Electronic Journal of Statistics
Vol. 15 (2021) 3678–3707
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1872

Crawling subsampling for multivariate spatial autoregression model in large-scale networks*

Wei Hu

Center for Applied Statistics and School of Statistics, Renmin University of China

Danyang Huang

Center for Applied Statistics and School of Statistics, Renmin University of China
e-mail: dyhuang@ruc.edu.cn

Bingyi Jing

Department of Mathematics, The Hong Kong University of Science and Technology

Bo Zhang

Center for Applied Statistics and School of Statistics, Renmin University of China
e-mail: mabzhang@ruc.edu.cn

Abstract: In network data analysis, multivariate spatial autoregression (MSAR) models may be used to analyze the autocorrelation among multiple responses. With large-scale networks, the estimation for MSAR on the entire network is computationally expensive. In this case, the subsampling method could be adopted. This approach selects a sample of nodes and then uses the estimate based on the sample to approximate the estimate on the full data. However, traditional sampling methods cannot obtain unbiased parameter estimates. Considering the second-order friend information of sampled nodes, we propose the crawling subsampling (CS) method for a general framework. Thereafter, based on the sampled data only, we construct the least-squares objective function. Under certain conditions, the computational complexity of optimizing the objective function is linear with the sample size n_s. The identification condition for the parameters on the sampled network is theoretically provided. The sample size order requirement is provided, and the asymptotic properties of the least-squares estimators are investigated. The numerical results for the simulated and real data are presented to demonstrate the performance of the proposed CS method and least-squares estimator for the MSAR model.

*Danyang Huang's research is partially supported by the National Natural Science Foundation of China (NSFC, 12071477), building world-class universities (disciplines) of the Renmin University of China. Bingyi Jing's research is partially supported by the HK RGC Grants GRF16302020, GRF 16304419, GRF 16305616, and the State Key Program of National Natural Science Foundation of China grant NSFC11631003. Bo Zhang is supported by the National Natural Science Foundation of China (NSFC,71873137), building world-class universities (disciplines) of the Renmin University of China.

Keywords and phrases: Large-scale network, multivariate spatial autoregression model, subsampling, least-squares estimation.

Received November 2020.

Contents

1 Introduction 3679
2 Subsampling for MSAR 3681
2.1 Model and notations 3681
2.2 Crawling subsampling scheme 3682
2.3 LSE based on subsample 3683
2.4 Asymptotic property of LSE based on CS 3686
3 Numerical studies 3688
3.1 Simulation settings 3688
3.2 Performance measurements and simulation results 3689
3.3 Third-party restaurant consumer review website data analysis . 3691
4 Concluding remarks 3695
A Appendix A 3696
A.1 Notations and Verification of (2.4) 3696
A.2 Proof of Proposition 1 3698
B Appendix B 3698
B.1 Proof of Theorem 1 3698
B.2 Proof of Theorem 2 3699
B.3 Proof of Theorem 3 3700
References 3704

1. Introduction

In recent years, there has been a surge of interest in network data analysis [42, 22, 23, 5]. In network data, the behaviors of individuals are influenced by those with whom they have relationships [7]. Thus, the responses of different individuals are dependent on the network structure [33]. The spatial autoregression (SAR) model has become popular in network analysis to gauge the influence of the network structure on individual responses [1]. To analyze the autocorrelation among multiple responses, the SAR model has recently been extended to multivariate cases [43, 11], resulting in the multivariate spatial autoregression (MSAR) model. We take Twitter as an example. The tweets on different topics (for example, investment and savings) naturally constitute multivariate responses for each user. Consider the entire network with N nodes. The quasi-maximum likelihood estimation has been proposed to estimate the MSAR model. However, its computational complexity is $O(N^3)$ [3, 2]. Thus, Zhu et al. [46] proposed a novel least-squares estimation (LSE) for MSAR, by means of which the computational complexity could be sharply reduced.

However, in large-scale networks, estimation on the entire network data may be infeasible. As a result, the estimation of network models is typically conducted based on samples [30]. We consider the samples collected from the entire network by some well-designed sampling schemes. Obtaining a relatively accurate estimate based on a subsample then becomes an important issue [6].

Subsampling is a classical method proposed by [28]. This approach selects the sample nodes and related information, and subsequently uses the estimate based on the sampled data. This method has been re-popularized in large-scale data analysis [24, 12, 44] for approximating the estimate on the full data. When dealing with independent data, numerous efficient subsampling methods are available for linear regression [25, 39] and logistic regression [40, 37]. For dependent data, such as time-series or network data, the subsampling method is popular but challenging [17, 4, 9]. This is because the sampled network inevitably loses the connections between certain individuals [30]. As a result, the consistency of the estimator on the sampled data hardly holds [30]. Moreover, sampling methods are typically designed for a particular model [34, 18, 30, 4].

In particular, for SAR-related models, the estimated autocorrelation parameter tends to be negatively biased on the sampled data [7]. As a result, appropriate network sampling schemes and estimation methods should be discussed. Zhou et al. [45] proposed the paired maximum likelihood estimator method for sampled data to obtain a consistent estimation for the SAR model. However, as Taylor expansion is conducted, this method requires the autocorrelation coefficient to approach zero, which can hardly be satisfied in all cases. Furthermore, Huang et al. [20] proposed the LSE estimator for the classical univariate SAR model with no covariates and discussed a simple sampled case. However, the required sample size and corresponding proof were not discussed in [20]. This served as the inspiration for the current work. It is worth noting that the MSAR model is an extension of the classical SAR model [46]. The manner in which to conduct consistent estimation for MSAR models based on sampled data becomes a problem of interest.

In this study, focusing on the MSAR model in large-scale networks, we propose the crawling subsampling (CS) method for a general framework. The proposed approach is an extension of the sampling method in [20]. First, we design a general three-step CS scheme. The sample collected can retain the necessary network structure information for estimation. Based on the selected sample, we discuss the parameter identification issue, and use LSE to obtain the estimate for the parameters. The computational complexity of optimizing the objective function is discussed. Thereafter, we provide the sample size order requirement. Finally, we establish the asymptotic properties of the estimator under certain technical conditions. The performance of the proposed method is demonstrated by various simulation examples and data analysis of a third-party restaurant review website.

The remainder of this paper is organized as follows. Section 2 introduces the CS method and investigates the theoretical properties of the LSE based on the subsample. The simulation and real dataset studies are presented in Section 3.

The paper is concluded in Section 4. All the technical details are presented in Appendix A and Appendix B.

2. Subsampling for MSAR

2.1. Model and notations

For a large-scale network with N nodes, $a_{i_1 i_2} = 1$ is defined if there exists a relationship from node i_1 to i_2 ($i_1 \neq i_2$), and $a_{i_1 i_2} = 0$ otherwise, where $1 \leq i_1, i_2 \leq N$. Thus, the adjacency matrix is recorded as $A = (a_{i_1 i_2}) \in \mathbb{R}^{N \times N}$ and $a_{i_1 i_1} = 0$ is assumed.

Furthermore, the row-normalized adjacency matrix is defined as $W = (w_{i_1 i_2}) \in \mathbb{R}^{N \times N}$, where $w_{i_1 i_2} = d_{i_1}^{-1} a_{i_1 i_2}$ and $d_{i_1} = \sum_{i_2=1}^{N} a_{i_1 i_2}$ is the nodal out-degree of node i_1. For the nodal features, assume that each node i has p-dimensional continuous responses ($\mathbb{Y}_i^\top \in \mathbb{R}^p$) and q-dimensional exogenous covariates ($\mathbb{X}_i^\top \in \mathbb{R}^q$), where $1 \leq i \leq N$. Herein, we define n_s to be a small sample size, and n_x is defined as the number of unsampled nodes followed by the sampled nodes. Next, we define $\mathbb{Y}_g = (\mathbb{Y}_{ij}) \in \mathbb{R}^{n_s \times p}$ as the response matrix of the sampled nodes and $\mathbb{Y}_{g_0} \in \mathbb{R}^{(n_s+n_x) \times p}$ as the response matrix of the sampled nodes and their following nodes (i.e., followees). We define $\mathbb{X}_g = (\mathbb{X}_{ik}) \in \mathbb{R}^{n_s \times q}$ as the exogenous covariate matrix of the sampled nodes and $\mathbb{E}_g = (\varepsilon_{ij}) \in \mathbb{R}^{n_s \times p}$ as the noise matrix of the sampled nodes, wherein we assume that ε_i is identically and independently distributed with mean $\mathbf{0}_p \in \mathbb{R}^p$ and $\mathrm{cov}(\varepsilon_i) = \Sigma_e \in \mathbb{R}^{p \times p}$.

