a corollary of Lemma 2 in the following.

Corollary 3. Under the same conditions in Lemma 2, if

$$\min\left\{\min_{i\in S_\alpha}\alpha_{0i}, \min_{i\in S_\beta}\beta_{0i}\right\} > \log\left[1 + c_{n,\tau/n(n-1)}(1+e^{\mu^-})(1+e^{\mu^+ + \overline{\alpha} + \overline{\beta}})\right],$$

then both $\min_{i\in S_\alpha} d_{i+} > \max_{j\in S_\alpha^c} d_{j+}$ and $\min_{i\in S_\beta} d_{+i} > \max_{j\in S_\beta^c} d_{+j}$ hold with probability at least $1-2\tau$.

We next evaluate the prediction risk for our estimator. We define $R(\mu,\boldsymbol{\alpha},\boldsymbol{\beta})$ as the risk of the parameter value $(\mu,\boldsymbol{\alpha},\boldsymbol{\beta})$. Specifically, we write

$$R(\mu,\boldsymbol{\alpha},\boldsymbol{\beta}) = \mathbb{E}\left[\frac{l(\mu,\boldsymbol{\alpha},\boldsymbol{\beta})}{\mathbb{E}[d_+]}\right].$$

Following the results proposed by Ref. [22], we can evaluate the performance of the estimator $(\hat{\mu}(\boldsymbol{s}),\hat{\boldsymbol{\alpha}}(\boldsymbol{s}),\hat{\boldsymbol{\beta}}(\boldsymbol{s}))$ by the excess risk as follows:

$$\mathcal{E}_s = R(\hat{\mu}(\boldsymbol{s}),\hat{\boldsymbol{\alpha}}(\boldsymbol{s}),\hat{\boldsymbol{\beta}}(\boldsymbol{s})) - \inf_{(\mu,\boldsymbol{\alpha},\boldsymbol{\beta})\in\Theta_s} R(\mu,\boldsymbol{\alpha},\boldsymbol{\beta}),$$

where $\Theta_s = (\mu,\boldsymbol{\alpha},\boldsymbol{\beta}) : |\mu| \le M_1, \boldsymbol{\alpha} \in [0,M_2]^n, \boldsymbol{\beta} \in [0,M_3]^n, \|\boldsymbol{\alpha}\|_0 \le s_1$, and $\|\boldsymbol{\beta}\|_0 \le s_2$. Then the following theorem gives an upper bound on the excess risk $\mathcal{E}_s$.

Theorem 2. We let $D_+ = \mathbb{E}[d_+]$. Then for any fixed $\tau \in (0,1)$, we have

CSTR: 32290.14.JUSTC-2024-0139
DOI: 10.52396/JUSTC-2024-0139
JUSTC, **2025**, 55(X):

$$\mathcal{E}_s \le \frac{2}{D_+}\left[M_1\left(\sqrt{2\mathrm{Var}(d_+)\log\frac{6}{\tau}}+\frac{1}{3}\log\frac{6}{\tau}\right)+ M_2 s_1\cdot\left(\sqrt{2\max_j \mathrm{Var}(d_{j+})\log\frac{6n}{\tau}}+\frac{1}{3}\log\frac{6n}{\tau}\right)+ M_3 s_2\cdot\left(\sqrt{2\max_k \mathrm{Var}(d_{+k})\log\frac{6n}{\tau}}+\frac{1}{3}\log\frac{6n}{\tau}\right)\right].$$

with probability $1-\tau$.

Moreover, under the same conditions of Theorem 1, i.e., the quantities M_1, M_2, M_3 are all of the order $O(\log n)$, we can obtain the following result.

Corollary 4. Specifically, under the reparameterization in Eq. (2) and the same conditions of Theorem 1, we can express the excess risk $\mathcal{E}_s$ bounds by

$$\mathcal{E}_s = O_p\left(\frac{\log n}{n^{1-\frac{\gamma}{2}}}+\frac{s_1(\log n)^{\frac{3}{2}}}{n^{\frac{3}{2}-\frac{\gamma+\alpha}{2}}}+\frac{s_2(\log n)^{\frac{3}{2}}}{n^{\frac{3}{2}-\frac{\gamma+\beta}{2}}}\right).$$

Noting that $s_1=\overline{o}(n^{1-\alpha})$ and $s_2=\overline{o}(n^{1-\beta})$, we can easily obtain that the excess risk $\mathcal{E}_s$ tends to be 0 with high probability as $n\to\infty$, which means that our method works significantly. In the next section, we will use this method to perform some simulations to show it.

4 Simulation results

In this section, we demonstrate the effectiveness of our estimator in performing simultaneous parameter estimation and model selection consistently. The first question is how to choose the parameter sparsity s_1 and s_2. To prevent overfitting the model, a natural approach is to use Bayesian information criterion (BIC). Precisely, we define

$$\mathrm{BIC}(s_1,s_2) = 2l(\hat{\mu},\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) + (s_1+s_2)\log(n^2). \tag{7}$$

Then, our goal is to choose s_1 and s_2 to minimize Eq. (7), which means

$$(s_1,s_2) = \arg\min\left\{\mathrm{BIC}(s_1,s_2): s_1=\sum_{i=1}^{K}a_i, s_2=\sum_{j=1}^{L}b_j\right\}, \tag{8}$$

where $K\in\{1,2,\cdots,m-1\}$ and $L\in\{1,2,\cdots,l-1\}$.

Next, we choose some situations to simulate the generation of networks and verify some interesting properties. These situations are given below:

(a) $\mu_0=-1.5, \alpha_{0i}=1.5, \beta_{0j}=1.5$ for $i\in S_\alpha, j\in S_\beta$;

(b) $\mu_0=-1.5, \alpha_{0i}=1.5, \beta_{0j}=\sqrt{\log n}$ for $i\in S_\alpha, j\in S_\beta$;

(c) $\mu_0=-\sqrt{\log n}, \alpha_{0i}=1.5, \beta_{0j}=1.5$ for $i\in S_\alpha, j\in S_\beta$;

(d) $\mu_0=-\sqrt{\log n}, \alpha_{0i}=1.5, \beta_{0j}=\sqrt{\log n}$ for $i\in S_\alpha, j\in S_\beta$;

where $n=50,100,200,400$. Moreover, for each situation, we choose parameter sparsity under three cases: (i) $s_1=10$, $s_2=10$, (ii) $s_1=10$, $s_2=15$, and (iii) $s_1=15$, $s_2=20$. In regard to the details of the simulation, we repeat 1000 times for each case to determine how many times our result of support matches the truth.

Table 1 reports the frequencies when the support is correctly identified for each case. These results indicate that as n grows, our judgment works well. In most cases, our accuracy rate can reach at least 0.8. For the cases where the wrong support sets are obtained, we can find that in most of the repetitions, our method has identified all the true support sets but overestimated with only a single index, which may be caused by the randomness of the generated graphs. Following the idea of the receiver operating characteristic, one can obtain that the ratio of the true positive rate (TPR) to the false positive rate (FPR) is very large. Therefore, we can recorrect our frequencies by adding the wrong case mentioned above to the truly identified cases, and the recorrection frequencies are computed in the parentheses in Table 1, which reaches nearly 1. In addition, we notice that when μ decreases, the frequencies when the support set is correctly identified might be lower. Based on the discussion in Section 3, we find the

Table 1. Simulation results for DSβM with self-loops on frequencies of the true support selected by BIC.