Then, we define the MSAR model on a sampled network. We denote the parameters $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_{j'j}) \in \mathbb{R}^{p \times p}$ and $B = (b_{kj}) \in \mathbb{R}^{q \times p}$. Thus, the MSAR model on the sampled network can be expressed as

$$\mathbb{Y}_g = W_{g_0} \mathbb{Y}_{g_0} \boldsymbol{\alpha} + \mathbb{X}_g B + \mathbb{E}_g, \tag{2.1}$$

where $W_{g_0} \in \mathbb{R}^{n_s \times (n_s+n_x)}$ in (2.1) only retains the rows of the sampled nodes in W; i.e., W_{g_0} records the edges between the sampled nodes and their followees. Following [11] and [46], $\boldsymbol{\alpha}_{jj}$ is referred to as the *intra-activity effect* or *endogenous effect*, which measures the impact from the same type of response variable of the connected neighbors; $\boldsymbol{\alpha}_{j'j} (j' \neq j)$ is known as the *extra-activity effect* or *cross-activity peer effect*, which measures the impact from the other response types; $b_{kj} (1 \leq k \leq q)$ is referred to as the *own effect*, which reflects the impact from the exogenous covariates. The detailed presentation of the MSAR model and related notations could also be found in [46]. The impact relationship among the responses and exogenous covariates in (2.1) can be interpreted as illustrated in Figure 1.

Subsequently, we discuss the parameter space. Let $|\lambda_{\max}(\boldsymbol{\alpha})|$ be the maximum absolute eigenvalue of the matrix $\boldsymbol{\alpha}$. Assuming that $|\lambda_{\max}(\boldsymbol{\alpha})| < 1$, the matrix $(I_{Np} - \boldsymbol{\alpha}^\top \otimes W)$ is invertible (as proven by Lemma 1 of [46]), where I_{Np} is the $(Np \times Np)$-dimensional identity matrix and $\otimes$ is the Kronecker product. In this study, based on the sample data, we focus on the estimation of $\boldsymbol{\alpha} \in \mathbb{R}^{p \times p}$ and $\boldsymbol{\beta} = \mathrm{vec}(B) \in \mathbb{R}^{pq}$ on this parameter space $\{|\lambda_{\max}(\boldsymbol{\alpha})| < 1\}$.

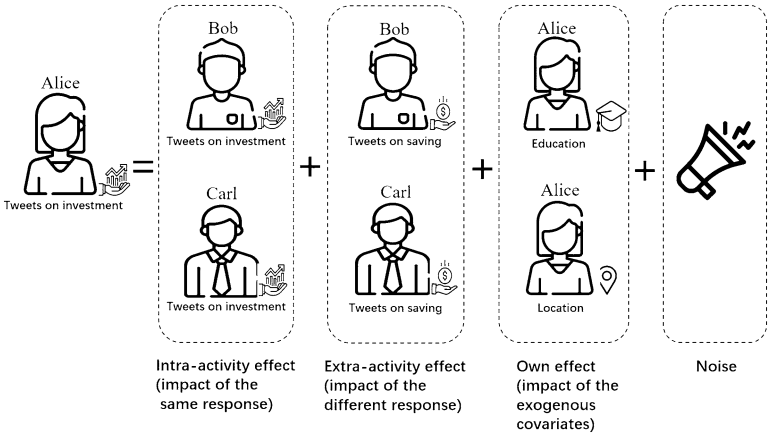

FIG 1. *Illustration of MSAR model. Assume that node i (Alice) has two different connected nodes (Bob and Carl), and there are two responses (tweets on investment and tweets on saving). The influential variables of Alice's tweets on investment can be divided into four parts, corresponding to the four terms on the right side of* (2.1)*. The first part consists of the tweets on investment of Alice's friends (Bob and Carl), which corresponds to the intra-activity effect; the second part consists of the tweets on saving of Alice's friends, which corresponds to the extra-activity effect; the third part consists of Alice's educational background and location, which corresponds to the own effect; and the fourth part is the noise term.*

2.2. Crawling subsampling scheme

In this section, we discuss the detailed procedures of the CS scheme. In the model of (2.1), only the nodes followed by the sampled nodes are considered (i.e., only part of the first-order friends). However, to calculate the least-squares-type objective function, which is defined as follows (refer to (2.2)-(2.4) for further details), the estimation method involves the first- and second-order friends of the sampled nodes. Thus, the proposed CS scheme aims to retain the network structure and the corresponding nodal features necessary for estimation.

Next, we focus on the sampling scheme. When using CS, the sampled data contain nodal features and double-layer supplementary information of the sampled nodes. We define g_i as a sampling indicator, with $g_i = 1$ if node i is sampled and $g_i = 0$ otherwise. We define $\mathcal{G}_y = \{i : g_i = 1, 1 \leq i \leq N\}$ as the collection of sampled nodes. The sampling scheme can be implemented in the following three steps. These steps are illustrated in Figure 2, and the related notations are presented in the following part.

STEP1: Obtain the original node set $\mathcal{G}_y$ via a convenient sampling method, such as simple random sampling. Denote the nodes in $\mathcal{G}_y$ as **sampled nodes**.

STEP2: Collect the followers and followees of sampled nodes, which are defined as $\mathcal{G}_{x_1}$ and $\mathcal{G}_{x_2}$, respectively. In other words, collect all nodes j that are outside $\mathcal{G}_y$ but are directly connected to some nodes in $\mathcal{G}_y$. Furthermore, collect the corresponding nodal features. The two node sets are expressed as $\mathcal{G}_{x_1} = \{j : a_{ji} = 1, j \notin \mathcal{G}_y \text{ and } i \in \mathcal{G}_y\}$ and $\mathcal{G}_{x_2} = \{j : a_{ij} = 1, j \notin \mathcal{G}_y$

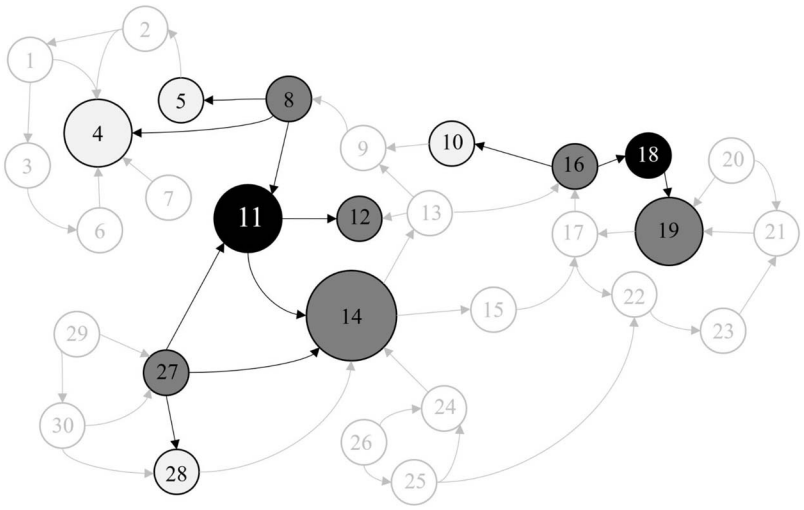

FIG 2. *Example of three steps of collecting nodes. The black nodes {11,18} are collected by* **STEP1***; the dark gray nodes {8,12,14,16,19,27} are collected by* **STEP2***; and the light gray nodes {4,5,10,28} are collected by* **STEP3**.

and $i \in \mathcal{G}_y\}$, respectively. Consider the information concerning the nodes in $\mathcal{G}_{x_1} \bigcup \mathcal{G}_{x_2}$ as the first layer of supplementary information.