Model	n	$s_1=10,\ s_2=10$	$s_1=10,\ s_2=15$	$s_1=15,\ s_2=20$
(a)	50	0.570(0.704)	0.534(0.650)	0.416(0.512)
	100	0.802(0.984)	0.828(0.992)	0.844(0.990)
	200	0.794(0.986)	0.844(0.984)	0.814(0.996)
	400	0.834(0.990)	0.860(0.990)	0.850(0.998)
(b)	50	0.676(0.806)	0.654(0.798)	0.572(0.668)
	100	0.788(0.990)	0.838(0.984)	0.830(0.996)
	200	0.798(0.996)	0.818(1.000)	0.820(0.988)
	400	0.804(0.996)	0.816(0.986)	0.802(0.988)
(c)	50	0.402(0.508)	0.326(0.404)	0.284(0.328)
	100	0.800(0.968)	0.808(0.968)	0.804(0.970)
	200	0.828(0.988)	0.804(0.998)	0.846(0.994)
	400	0.800(0.996)	0.798(0.994)	0.836(0.998)
(d)	50	0.550(0.662)	0.548(0.678)	0.518(0.608)
	100	0.786(0.990)	0.812(0.978)	0.814(0.972)
	200	0.828(0.990)	0.812(0.988)	0.810(0.994)
	400	0.850(0.998)	0.820(0.996)	0.818(0.996)

The numbers in parentheses represent the frequencies after the correction.

CSTR: 32290.14.JUSTC-2024-0139
DOI: 10.52396/JUSTC-2024-0139
JUSTC, **2025**, 55(X):

support set mainly via the degree heterogeneity. However, an excessively small value of μ may contribute to a nearly sparse network, which means that the degree heterogeneity is hard to detect.

Following the interpretation in Ref. [23], a generalized information criterion (GIC) is constructed as follows:

$$\mathrm{GIC}(s_1, s_2) = 2l(\hat{\mu}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) + a_n \cdot (s_1 + s_2), \tag{9}$$

where a_n is a positive sequence that depends on the sample size n and controls the penalty on model complexity. The rationale of the information criteria for model selection is that the true model can uniquely optimize the information criterion (9) by appropriately choosing a_n[24]. In BIC, we choose $a_n = \log n$, but from the discussion above, we can find that the penalty term is not high enough. Therefore, we can choose $a_n = 2\log n$, which means

$$\mathrm{GIC}(s_1, s_2) = 2l(\hat{\mu}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) + 2(s_1 + s_2)\log n^2. \tag{10}$$

By changing the information criterion from BIC to GIC, Fig. 1 shows that the accuracy rate reaches nearly 1. Based on the result, using GIC might be a nice approach when we use DSβM to model the real network. Next, we consider the error of the global and local density parameter estimators, which is evaluated by the mean of the absolute error. The results in Figs. 2–4 indicate that our solution works well. This result is natural since our method is unbiased, which has also been confirmed in Section 3.

From the simulation results, we can conclude that our method works well in practice for various networks with degree heterogeneity. Moreover, even when the network is nearly sparse, our model can also fit it excellently. Table 1 and Figs. 1–4 provide strong support for our conclusions.

Finally, considering that most of the directed networks do not have self-loops, we might also apply our method to these cases to judge the discussions mentioned in Section 3, although our monotonicity lemma might no longer be applicable. The main approach is to remove the self-loops in the original simulation, then we compare the results between these

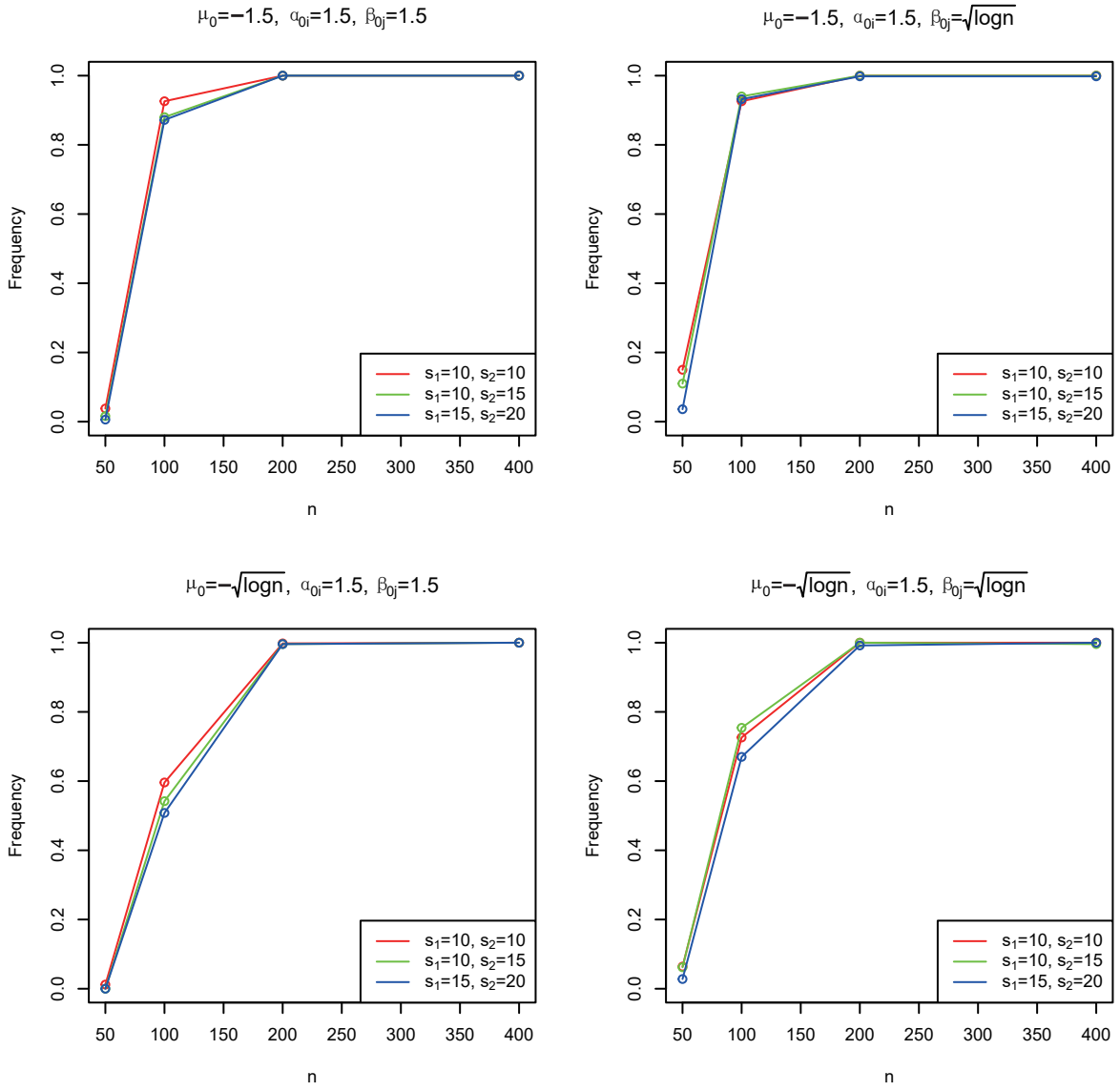


Fig. 1. Simulation results for DSβM with self-loops on frequencies of the true support selected by GIC.

CSTR: 32290.14.JUSTC-2024-0139
DOI: 10.52396/JUSTC-2024-0139
JUSTC, **2025**, 55(X):

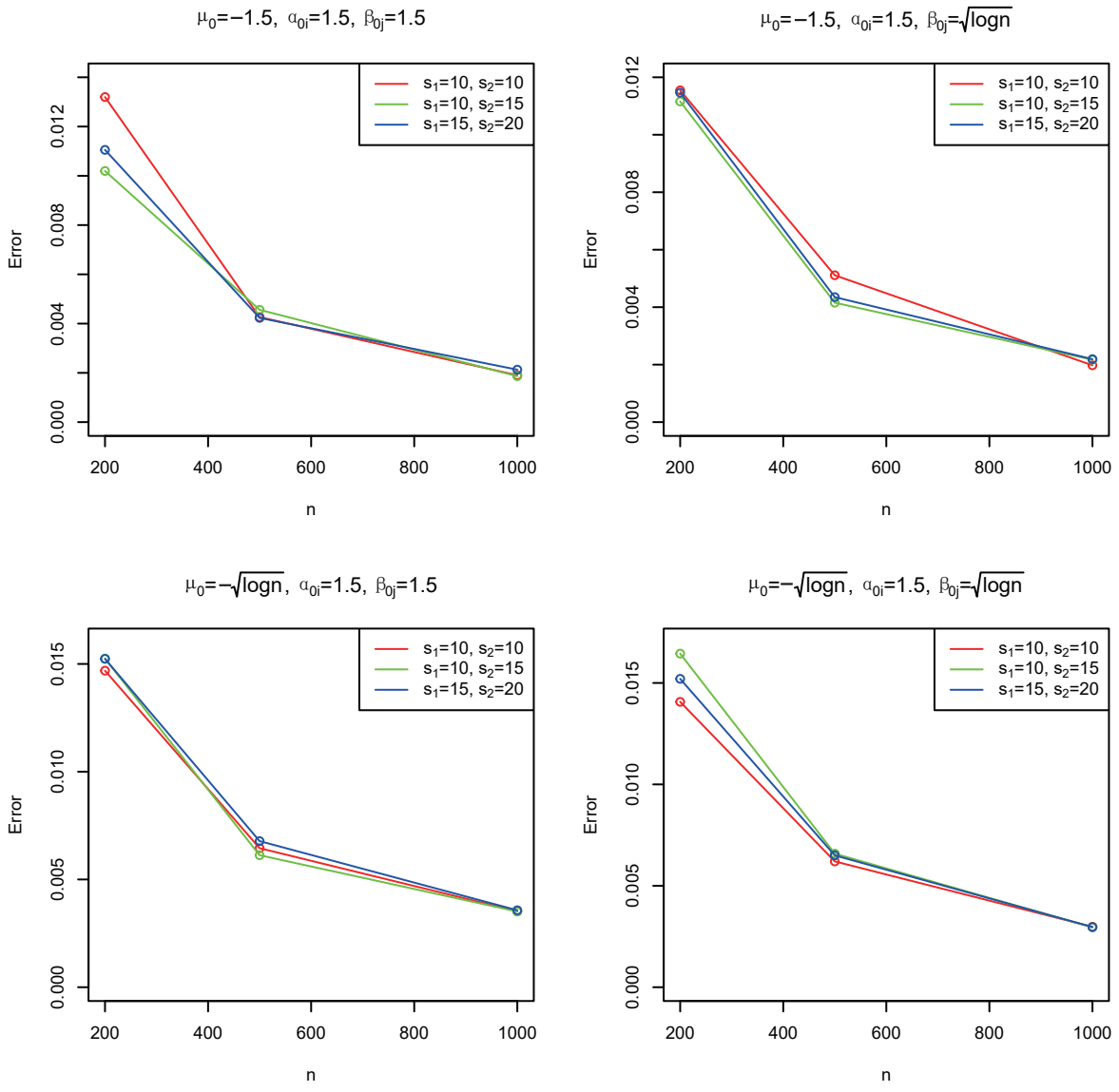
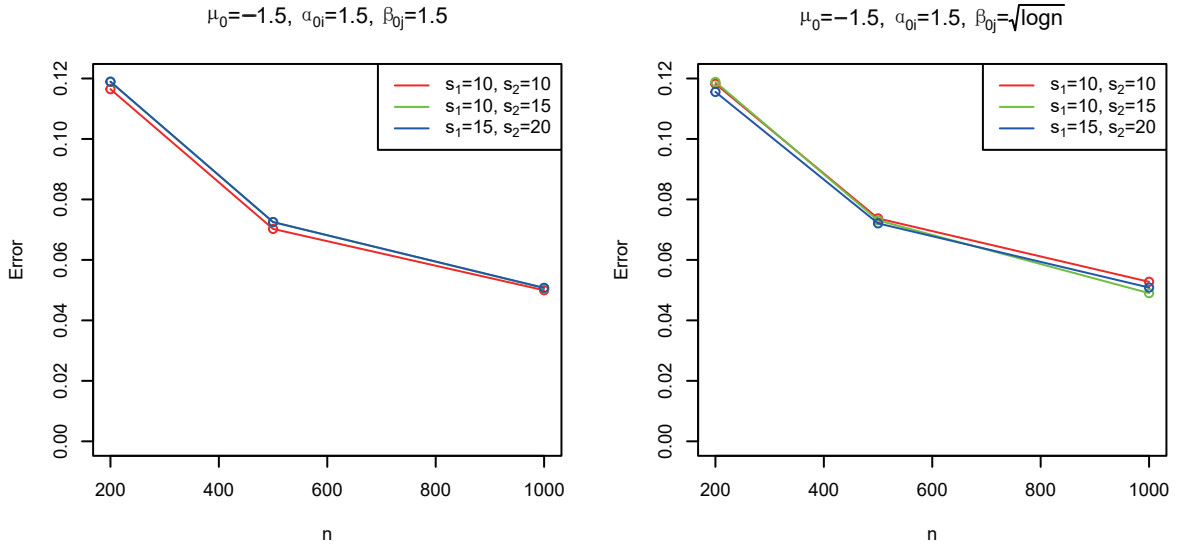


Fig. 2. Simulation results for DSβM with self-loops on the mean error of $|\hat{\mu}(\hat{s}_1,\hat{s}_2)-\mu_0|$ with (s_1,s_2) truly selected by BIC.