STEP3: Collect the indirectly connected nodes of sampled nodes, defined as $\mathcal{G}_{x_3}$. In other words, collect all the nodes that are outside $\mathcal{G}_y \bigcup \mathcal{G}_{x_1} \bigcup \mathcal{G}_{x_2}$ but are indirectly connected to some nodes in $\mathcal{G}_y$. Further, collect the corresponding nodal features. By searching for all the nodes followed by nodes $j \in \mathcal{G}_{x_1}$, we obtain the indirectly connected node set, expressed as $\mathcal{G}_{x_3} = \{k : a_{jk} = 1, j \in \mathcal{G}_{x_1}$, and $k \notin \mathcal{G}_y \bigcup \mathcal{G}_{x_1} \bigcup \mathcal{G}_{x_2}\}$. Consider the information concerning the nodes in $\mathcal{G}_{x_3}$ as the second layer of supplementary information.

Denote $\mathcal{G}_x = \mathcal{G}_{x_1} \bigcup \mathcal{G}_{x_2} \bigcup \mathcal{G}_{x_3}$ as the supplementary node set, which records the supplementary nodes outside the sampled nodes. Finally, based on the sampled data, we can construct the least-squares objective function in the next section.

2.3. LSE based on subsample

In this section, based on the sampled data, we construct the least-squares objective function. Moreover, the computational complexity for optimizing the objective function and the identification issue for the parameters are discussed.

First, we define the best linear predictor (BLP) of the response $\mathbb{Y}_{i_1 j_1}$ only on the sampled data collected above. For any sampled node $i_1 \in \mathcal{G}_y$ and $1 \leq j_1 \leq p$, we define $\mathbb{Y}_{-(i_1,j_1)} = \{\mathbb{Y}_{ij} : (i,j) \neq (i_1,j_1), i \in \mathcal{G}_y \bigcup \mathcal{G}_x, 1 \leq j \leq p\}$. Based on the sampled data, which consist of the nodes in $\mathcal{G}_y \bigcup \mathcal{G}_x$, necessary connections, and corresponding nodal features, we construct the BLP of $\mathbb{Y}_{i_1 j_1}$, given $\mathbb{Y}_{-(i_1,j_1)}$

[31]. The BLP, defined as $F\left\{\mathbb{Y}_{i_1j_1}|\mathbb{Y}_{-(i_1,j_1)}\right\}$, can be obtained as follows:

$$F\left\{\mathbb{Y}_{i_1j_1}|\mathbb{Y}_{-(i_1,j_1)}\right\} = \boldsymbol{\mu}_{i_1j_1} + \sum_{(i_2,j_2)\in\mathbb{S}_{xyp}\setminus(i_i,j_1)} r_{i_1j_1i_2j_2}\left(\mathbb{Y}_{i_2j_2} - \boldsymbol{\mu}_{i_2j_2}\right), \tag{2.2}$$

where

$$r_{i_1j_1i_2j_2} = \frac{\boldsymbol{\alpha}_{j_1\cdot}\Omega_{e,\cdot j_2}w_{i_2i_1} + \Omega_{e,j_2\cdot}\boldsymbol{\alpha}_{\cdot j_1}w_{i_1i_2} - \left(\boldsymbol{\alpha}_{j_1\cdot}\Omega_e\boldsymbol{\alpha}_{j_2\cdot}^{\top}\right)\left(\sum_{i\in\mathcal{G}_y\bigcup\mathcal{G}_x} w_{ii_1}w_{ii_2}\right)}{\omega_{e,j_1j_1} + \left(\boldsymbol{\alpha}_{j_1\cdot}\Omega_e\boldsymbol{\alpha}_{j_1\cdot}^{\top}\right)\left(\sum_{i\in\mathcal{G}_y\bigcup\mathcal{G}_x} w_{ii_1}^2\right)}, \tag{2.3}$$

in which $\mathbb{S}_{xyp} = ((\mathcal{G}_x\cup\mathcal{G}_y)\times\{1,\ldots,p\})$, $\boldsymbol{\mu}_{i_1j_1} = E(\mathbb{Y}_{i_1j_1})$ and $\Omega_e = \Sigma_e^{-1}$. Moreover, $\Omega_{e,\cdot j_2}, \Omega_{e,j_2\cdot}$ represent the j_2th column and j_2th row of the matrix Ω_e, respectively. Similarly, $\boldsymbol{\alpha}_{j_1\cdot}$ and $\boldsymbol{\alpha}_{\cdot j_2}$ represent the j_1th row and j_2th column of the matrix $\boldsymbol{\alpha}$, respectively. The verification of (2.2) is presented in Appendix A.1. Three types of unsampled nodes are used for $r_{i_1j_1i_2j_2}\neq 0$ in (2.3), as illustrated in Figure 3. It should be noted that the calculation of any $\boldsymbol{\mu}_{i_1j_1}$ in (2.2) still involves the nodal exogenous variables of the entire network. However, in the proposed calculation of the objective function (2.4), the mean term $\boldsymbol{\mu}_{i_1j_1}$ will be eliminated. We can calculate and optimize the objective function based only on the sampled data.

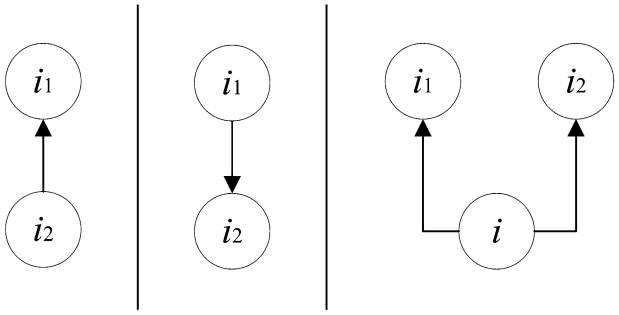

FIG 3. *Three types of unsampled nodes i_2 involved in (2.3). In this case, i_1 represents the sampled nodes ($\mathcal{G}_y$) and i_2 represents the supplementary nodes ($\mathcal{G}_x$). The three types are **(1)** nodes i_1 followed by i_2 ($w_{i_2i_1}\neq 0$), representing $i_2\in\mathcal{G}_{x_1}$ in the left panel; **(2)** nodes i_2 followed by i_1 ($w_{i_1i_2}\neq 0$), representing $i_2\in\mathcal{G}_{x_2}$ in the middle panel; and **(3)** i_1 and i_2 indirectly followed by a third node i ($\sum_{i\in\mathcal{G}_y\bigcup\mathcal{G}_x} w_{ii_1}w_{ii_2}\neq 0$), representing $i_2\in\mathcal{G}_{x_3}$ in the right panel. As a result, the three types of unsampled nodes involved in (2.3) correspond exactly to the supplementary nodes $\mathcal{G}_{x_1},\mathcal{G}_{x_2},\mathcal{G}_{x_3}$ collected by **STEP 2** and **STEP 3** in Figure 2.*

Next, we discuss the objective function based only on the sampled data. For $1\leq i\leq N$ and $1\leq j\leq p$, we define $\boldsymbol{\mu}_c = \text{vec}\{F(\mathbb{Y}_{ij}|\mathbb{Y}_{-(i,j)})\}\in\mathbb{R}^{Np}$, $\boldsymbol{\theta} = \left(\text{vec}(\boldsymbol{\alpha})^{\top},\boldsymbol{\beta}^{\top}\right)^{\top}\in\mathbb{R}^{p^2+pq}$, the sampling matrix $g = \text{diag}(g_1,\cdots,g_N)\in\mathbb{R}^{N\times N}$, and $G = I_p\otimes g$. Then, for the sampled nodes, the least-squares objective function can be written as

$$L(\boldsymbol{\theta}) = \sum_{i\in\mathcal{G}_y,1\leq j\leq p}\left\{\mathbb{Y}_{ij} - F\left(\mathbb{Y}_{ij}|\mathbb{Y}_{-(i,j)}\right)\right\}^2 = ||G(\mathcal{Y}-\boldsymbol{\mu}_c)||^2. \tag{2.4}$$

The verification of (2.4) is provided in Appendix A.1. It should be emphasized that the G matrix in (2.4) is simply presented for convenience of the proof. We denote $H = G(\mathcal{Y} - \boldsymbol{\mu}_c) = mGS^\top(\Omega_e \otimes I_N)(S\mathcal{Y} - \widetilde{\mathbb{X}}\boldsymbol{\beta}) \in \mathbb{R}^{Np}$, where $S = I_{Np} - \boldsymbol{\alpha}^\top \otimes W$ and $\widetilde{\mathbb{X}} = I_p \otimes \mathbb{X}$. Subsequently, we have $L = H^\top H$. Only the sampled data are involved in the actual calculation of H. Furthermore, the values of the unsampled nodes in H are all equal to 0. We focus on the objective function $L(\boldsymbol{\theta})$ in the following parts.

Theorem 1. *We assume that (2.4) can be optimized by the Newton–Raphson algorithm. Further, we assume that the nodal in-degree $\sum_j a_{ji}$ is bounded by a finite constant $d_{\max}$ (i.e., $\max_i \sum_j a_{ji} \leq d_{\max}$). Thus, the computational complexity of optimizing (i.e., minimizing) the objective function (2.4) is $O(n_s)$. Moreover, if the maximum value of the nodal in-degree is $O(\log n_s)$ [36], the computational complexity is $O(n_s \log n_s)$.*

The proof of Theorem 1 is presented in Appendix B.1. According to Theorem 1, in the ideal case, the computational complexity of optimizing the objective function is linear in the sample size n_s. Even if the condition is not satisfied, the computational complexity is relatively low because the number of nodes involved is significantly smaller than that of the entire network. It is remarkable that the convergence of Newton–Raphson algorithm is guaranteed by [14] and [29], which requires the objective function to be twice differentiable and the second order derivative to be nonsingular. As a result, under Condition (C1) and the assumption in Theorem 2, which we will introduce later, the algorithm converges.

Subsequently, based on the sampled data, we investigate the identification of the parameters. The true parameters of $\boldsymbol{\alpha}, \boldsymbol{\beta}$ only in this identification part are defined as $\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0$, respectively. We define $S_g = I_{n_s p} - \boldsymbol{\alpha}^\top \otimes W_g \in \mathbb{R}^{n_s p \times n_s p}$, $S_{g0} = I_{n_s p} - \boldsymbol{\alpha}_0^\top \otimes W_g$. Furthermore, we define $V_i (i = 1, \cdots, p)$ to be a $1 \times p$ row vector, in which the ith entry is 1 and the others are 0. We define $M_p = (V_p \otimes I_{n_s}) S_{g0}^{-1} \{I_p \otimes (W_g \mathbb{X}_g)\} \boldsymbol{\beta}_0$, $M_g = (M_1, \cdots, M_p) \in \mathbb{R}^{n_s \times p}$, and $\mathbb{X}_g^* = (M_g, \mathbb{X}_g) \in \mathbb{R}^{n_s \times (p+q)}$. Thereafter, we make the following assumption.

(C1) (Identification condition): Assume that the limit $\lim_{n_s \to \infty} n_s^{-1}(\mathbb{X}_g^{*\top} \mathbb{X}_g^*)$ exists and is nonsingular.

Note that condition (C1) corroborates with the identification condition in the existing literature [43, 11]. The identification on the sampled data can be proven under condition (C1).

Theorem 2. *Assume that there exists $\delta > 0$ such that*

$$\min_{\{|\lambda_{\max}(\boldsymbol{\alpha})| \leq 1-\delta\}} \{\lambda_{\min}(S_g S_g^\top)\} \geq c_s, \tag{2.5}$$

where c_s is a positive constant. Then, under condition (C1), in the parameter space $\{|\lambda_{\max}(\boldsymbol{\alpha})| \leq 1 - \delta\}$, the parameters $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ can be identified in the sampled network.

The proof of Theorem 2 is presented in Appendix B.2. According to Theorem 2, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ can be identified in the sampled network.

2.4. Asymptotic property of LSE based on CS

In this section, for the least-squares estimators on the sampled data, we discuss the required sample size. Then, we investigate the asymptotic distribution under the identifiability assumption. First, recall that L is the least-squares objective function, and the sample size is n_s. To give the form of the asymptotic distribution, we define $\mathcal{E} = \text{vec}(\mathbb{E}) \in \mathbb{R}^{Np}$, $\widetilde{\mathcal{E}} = (\Sigma_e^{-1/2} \otimes I_N)\mathcal{E}$ and $\widetilde{S} = (\Sigma_e^{-1} \otimes I_N)S$; we have the objective function $L = H^\top H$. Moreover, we define $H = H^* \widetilde{\mathcal{E}}$, where $H^* = mG\widetilde{S}^\top \left(\Sigma_e^{1/2} \otimes I_N\right)$. We denote $\boldsymbol{\alpha}_{j_1 j_2}$ as the j_1th row, j_2th column element of the matrix $\boldsymbol{\alpha}$. With respect to $\boldsymbol{\alpha}_{j_1 j_2}$ and the vector $\boldsymbol{\beta}$, we denote the derivatives $H^{\boldsymbol{\alpha}}_{j_1 j_2} = M^*_{j_1 j_2}\widetilde{\mathcal{E}} + V^*_{j_1 j_2}$, $H_{\boldsymbol{\beta}} = -mG\widetilde{S}^\top \widetilde{\mathbb{X}}$. The other related matrices and vectors are expressed as $M_{\boldsymbol{\beta}} = 2H^\top_{\boldsymbol{\beta}} H^*$, $V^*_{j_1 j_2} = mG\widetilde{S}^\top S^{\boldsymbol{\alpha}}_{j_1 j_2} S^{-1}(\widetilde{\mathbb{X}}\boldsymbol{\beta})$, $M^*_{j_1 j_2} = \{m^{\boldsymbol{\alpha}}_{j_1 j_2} G\widetilde{S}^\top + mG\widetilde{S}^{\boldsymbol{\alpha}\top}_{j_1 j_2} + mG\widetilde{S}^\top S^{\boldsymbol{\alpha}}_{j_1 j_2} S^{-1}\}(\Sigma_e^{1/2} \otimes I_N)$, $M_{j_1 j_2} = H^{*\top} M^*_{j_1 j_2}$, and $V_{j_1 j_2} = H^{*\top} V^*_{j_1 j_2}$.

To establish the asymptotic property based on the crawling subsample, we focus on the following technical conditions on large-scale network data.

(C2) (Network structure)

(C2.1) (Connectivity of the entire network) Let the set of all nodes $\{1, \cdots, N\}$ be the state space of a Markov chain, with the transition probability given by W. The Markov chain is assumed to be irreducible and aperiodic. Furthermore, $\pi = (\pi_i)^\top \in \mathbb{R}^N$ is defined as the stationary distribution vector of the Markov chain (i.e., $\pi_i \geq 0, \sum_{i=1}^N \pi_i = 1$, and $W^\top \pi = \pi$). Then, assume that $\sum_{i=1}^N \pi_i^2 = O\left(N^{-1/2-\delta}\right)$, where $0 < \delta \leq 1/2$ is a positive constant.

(C2.2) (Uniformity of the sampled network) Assume that $|\lambda_{\max}\{g(W+W^\top)g\}| = O(\log n_s)$.