CSTR: 32290.14.JUSTC-2024-0139
DOI: 10.52396/JUSTC-2024-0139
JUSTC, **2025**, 55(X):

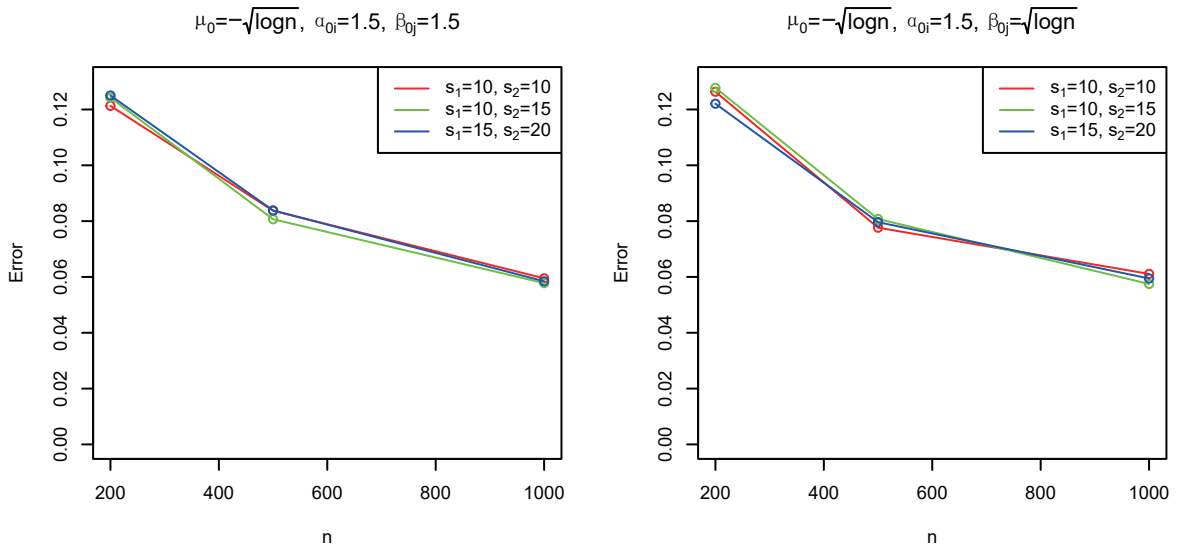


Fig. 3. Simulation results for DSβM with self-loops on the mean error of $|\hat{\alpha}(\hat{s}_1,\hat{s}_2)-\alpha_0|$ with (s_1,s_2) truly selected by BIC.

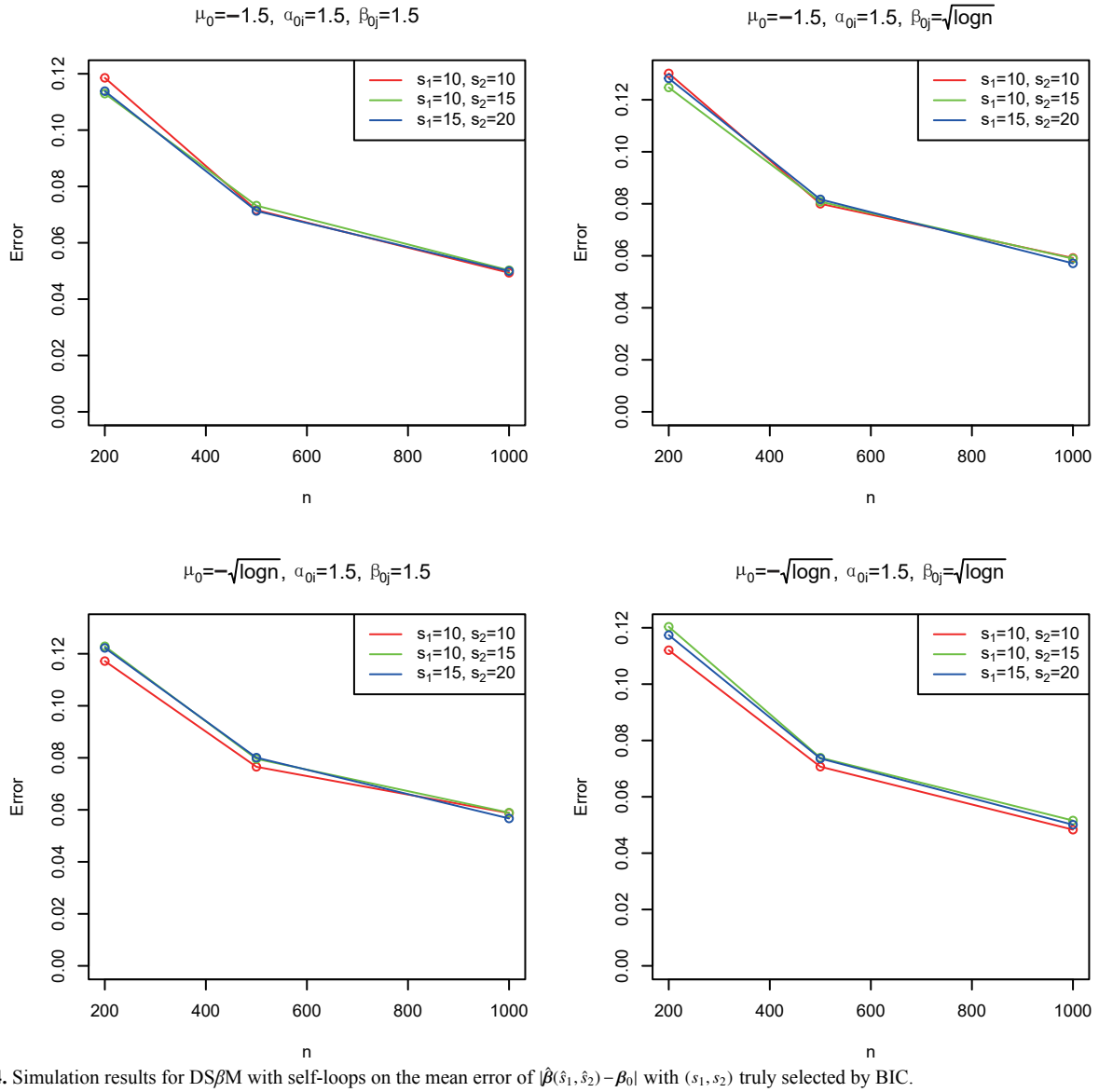


Fig. 4. Simulation results for DSβM with self-loops on the mean error of $|\hat{\boldsymbol{\beta}}(\hat{s}_1,\hat{s}_2)-\boldsymbol{\beta}_0|$ with (s_1,s_2) truly selected by BIC.

CSTR: 32290.14.JUSTC-2024-0139
DOI: 10.52396/JUSTC-2024-0139
JUSTC, **2025**, 55(X):

Table 2. Simulation results for DSβM with no self-loops on frequencies of the true support selected by BIC.