Conditions (C2.1) and (C2.2) are two separate assumptions regarding the overall network structure. Condition (C2.1) is similar to those in other studies [20, 46]. Condition (C2.1) requires certain connectivity for the network structure. If the network is fully connected following a limited number of steps, the irreducibility and aperiodicity can be satisfied simultaneously. This condition is satisfied if the well-known theory of six-degree separation [26] holds. Moreover, δ represents the global influence, typically for a large-scale network, where there is no dominant node (i.e., $\delta \to 1/2$). Condition (C2.2) imposes a certain uniformity assumption only on the sampled network (i.e., Wg, $gW^\top$). (C2.2) requires $\lambda_{\max}\{g(W + W^\top)g\}$ to be slowly diverging with a rate of $O(\log n_s)$. Naturally, condition (C2.2) can be considered as the generalization of the uniformity assumption of the entire network [20, 46, 21].

Next, we discuss the sample size required for the subsampling method.

Proposition 1. *Under condition (C2) and (C4), there exists* $\epsilon > 0$*, and we assume that the sample size in* $\mathcal{G}_y$ *satisfies*

$$n_s = O(N^{1-2\delta+\epsilon}).$$

Thus, we have $n_s^{-2}\operatorname{tr}(M_{k_1k_2}^{*\top}M_{j_1j_2}^{*}M_{j_1j_2}^{*\top}M_{k_1k_2}^{*}) \to 0$ *and* $n_s^{-2}V_{j_1j_2}^{*\top}M_{k_1k_2}^{*}M_{k_1k_2}^{*\top}V_{j_1j_2}^{*} \to 0$.

The proof of Proposition 1 is presented in Appendix A.2. Proposition 1 indicates that the sample size order is related to the nodes with global influence. The condition of the sample size n_s is necessary in Theorem 3. The required sample size n_s will be smaller for networks with smaller global influence. Thus, we state the condition as follows:

(C3) (Order of sample size) Assume that there exists $\epsilon > 0$ such that $n_s = O(N^{1-2\delta+\epsilon})$.

(C4) (Covariates): For arbitrary $\boldsymbol{\beta}_r \in \mathbb{R}^{pq}$ and $R \in \mathbb{R}^{Np\times Np}$, define $R_g = GRG$ and assume that $|N^{-1}\{(I_p \otimes \mathbb{X})\boldsymbol{\beta}_r\}^\top R\{(I_p \otimes \mathbb{X})\boldsymbol{\beta}_r\}| \le N^{-1}c_{\boldsymbol{\beta}}\operatorname{tr}(R_g)$ as $N \to \infty$, where $c_{\boldsymbol{\beta}}$ is a positive constant only related to $\boldsymbol{\beta}_r$.

(C5) (Noise term) Let $\widetilde{\varepsilon}_i = \left(\Sigma_e^{1/2\top}\right)^{-1}\varepsilon_i = (\widetilde{\varepsilon}_{i1}, \cdots, \widetilde{\varepsilon}_{ip})^\top \in \mathbb{R}^p$. Assume that $E\left(\widetilde{\varepsilon}_{ij}^4\right) = \kappa_4$ and $E\left(\widetilde{\varepsilon}_{ij_1}\widetilde{\varepsilon}_{ij_2}\widetilde{\varepsilon}_{ij_3}\right) = 0$ for $1 \le i \le N$ and $1 \le j_1, j_2, j_3 \le p$, where κ_4 is a finite constant.

(C6) (Law of large numbers) For $1 \le i_1, i_2 \le p^2$, assume that the following six limits exist:

$$\begin{aligned}
\Sigma_{1\boldsymbol{\alpha}}^{(i_1,i_2)} &= \lim_{n_s\to\infty} n_s^{-1}\left\{4\operatorname{tr}\left(M_{j_1j_2}M_{k_1k_2}^\top\right) + 4\operatorname{tr}\left(M_{j_1j_2}M_{k_1k_2}\right)\right.\\
&\quad \left. + 4\left(\kappa_4 - 3\right)\operatorname{tr}\left\{\operatorname{diag}\left(M_{j_1j_2}\right)\operatorname{diag}\left(M_{k_1k_2}\right)\right\} + 4V_{j_1j_2}^\top V_{k_1k_2}\right\},\\
\Sigma_{1\boldsymbol{\alpha}\boldsymbol{\beta}}^{(i_1,\cdot)} &= 2\lim_{n_s\to\infty} n_s^{-1}\left(V_{j_1j_2}^\top M_{\boldsymbol{\beta}}^\top\right), \quad \Sigma_{1\boldsymbol{\beta}} = \lim_{n_s\to\infty} n_s^{-1}\left(M_{\boldsymbol{\beta}}M_{\boldsymbol{\beta}}^\top\right)\\
\Sigma_{2\boldsymbol{\alpha}}^{(i_1,i_2)} &= 2\lim_{n_s\to\infty} n_s^{-1}\left\{\operatorname{tr}\left(M_{j_1j_2}^{*\top}M_{k_1k_2}^{*}\right) + V_{j_1j_2}^{*\top}V_{k_1k_2}^{*}\right\},\\
\Sigma_{2\boldsymbol{\alpha}\boldsymbol{\beta}}^{(i_1,\cdot)} &= 2\lim_{n_s\to\infty} n_s^{-1}V_{j_1j_2}^{*\top}H_{\boldsymbol{\beta}}, \quad \Sigma_{2\boldsymbol{\beta}} = 2\lim_{n_s\to\infty} n_s^{-1}H_{\boldsymbol{\beta}}^\top H_{\boldsymbol{\beta}}.
\end{aligned} \tag{2.6}$$

In the overall network data, conditions (C4) and (C5) set regularity conditions on the exogenous covariates and noise terms, respectively. In (C6), note that $\Sigma_{1\boldsymbol{\alpha}}^{(i_1,i_2)}$ represents the i_1th row, i_2th column element in $\Sigma_{1\boldsymbol{\alpha}}^{(i_1,i_2)}$, whereas $\Sigma_{1\boldsymbol{\alpha}\boldsymbol{\beta}}^{(i_1,\cdot)}$ represents the i_1 column of $\Sigma_{1\boldsymbol{\alpha}\boldsymbol{\beta}}$, $\Sigma_{1\boldsymbol{\beta}} \in \mathbb{R}^{(pq)\times(pq)}$. Similarly, $\Sigma_{2\boldsymbol{\alpha}}^{(i_1,i_2)}$ represents the i_1th row, i_2th column element in $\Sigma_{2\boldsymbol{\alpha}}^{(i_1,i_2)}$, whereas $\Sigma_{2\boldsymbol{\alpha}\boldsymbol{\beta}}$ represents the i_1 column of $\Sigma_{1\boldsymbol{\alpha}\boldsymbol{\beta}}$, $\Sigma_{1\boldsymbol{\beta}} \in \mathbb{R}^{(pq)\times(pq)}$.

Theorem 3. *Assuming conditions (C1) to (C6), the true parameters are rewritten as* $\boldsymbol{\theta} = \left(\operatorname{vec}(\boldsymbol{\alpha})^\top, \boldsymbol{\beta}^\top\right)^\top \in \mathbb{R}^{p^2+pq}$*, and* $\hat{\boldsymbol{\theta}}$ *represents the least-squares estimator. For* $1 \le i_1, i_2 \le p^2$*, let* $\Sigma_{1\boldsymbol{\alpha}} = (\Sigma_{1\boldsymbol{\alpha}}^{(i_1,i_2)}) \in \mathbb{R}^{p^2\times p^2}$*,* $\Sigma_{1\boldsymbol{\alpha}\boldsymbol{\beta}} = \left(\Sigma_{1\boldsymbol{\alpha}\boldsymbol{\beta}}^{(i_1,\cdot)}\right) \in \mathbb{R}^{p^2\times(pq)}$*,* $\Sigma_{1\boldsymbol{\beta}} \in \mathbb{R}^{(pq)\times(pq)}$*,* $\Sigma_{2\boldsymbol{\alpha}} = (\Sigma_{2\boldsymbol{\alpha}}^{(i_1,i_2)}) \in \mathbb{R}^{p^2\times p^2}$*,* $\Sigma_{2\boldsymbol{\alpha}\boldsymbol{\beta}} = \left(\Sigma_{2\boldsymbol{\alpha}\boldsymbol{\beta}}^{(i_1,\cdot)}\right) \in$

$\mathbb{R}^{p^2\times(pq)}$, $\Sigma_{2\boldsymbol{\beta}} \in \mathbb{R}^{(pq)\times(pq)}$. *Then, in the parameter space* $\{|\lambda_{\max}(\boldsymbol{\alpha})| < 1\}$, *as* $n_s \to \infty$, *we have*

$$\sqrt{n_s}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \to_d N\left(\mathbf{0}_{p^2+pq}, \left(\Sigma_{2s}\right)^{-1}\Sigma_{1s}\left(\Sigma_{2s}\right)^{-1}\right), \tag{2.7}$$

where

$$\Sigma_{1s} = \begin{pmatrix} \Sigma_{1\boldsymbol{\alpha}} & \Sigma_{1\boldsymbol{\alpha\beta}} \\ \Sigma_{1\boldsymbol{\alpha\beta}}^{\top} & \Sigma_{1\boldsymbol{\beta}} \end{pmatrix}, \quad \Sigma_{2s} = \begin{pmatrix} \Sigma_{2\boldsymbol{\alpha}} & \Sigma_{2\boldsymbol{\alpha\beta}} \\ \Sigma_{2\boldsymbol{\alpha\beta}}^{\top} & \Sigma_{2\boldsymbol{\beta}} \end{pmatrix}. \tag{2.8}$$

The detailed expression of (2.8) is provided in (2.6). The proof of Theorem 3 is presented in Appendix B.3. According to Theorem 3, $\hat{\boldsymbol{\theta}}$ is $\sqrt{n_s}$-consistent and asymptotically normally distributed on the sampled data. The estimate is obtained by optimizing (2.4) using the Newton–Raphson algorithm.

3. Numerical studies

3.1. Simulation settings

To demonstrate the performance of the parameter estimation on finite sampled data, we compare four sampling methods on three different networks. The first three classical sampling methods are simple random sampling (SRS), snowball sampling (SN), and Metropolis–Hastings random walk (MHRW). The final sampling method is CS. We consider the three traditional sampling methods and the generation mechanism of the simulated data. We define p_i to represent the sampling probability of node i.

(1) SRS: A simple random sample is drawn by selecting n_s nodes from the total N nodes with equal probability $p_i = 1/N$. Thus, SRS roughly ignores the network topology.

(2) SN: SN [16] is a traditional network sampling method for capturing the k-order friends of the initial seed node. The SN sample can be obtained as follows: **(i)** First, randomly select a node i and collect all the nodes directly connected to i. Denote the sampled node set as $\mathcal{G}_y^1 = \{i\}\bigcup\{j : 1 \le j \le N, a_{ij} = 1 \text{ or } a_{ji} = 1\}$. **(ii)** Randomly select a node i' in $\mathcal{G}_y^t$, where t represents the tth step in the sampling. Thereafter, collect the nodes connected to i'. Denote the sampled node set as $\mathcal{G}_y^{t+1} = \mathcal{G}_y^t\bigcup\{k : 1 \le k \le N, a_{i'k} = 1 \text{ or } a_{ki'} = 1\}$. **(iii)** Repeat **(ii)** until the sample size reaches n_s.

(3) MHRW: MHRW, which is based on *random walk* sampling, is another network sampling method for collecting nodes adaptively by means of link tracing [32, 41]. The MHRW sample can be obtained as follows: **(i)** First, randomly select a node i, with $\mathcal{G}_y^1 = \{i\}$. **(ii)** Randomly select a node i' in $\mathcal{G}_y^t$ from the connected nodes (i.e., j) of node i'. MHRW adaptively selects a node with unequal probability p_j, where $p_j = d_i^{-1}\min(1, d_i/d_j)$ and d_i is the out-degree of node i. Denote $\mathcal{G}_y^{t+1} = \mathcal{G}_y^t\bigcup\{j\}$. **(iii)** Repeat **(ii)** until the sample size reaches n_s.

We consider the following three simulation examples for the network structure. The three examples have different generation mechanisms for the network structure A. **(1)** The first example is the *dyad independence network* [19],

where a dyad is defined as $\mathcal{A}_{ij} = (a_{ij}, a_{ji})$ for the ith and jth nodes. We set $P(\mathcal{A}_{ij} = (1,1)) = 20N^{-1}$ and $P(\mathcal{A}_{ij} = (1,0)) = P(\mathcal{A}_{ij} = (0,1)) = 0.5N^{-0.8}$. **(2)** The second example is the *stochastic block model* [38, 27]. We define $K = 100$ as the total number of blocks. Then, we set $P(a_{ij} = 1) = 0.9N^{-1}$ if the ith and jth nodes are in the same block, and $P(a_{ij} = 1) = 0.3N^{-1}$ otherwise. **(3)** The third example is the *power-law distribution network* [10]. The in-degree of node i is denoted as d_i (i.e., $d_i = \sum_{j=1}^{N} a_{ji}$). Moreover, d_i follows the discrete power-law distribution (i.e., $P(d_i = k) = c_n k^{-c_e}$), where c_n is a normalizing constant. We set the exponent parameter $c_e = 2.5$.

Finally, we consider the generation mechanism of the noise matrix $\mathbb{E}$ and the exogenous covariate matrix $\mathbb{X}$. We set $p = 2$, $q = 2$, $\boldsymbol{\alpha} = (0.3, 0; -0.2, 0.1)$, and $\boldsymbol{\beta} = (-0.5, 1.3, 1, 0.3)^\top$. For the noise matrix, $\mathbb{E}_{i\cdot}$ (i.e., $i = 1, ..., N$) is generated independently: **(1)** multivariate normal distribution with mean $(0,0)^\top$ and covariance $\Sigma_e = (0.4, 0.1; 0.1, 0.6) \in \mathbb{R}^{2\times 2}$; and **(2)** t-distribution with the same mean and covariance as (1) and a degree equal to 5. For the exogenous covariates of node i, $\mathbb{X}_{i\cdot}$ is generated from a multivariate normal distribution with mean $(0,0)^\top$ and covariance $\Sigma_x = (0.5^{|j_1 - j_2|}) \in \mathbb{R}^{2\times 2}$.

3.2. Performance measurements and simulation results

We compare the different sampling methods in terms of the bias and efficiency of the estimates. Consider the overall network size (i.e., $N = 2000$) and different sample sizes (n_s=200, 500). In each simulation example, we repeat the experiment $R = 500$ times to compare the performance. For the true parameter $\boldsymbol{\alpha}_{jk} (1 \le j, k \le p)$, the corresponding estimator is recorded as $\hat{\boldsymbol{\alpha}}_{jk}^{(r)}$, where r represents the rth replication (i.e., $r = 1, ..., R$). For the parameter $\boldsymbol{\alpha}_{jk}$, the average bias is constructed as $R^{-1}\sum_{r=1}^{R}(\hat{\boldsymbol{\alpha}}_{jk}^{(r)} - \boldsymbol{\alpha}_{jk})$. The standard deviation is constructed as $[R^{-1}\{\hat{\boldsymbol{\alpha}}_{jk}^{(r)} - R^{-1}\sum_{r=1}^{R}(\hat{\boldsymbol{\alpha}}_{jk}^{(r)})\}^2]^{1/2}$. The root mean square error (RMSE) is calculated by $\{R^{-1}\sum_{r=1}^{R}(\hat{\boldsymbol{\alpha}}_{jk}^{(r)} - \boldsymbol{\alpha}_{jk})^2\}^{1/2}$. Furthermore, $\widehat{\mathrm{SE}}_{jk}^{(r)}$ is denoted as the $\{(j-1)p + k\}$th diagonal element of the asymptotic covariance matrix in (2.7). Then, the empirical 95% confidence interval for $\boldsymbol{\alpha}_{jk}$ is constructed as $\mathrm{CI}_{jk}^{(r)} = (\hat{\boldsymbol{\alpha}}_{jk}^{(r)} - z_{0.975}n_s^{-1}\widehat{\mathrm{SE}}_{jk}^{(r)}, \hat{\boldsymbol{\alpha}}_{jk}^{(r)} + z_{0.975}n_s^{-1}\widehat{\mathrm{SE}}_{jk}^{(r)})$, where z_α is the αth lower quantile of a standard normal distribution. As a result, using the empirical 95% confidence interval, the empirical coverage probability is determined as $\mathrm{CP}_{jk} = R^{-1}\sum_{r=1}^{R} I(\boldsymbol{\alpha}_{jk} \in \mathrm{CI}_{jk}^{(r)})$. Moreover, the *average running time* with 500 replications can be obtained.