Model	n	$s_1 = 10,\ s_2 = 10$	$s_1 = 10,\ s_2 = 15$	$s_1 = 15,\ s_2 = 20$
(a)	50	0.486(0.704)	0.468(0.650)	0.314(0.512)
	100	0.814(0.984)	0.836(0.992)	0.842(0.990)
	200	0.798(0.986)	0.824(0.984)	0.820(0.996)
	400	0.824(0.990)	0.868(0.990)	0.850(0.998)
(b)	50	0.614(0.806)	0.574(0.798)	0.480(0.668)
	100	0.792(0.990)	0.838(0.984)	0.824(0.996)
	200	0.796(0.996)	0.824(1.000)	0.814(0.988)
	400	0.798(0.996)	0.826(0.986)	0.810(0.988)
(c)	50	0.324(0.508)	0.230(0.404)	0.192(0.328)
	100	0.802(0.968)	0.810(0.968)	0.804(0.970)
	200	0.826(0.988)	0.798(0.998)	0.850(0.994)
	400	0.810(0.996)	0.798(0.994)	0.836(0.998)
(d)	50	0.470(0.662)	0.474(0.678)	0.402(0.608)
	100	0.778(0.990)	0.806(0.978)	0.826(0.972)
	200	0.822(0.990)	0.810(0.988)	0.822(0.994)
	400	0.852(0.998)	0.824(0.996)	0.820(0.996)

The numbers in parentheses represent the frequencies after the correction.

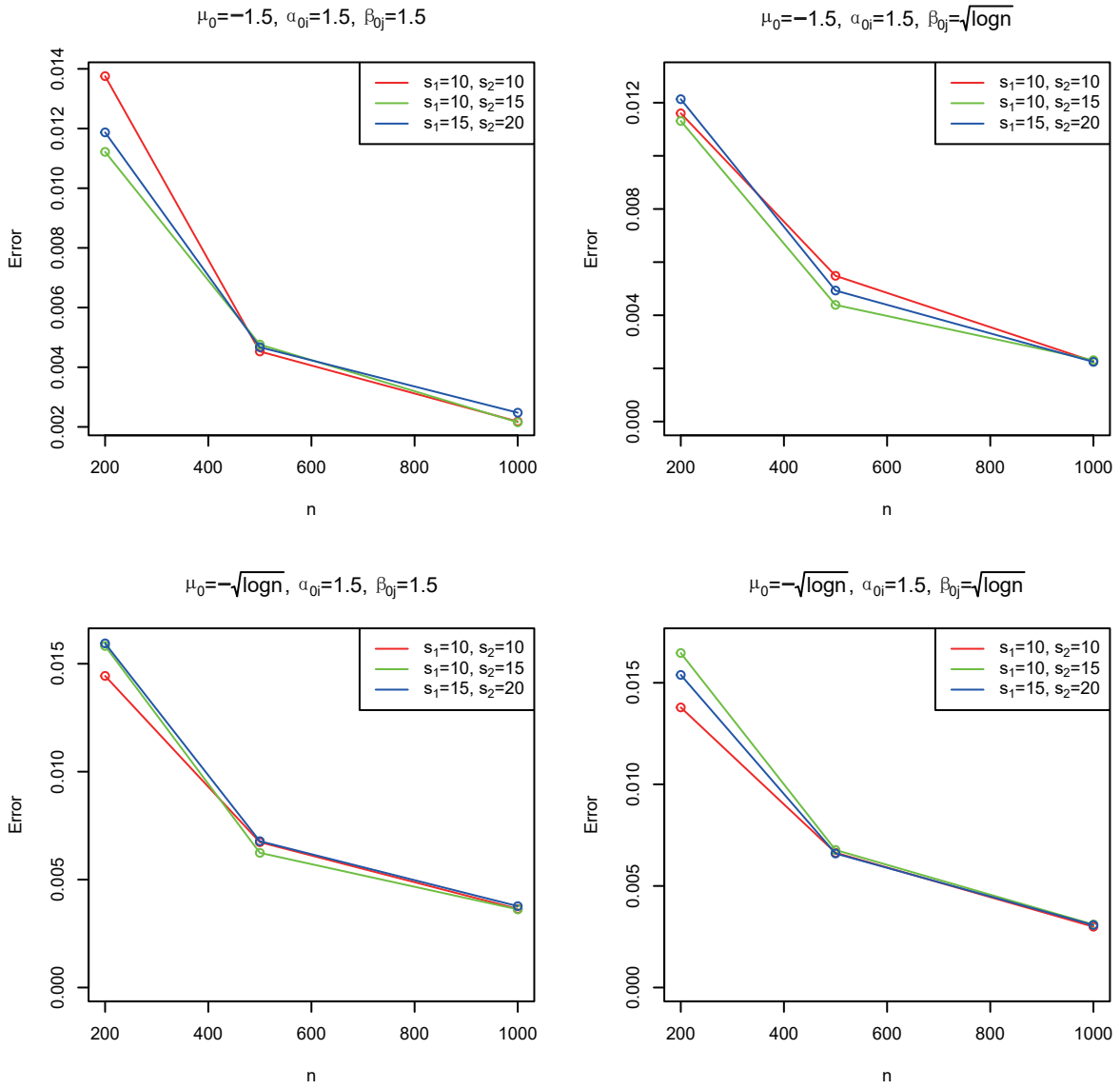


Fig. 5. Simulation results for DSβM with no self-loops on the mean error of $|\hat{\mu}(\hat{s}_1, \hat{s}_2) - \mu_0|$ with (s_1, s_2) truly selected by BIC.

CSTR: 32290.14.JUSTC-2024-0139
DOI: 10.52396/JUSTC-2024-0139
JUSTC, **2025**, 55(X):