The simulation results are obtained using the four sampling methods. Tables 1 to 2 report the simulation results with 500 replications for examples 1 to 3, with error terms in the normal distribution. As the simulation results are similar for the t-distribution, Tables 4 to 5 only present the results for the CS method with the different examples.

First, by using the CS method, without the entire network data, we only need $n_s = 200$ for the estimates to perform effectively. Secondly, in each simulation model, the maximum bias of the CS method is 0.015. This is because the CS

TABLE 1

Simulation results with 500 replications for example 1 (dyad independence network). The error term $\varepsilon_i \sim N(\mathbf{0}_2, \Sigma_e)$. The estimates on the entire network and average running time are abbreviated as EWN and T, respectively.

n_s=N	Index	$\boldsymbol{\alpha}_{11}$	$\boldsymbol{\alpha}_{21}$	$\boldsymbol{\alpha}_{12}$	$\boldsymbol{\alpha}_{22}$	B_{11}	B_{21}	B_{12}	B_{22}
	Bias	0.003	-0.001	0.002	-0.001	0.001	-0.001	0.002	-0.002
EWN	RMSE	0.063	0.046	0.083	0.054	0.017	0.017	0.020	0.021
T:143.2s	CP	0.934	0.946	0.948	0.958	0.962	0.972	0.936	0.950
n_s=200	Index	$\boldsymbol{\alpha}_{11}$	$\boldsymbol{\alpha}_{21}$	$\boldsymbol{\alpha}_{12}$	$\boldsymbol{\alpha}_{22}$	B_{11}	B_{21}	B_{12}	B_{22}
	Bias	0.006	-0.005	-0.002	0.004	0.001	0.000	0.005	-0.002
CS	RMSE	0.165	0.144	0.246	0.141	0.053	0.050	0.064	0.063
T: 14.0s	CP	0.952	0.954	0.946	0.948	0.952	0.958	0.950	0.960
	Bias	-0.242	0.162	0.002	-0.084	0.003	-0.001	0.003	0.004
SRS	RMSE	0.257	0.175	0.111	0.110	0.057	0.057	0.067	0.067
T: 13.6s	CP	0.166	0.268	0.936	0.780	0.956	0.954	0.946	0.952
	Bias	-0.228	0.155	0.001	-0.087	0.001	-0.002	-0.007	0.006
SN	RMSE	0.246	0.171	0.116	0.116	0.052	0.050	0.070	0.064
T: 14.2s	CP	0.326	0.404	0.946	0.808	0.948	0.950	0.938	0.948
	Bias	-0.228	0.156	-0.006	-0.081	-0.004	0.006	0.000	0.005
MHRW	RMSE	0.250	0.174	0.128	0.114	0.056	0.052	0.067	0.062
T: 13.9s	CP	0.342	0.442	0.940	0.836	0.950	0.954	0.960	0.944
n_s=500	Index	$\boldsymbol{\alpha}_{11}$	$\boldsymbol{\alpha}_{21}$	$\boldsymbol{\alpha}_{12}$	$\boldsymbol{\alpha}_{22}$	B_{11}	B_{21}	B_{12}	B_{22}
	Bias	0.000	0.004	-0.011	0.008	-0.001	0.000	-0.003	0.001
CS	RMSE	0.106	0.091	0.156	0.087	0.033	0.032	0.041	0.039
T: 21.7s	CP	0.962	0.958	0.950	0.946	0.952	0.952	0.972	0.960
	Bias	-0.157	0.106	-0.002	-0.054	-0.003	0.004	0.000	-0.002
SRS	RMSE	0.182	0.125	0.115	0.088	0.033	0.034	0.041	0.043
T: 25.2s	CP	0.550	0.646	0.934	0.884	0.970	0.948	0.960	0.934
	Bias	-0.151	0.098	-0.002	-0.045	-0.002	0.001	0.000	0.000
SN	RMSE	0.174	0.120	0.113	0.087	0.031	0.031	0.040	0.038
T: 26.6s	CP	0.554	0.652	0.938	0.894	0.954	0.950	0.938	0.954
	Bias	-0.154	0.114	-0.010	-0.060	-0.003	0.002	-0.001	0.001
MHRW	RMSE	0.174	0.129	0.110	0.089	0.031	0.033	0.038	0.038
T: 26.5s	CP	0.554	0.558	0.950	0.864	0.952	0.940	0.966	0.960

method leads to an unbiased estimation. As n_s increases, the RMSE for each parameter decreases. Moreover, the empirical coverage probabilities of the CS estimators are stable at a level of 95%, which implies that the estimated standard error effectively approximates the true standard error SE_{jk}. In contrast, for the other three traditional sampling methods, the diagonal elements of $\boldsymbol{\alpha}$ are substantially underestimated. Moreover, the empirical coverage probabilities are far from the level of 95%. Thus, we can conclude that it is necessary to consider the related information among the sampled nodes $\mathcal{G}_y$ (i.e., the double-layer

TABLE 2

Simulation results with 500 replications for example 2 (stochastic block network). The error term $\varepsilon_i \sim N(\mathbf{0}_2, \Sigma_e)$. The estimates on the entire network and average running time are abbreviated as EWN and ART, respectively.

n_s=N	Index	$\boldsymbol{\alpha}_{11}$	$\boldsymbol{\alpha}_{21}$	$\boldsymbol{\alpha}_{12}$	$\boldsymbol{\alpha}_{22}$	B_{11}	B_{21}	B_{12}	B_{22}
	Bias	0.000	0.001	0.001	-0.001	-0.002	-0.001	0.000	0.000
EWN	RMSE	0.015	0.012	0.018	0.013	0.018	0.018	0.021	0.019
T: 23.6s	CP	0.952	0.954	0.946	0.948	0.952	0.958	0.950	0.960
n_s=200	Index	$\boldsymbol{\alpha}_{11}$	$\boldsymbol{\alpha}_{21}$	$\boldsymbol{\alpha}_{12}$	$\boldsymbol{\alpha}_{22}$	B_{11}	B_{21}	B_{12}	B_{22}
	Bias	-0.001	0.002	-0.001	0.003	-0.004	0.004	-0.001	0.003
CS	RMSE	0.042	0.035	0.059	0.036	0.053	0.055	0.068	0.068
T: 14.6s	CP	0.936	0.952	0.954	0.946	0.946	0.950	0.958	0.958
	Bias	-0.094	0.059	-0.014	-0.027	-0.079	0.032	0.009	0.004
SRS	RMSE	0.126	0.091	0.100	0.071	0.101	0.073	0.068	0.067
T: 11.3s	CP	0.846	0.870	0.952	0.942	0.780	0.920	0.948	0.954
	Bias	-0.060	0.046	-0.013	-0.014	-0.017	0.019	-0.001	-0.002
SN	RMSE	0.080	0.064	0.067	0.046	0.061	0.062	0.065	0.066
T: 11.4s	CP	0.804	0.794	0.942	0.938	0.950	0.954	0.946	0.948
	Bias	-0.029	0.008	-0.016	-0.003	-0.030	0.022	0.010	0.001
MHRW	RMSE	0.074	0.055	0.063	0.045	0.064	0.057	0.068	0.063
T: 11.6s	CP	0.796	0.854	0.955	0.937	0.929	0.941	0.937	0.941
n_s=500	Index	$\boldsymbol{\alpha}_{11}$	$\boldsymbol{\alpha}_{21}$	$\boldsymbol{\alpha}_{12}$	$\boldsymbol{\alpha}_{22}$	B_{11}	B_{21}	B_{12}	B_{22}
	Bias	-0.002	0.003	-0.001	0.001	0.000	0.001	0.002	-0.001
CS	RMSE	0.024	0.024	0.036	0.023	0.036	0.035	0.042	0.040
T: 14.7s	CP	0.956	0.956	0.946	0.964	0.946	0.972	0.964	0.948
	Bias	-0.042	0.030	-0.007	-0.012	-0.004	0.015	-0.001	-0.004
SRS	RMSE	0.049	0.037	0.035	0.025	0.041	0.040	0.042	0.042
T: 12.1s	CP	0.682	0.746	0.948	0.910	0.932	0.928	0.958	0.950
	Bias	-0.037	0.031	-0.012	-0.011	0.020	-0.015	-0.006	0.004
SN	RMSE	0.046	0.039	0.037	0.026	0.040	0.039	0.040	0.040
T: 12.7s	CP	0.704	0.746	0.944	0.928	0.932	0.932	0.956	0.958
	Bias	-0.043	0.028	-0.006	-0.012	-0.016	0.009	0.002	-0.001
MHRW	RMSE	0.050	0.036	0.036	0.026	0.037	0.037	0.042	0.042
T: 12.4s	CP	0.642	0.760	0.946	0.912	0.946	0.952	0.942	0.944