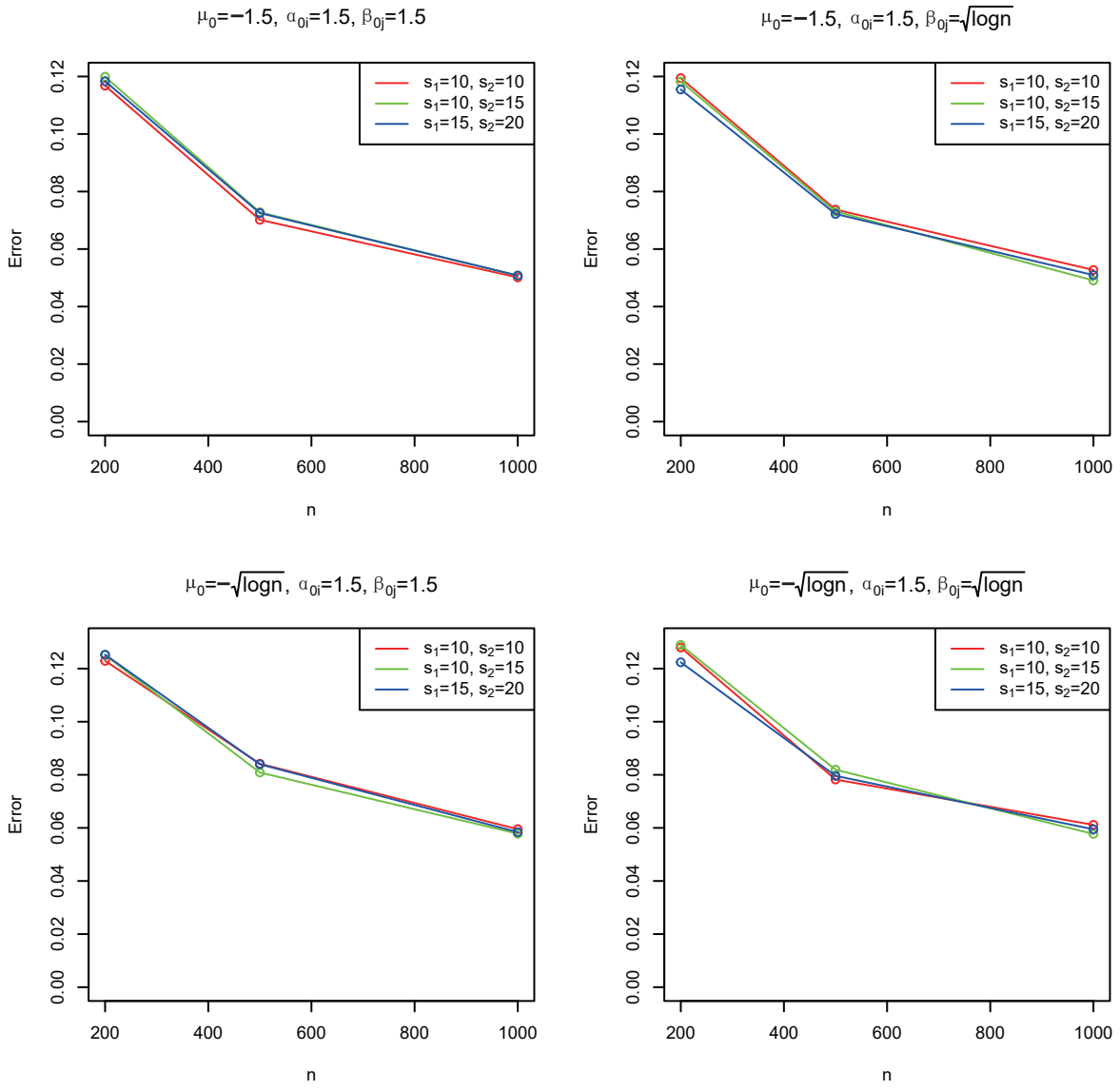


Fig. 6. Simulation results for DSβM with no self-loops on the mean error of $|\hat{\boldsymbol{\alpha}}(\hat{s}_1,\hat{s}_2)-\boldsymbol{\alpha}_0|$ with (s_1,s_2) truly selected by BIC.

two simulation groups. Table 2 reports the frequencies when the support is correctly identified for each case. From the results we can find that the frequencies after removing self-loops do not change substantially. When the number of nodes in the network is 400, the difference in frequencies is even within 0.01. Moreover, as n tends to grow larger, the difference between the two simulations becomes smaller, which means that our method still works well. Meanwhile, we also put Figs. 5–7 that describe the corresponding parameter estimation. By comparing the results with Figs. 2–4, we can find that there is little difference between the two errors, so we could try to apply this method to real networks, which will be stated in the next section. We must notice that the percentage of nodes with self-loops in the network is really low. That is to say, in our simulation, less than 10% of nodes have self-loops on average, leading to the fact that the structure of the network does not change much after removing the self-loops, and if we set the support of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to different nodes, this proportion could be much lower. This also explains the slight difference between the two simulation groups, and makes our method more reasonable.

5 Data analysis

In this section, we analyze the political blogs data collected by Ref. [25]. This dataset can be downloaded from http://www-personal.umich.edu/~mejn/netdata/. This dataset describes a directed network of hyperlinks between weblogs on US politics, recorded in 2005 by Adamic and Glance. Each node in the dataset has an attribute description (denoted by 0 or 1), which indicates democratic or conservative. The original data contain 1490 nodes and 19090 edges. We mainly focus on the edge connection rule of this network to fit our DSβM. First, we remove the nodes whose in-degree or out-degree is zero. Then we can obtain a subgraph without so many isolated nodes, which contains 831 nodes and 16110 edges.

Now we try to fit our model to this network. First, we have to choose the parameter sparsity s_1 and s_2. In this data analysis, we consider $\max\{s_1,s_2\}\le n/2$ (n is the total number of

CSTR: 32290.14.JUSTC-2024-0139
DOI: 10.52396/JUSTC-2024-0139
JUSTC, **2025**, 55(X):

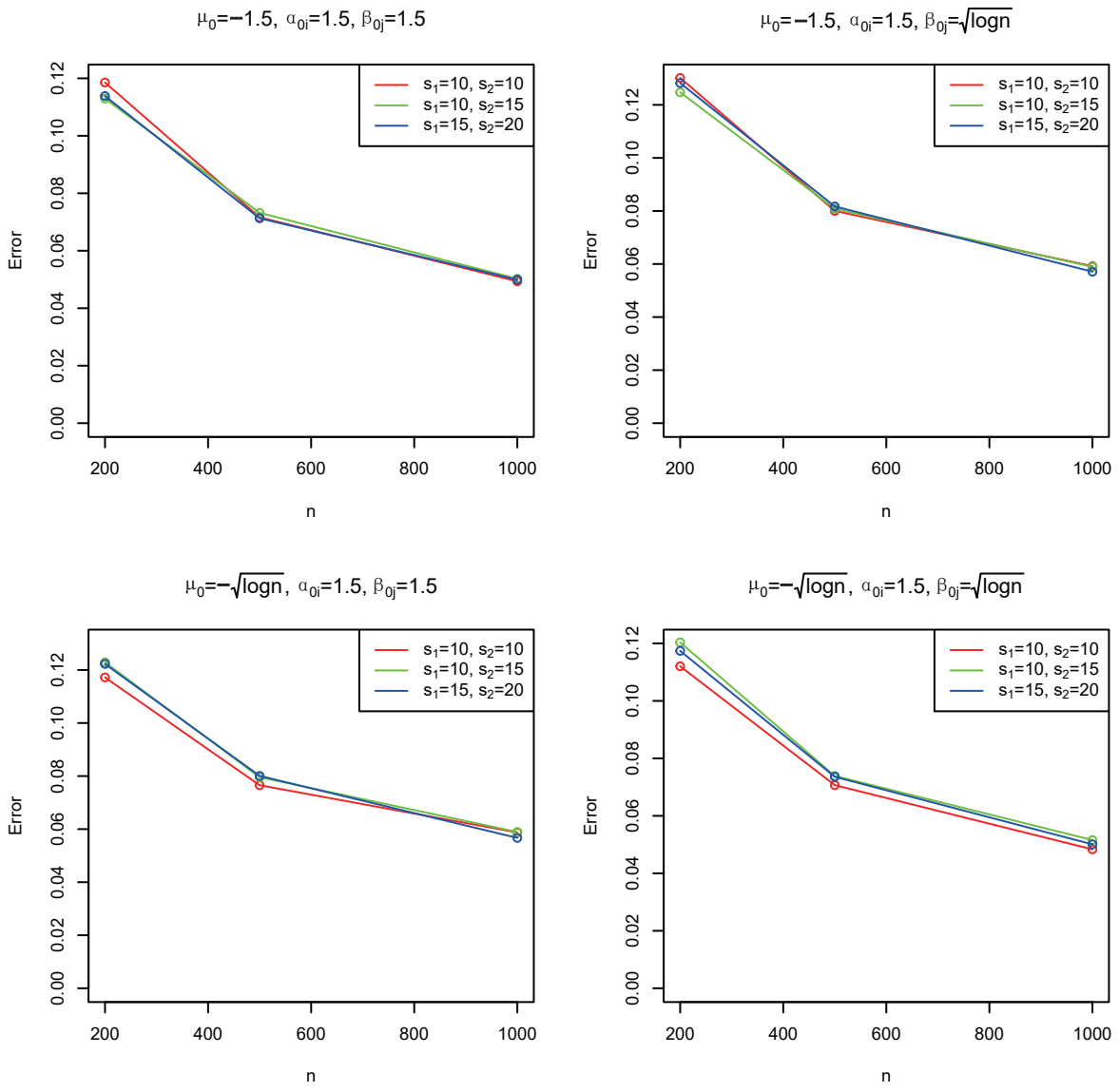


Fig. 7. Simulation results for DSβM with no self-loops on the mean error of $|\hat{\boldsymbol{\beta}}(\hat{s}_1,\hat{s}_2)-\boldsymbol{\beta}_0|$ with (s_1,s_2) truly selected by BIC.

Table 3. The top 30 largest estimators of α_i and β_j in the political blogs dataset.

Node	d_{i+}	$\hat{\alpha}_i$	Node	d_{+j}	$\hat{\beta}_j$	Node	d_{i+}	$\hat{\alpha}_i$	Node	d_{+j}	$\hat{\beta}_j$
436	217	4.448	78	273	5.043	521	85	2.924	629	122	3.767
218	123	3.472	556	248	4.870	35	84	2.907	373	119	3.733
181	122	3.459	35	228	4.723	468	84	2.907	92	113	3.662
249	119	3.421	318	223	4.686	28	82	2.873	239	113	3.662
454	114	3.355	674	200	4.505	608	82	2.873	452	113	3.662
172	105	3.230	436	197	4.481	62	79	2.821	232	109	3.613
254	104	3.216	509	188	4.406	320	78	2.804	711	108	3.601
530	104	3.216	615	182	4.355	437	78	2.804	258	103	3.537
587	102	3.187	360	181	4.347	556	77	2.786	204	101	3.511
760	100	3.158	800	171	4.258	238	76	2.768	140	100	3.498
75	97	3.113	595	163	4.186	279	75	2.750	530	100	3.498
53	92	3.037	550	144	4.002	631	75	2.750	726	100	3.498
489	89	2.989	154	140	3.961	826	74	2.731	815	100	3.498
308	88	2.973	826	128	3.834	38	71	2.675	587	99	3.485
103	86	2.940	319	125	3.801	307	71	2.675	816	99	3.485

The global density parameter $\mu = -6.517$.

CSTR: 32290.14.JUSTC-2024-0139
DOI: 10.52396/JUSTC-2024-0139
JUSTC, **2025**, 55(X):

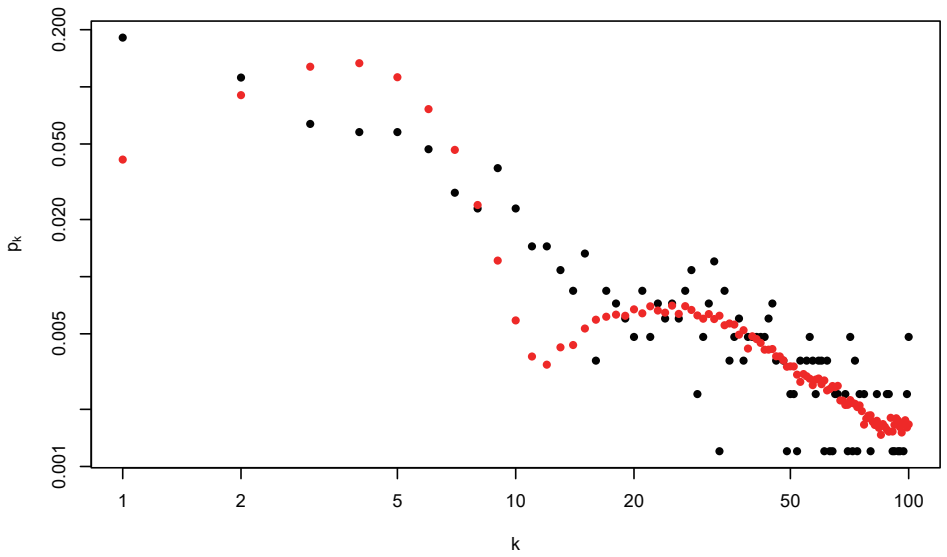


Fig. 8. Degree distribution for the in-degree case. The black points represent the real data distribution and the red points represent the average simulation result distribution.

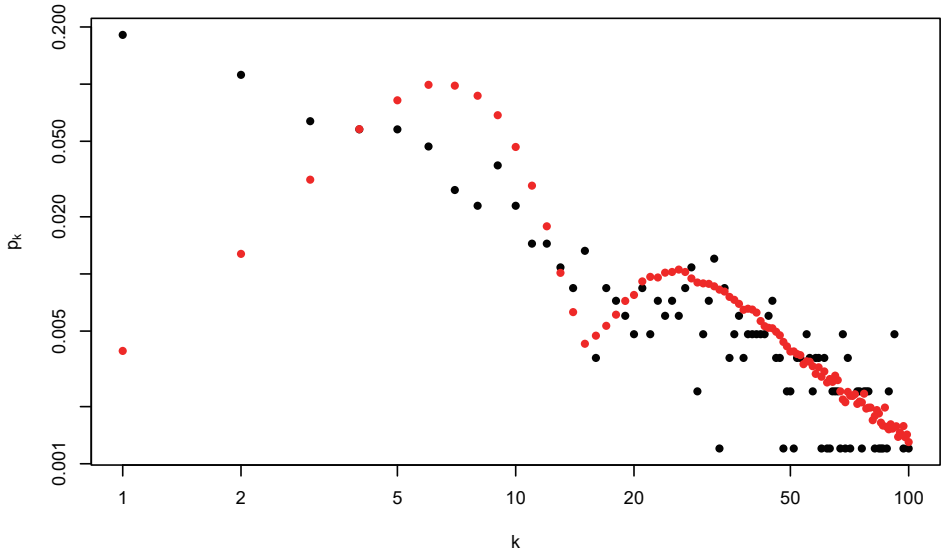


Fig. 9. Degree distribution for the out-degree case. The black points represent the real data distribution and the red points represent the average simulation result distribution.

nodes). Using the BIC proposed in Section 4, we can obtain that 35% of the nodes for the out-degree case and 32% for the in-degree case have nonzero parameters. We put some of the MLE estimators in Table 3. We can find that the estimators just match our results. That is, the larger the estimators are, the larger the degrees will be, which makes sense intuitively.

In addition, we use the degree distribution to evaluate our model. We repeat 100 times to generate networks, calculate the average degree distribution, and then compare the results with the real data. Figs. 8 and 9 show the in-degree and out-degree distributions. The results indicate that our DSβM can fit these data to a certain extent.

Finally, about this dataset, different studies have been conducted on it. Ref. [25] studied the linking patterns and discussion topics. Their aim was to measure the degree of interaction between democratic and conservative blogs, and to uncover any differences in the structure of the two communities. They found differences in the behavior of democratic and conservative blogs, with conservative blogs linking to each other more frequently and in a denser pattern. Although their method is different from ours, their conclusions just match those in our paper in a sense.

6 Conclusions

In this paper, we have improved the sparse β model for the directed network by allowing self-loops, which can capture degree heterogeneity. We have derived several interesting properties such as consistency and asymptotic normality for our model by reparameterizing the global and local density parameters when the support is known. For the other case when the support is unknown, we use the l_0 penalized likelihood approach to estimate the parameter. The only problem

CSTR: 32290.14.JUSTC-2024-0139
DOI: 10.52396/JUSTC-2024-0139
JUSTC, **2025**, 55(X):

for this method is the heavy time complexity. Fortunately, the monotonicity lemma can help us solve this problem. This lemma's contribution is significant, which means that we only have to select the best solution from at most n^2 models. Then we denote the condition to identify whether the estimator from our approach can work well with high probability. In addition, we use the excess risk to evaluate the prediction risk for our estimator. Finally, we also apply our method to those directed networks without self-loops and perform real data analysis to judge our results, although our monotonicity lemma might no longer be applicable. In summary, through our improvement, the SβM proposed by Ref. [12] can be applied for directed networks, thus modeling nearly sparse directed networks significantly.

However, our final purpose is to model the p_1 model, which was first proposed by Ref. [1]. Unfortunately, although we have proposed the model (1) for directed networks, which is a special case of the p_1 model (also known as the p_0 model), there are not many real networks that have self-loops. If we delete the self-loop assumption, we can find that the monotonicity lemma is no longer applicable, leading to the result that the time complexity cannot be reduced. Seeing this situation, we may choose other penalized likelihood estimations. Ref. [17] proposed a new parameter-sparse random graph model for density-sparse directed networks, which is called the SRGM, with l_1-norm penalized likelihood estimation. Similar contributions have also been made to networks with covariates[13, 16]. Based on this situation, our next step is to improve our model without self-loops, and then focus on the parameter estimation for the p_1 model, which is used to fit the real sparse network well.

Supporting information

The supporting information for this article can be found online at http://doi.org/10.52396/JUSTC-2024-0139. It includes the proofs of all the theorems and lemmas.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (11771418).

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Zi'ang Xu is a postgraduate student of the University of Science and Technology of China. His research mainly focuses on random network.

Qunqiang Feng is currently an Associate Professor at the University of Science and Technology of China (USTC). He received his Ph.D degree from USTC in 2006. His research mainly focuses on applied probability, random network models, and network data analysis.

References

[1] Holland P W, Leinhardt S. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, **1981**, *76* (373): 33–50.
[2] Barabási A-L, Pósfai M. Network Science. Cambridge, UK: Cambridge University Press, **2016**.
[3] De Paula A. Econometrics of network models. In: Advances in Economics and Econometrics: Eleventh World Congress, Cambridge, UK: Cambridge University Press (Econometric Society Monographs), **2017**: 268–323.
[4] Lewis K, Gonzalez M, Kaufman J. Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, **2012**, *109* (1): 68–72.
[5] Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nature Methods*, **2012**, *9* (5): 471–472.
[6] Newman M. Networks. Oxford, UK: Oxford University Press, **2018**.
[7] Holland P W, Laskey K B, Leinhardt S. Stochastic blockmodels: First steps. *Social Networks*, **1983**, *5* (2): 109–137.
[8] Fienberg S E, Meyer M M, Wasserman S S. Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association*, **1985**, *80* (389): 51–67.
[9] Aicher C, Jacobs A Z, Clauset A. Learning latent block structure in weighted networks. *Journal of Complex Networks*, **2015**, *3* (2): 221–248.
[10] Chatterjee S, Diaconis P, Sly A. Random graphs with a given degree sequence. *Annals of Applied Probability*, **2011**, *21* (4): 1400–1435.
[11] Yan T, Xu J. A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices. *Biometrika*, **2013**, *100* (2): 519–524.
[12] Chen M, Kato K, Leng C. Analysis of networks via the sparse β-model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **2021**, *83*: 887–910.
[13] Zhang Y, Wang Q, Zhang Y, et al. L-2 regularized maximum likelihood for β-model in large and sparse networks. arXiv: 2110.11856, **2021**.
[14] Yan T, Li Y, Xu J, et al. Wilks' theorems in the β-model. arXiv: 2211.10055, **2022**.
[15] Yan T, Leng C, Zhu J. Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *Annals of Statistics*, **2016**, *44* (1): 31–57.
[16] Yan T, Jiang B, Fienberg S E, et al. Statistical inference in a directed network model with covariates. *Journal of the American Statistical Association*, **2019**, *114* (526): 857–868.
[17] Stein S, Leng C. An annotated graph model with differential degree heterogeneity for directed networks. *Journal of Machine Learning Research*, **2023**, *24* (1): 5308–5376.
[18] Yan T. Directed networks with a differentially private bi-degree sequence. *Statistica Sinica*, **2021**, *31* (4): 2031–2050.
[19] Mukherjee R, Mukherjee S, Sen S. Detection thresholds for the β-model on sparse graphs. *Annals of Statistics*, **2018**, *46*: 1288–1317.
[20] Krivitsky P N, Handcock M S, Morris M. Adjusting for network size and composition effects in exponential-family random graph models. *Statistical Methodology*, **2011**, *8* (4): 319–339.
[21] Kolaczyk E D, Krivitsky P N. On the question of effective sample size in network modeling: An asymptotic inquiry. *Statistical Science*, **2015**, *30* (2): 184–198.
[22] Greenshtein E, Ritov Y. Persistence in high-dimensional linear predictor selection and the virtue of overparametrization. *Bernoulli*, **2004**, *10* (6): 971–988.
[23] Nishii R. Asymptotic properties of criteria for selection of variables in multiple regression. *Annals of Statistics*, **1984**, *12*: 758–765.
[24] Fan Y, Tang C Y. Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **2013**, *75* (3): 531–552.
[25] Adamic L A, Glance N. The political blogosphere and the 2004 US Election: Divided they blog. In: Proceedings of the 3rd International Workshop on Link Discovery. New York: ACM, **2005**: 36–43.

CSTR: 32290.14.JUSTC-2024-0139
DOI: 10.52396/JUSTC-2024-0139
JUSTC, **2025**, 55(X):