supplementary information). Moreover, in the face of a large-scale network with limited storage and computational resources, our method can obtain consistent estimates on the sampled data.

3.3. Third-party restaurant consumer review website data analysis

In this section, we present the CS implementation for the MSAR model on third-party restaurant consumer review website data. On the website, users can comment on the merchant that they visited. We regard the online user behavior

TABLE 3

Simulation results with 500 replications for example 3 (power-law distribution network). The error term $\varepsilon_i \sim N(\mathbf{0}_2, \Sigma_e)$. The estimates on the entire network and average running time are abbreviated as EWN and ART, respectively.

n_s=N	Index	$\boldsymbol{\alpha}_{11}$	$\boldsymbol{\alpha}_{21}$	$\boldsymbol{\alpha}_{12}$	$\boldsymbol{\alpha}_{22}$	B_{11}	B_{21}	B_{12}	B_{22}
	Bias	0.001	0.001	0.000	0.000	0.000	0.001	0.000	0.000
EWN	RMSE	0.020	0.013	0.024	0.017	0.018	0.017	0.021	0.021
T: 23.7s	CP	0.952	0.954	0.946	0.948	0.952	0.958	0.950	0.960
n_s=200	Index	$\boldsymbol{\alpha}_{11}$	$\boldsymbol{\alpha}_{21}$	$\boldsymbol{\alpha}_{12}$	$\boldsymbol{\alpha}_{22}$	B_{11}	B_{21}	B_{12}	B_{22}
	Bias	0.002	-0.004	0.002	-0.002	-0.003	0.003	0.008	-0.002
CS	RMSE	0.056	0.039	0.067	0.047	0.054	0.057	0.070	0.065
T: 13.1s	CP	0.948	0.956	0.946	0.958	0.950	0.952	0.948	0.938
	Bias	-0.127	0.088	-0.006	-0.047	0.010	0.000	-0.002	0.003
SRS	RMSE	0.153	0.105	0.103	0.079	0.059	0.058	0.066	0.064
T: 10.6s	CP	0.630	0.732	0.966	0.882	0.954	0.950	0.950	0.948
	Bias	-0.105	0.072	0.001	-0.039	-0.046	0.025	-0.022	0.013
SN	RMSE	0.122	0.085	0.077	0.063	0.908	0.480	0.620	0.330
T: 10.8s	CP	0.610	0.668	0.940	0.890	0.968	0.960	0.954	0.948
	Bias	-0.115	0.080	-0.004	-0.043	0.008	0.011	0.020	-0.001
MHRW	RMSE	0.132	0.093	0.073	0.068	0.319	0.068	0.262	0.070
T: 10.7s	CP	0.588	0.616	0.950	0.888	0.954	0.968	0.950	0.942
n_s=500	Index	$\boldsymbol{\alpha}_{11}$	$\boldsymbol{\alpha}_{21}$	$\boldsymbol{\alpha}_{12}$	$\boldsymbol{\alpha}_{22}$	B_{11}	B_{21}	B_{12}	B_{22}
	Bias	0.002	-0.001	-0.001	0.004	-0.001	0.002	0.002	0.001
CS	RMSE	0.035	0.022	0.042	0.030	0.034	0.032	0.038	0.039
T: 13.5s	CP	0.970	0.956	0.936	0.946	0.960	0.926	0.956	0.954
	Bias	-0.092	0.074	-0.005	-0.037	0.001	-0.005	0.004	0.006
SRS	RMSE	0.102	0.078	0.048	0.047	0.034	0.034	0.041	0.042
T: 11.1s	CP	0.434	0.178	0.960	0.714	0.958	0.968	0.946	0.964
	Bias	-0.089	0.061	0.004	-0.030	0.009	0.013	-0.001	-0.001
SN	RMSE	0.097	0.066	0.042	0.043	0.037	0.038	0.043	0.043
T: 11.3s	CP	0.404	0.392	0.962	0.806	0.954	0.942	0.950	0.946
	Bias	-0.101	0.069	-0.001	-0.040	-0.013	0.002	0.001	-0.003
MHRW	RMSE	0.108	0.074	0.041	0.049	0.035	0.033	0.043	0.039
T: 11.3s	CP	0.286	0.284	0.956	0.718	0.960	0.968	0.944	0.954

features (namely, *activity* and *interaction*) as responses. The user *activity* represents the number of users' comments for a merchant over the past month. The user *interaction* represents the communications among users, which is measured by the number of likes and comments between users. The log-transformed values of the *activity* and *interaction* are aggregated, following which the responses are normalized with mean 0 and variance 1. The histograms of the log-transformed *activity* and *interaction* are presented in Figure 4. We collect three user features as the exogenous nodal covariates: the registration time, VIP, and gender

TABLE 4

CS with 500 replications for three examples. The error term ε_i follows t(5) with mean $\mathbf{0}_2$ and covariance Σ_e. The results of the average running time are similar to those of the former cases.

n_s=200	$\boldsymbol{\alpha}_{11}$	$\boldsymbol{\alpha}_{21}$	$\boldsymbol{\alpha}_{12}$	$\boldsymbol{\alpha}_{22}$	B_{11}	B_{21}	B_{12}	B_{22}
EXAMPLE 1: DYAD INDEPENDENCE NETWORK								
Bias	-0.015	0.002	-0.015	-0.012	0.002	0.003	0.008	-0.005
RMSE	0.183	0.172	0.282	0.158	0.073	0.070	0.085	0.082
CP	0.956	0.948	0.938	0.958	0.954	0.928	0.960	0.956
EXAMPLE 2: STOCHASTIC BLOCK NETWORK								
Bias	0.000	0.003	-0.005	0.000	0.001	0.001	0.008	-0.001
RMSE	0.048	0.042	0.069	0.039	0.070	0.066	0.091	0.084
CP	0.956	0.930	0.952	0.936	0.954	0.948	0.952	0.952
EXAMPLE 3: POWER LAW NETWORK								
Bias	0.005	-0.001	0.003	0.004	-0.004	0.010	0.001	0.006
RMSE	0.067	0.042	0.075	0.053	0.070	0.070	0.081	0.083
CP	0.960	0.958	0.966	0.960	0.946	0.938	0.938	0.938
n_s=500	$\boldsymbol{\alpha}_{11}$	$\boldsymbol{\alpha}_{21}$	$\boldsymbol{\alpha}_{12}$	$\boldsymbol{\alpha}_{22}$	B_{11}	B_{21}	B_{12}	B_{22}
EXAMPLE 1: DYAD INDEPENDENCE NETWORK								
Bias	0.010	0.004	-0.012	0.007	0.000	0.002	0.000	0.002
RMSE	0.117	0.103	0.177	0.098	0.041	0.041	0.050	0.053
CP	0.950	0.956	0.930	0.958	0.952	0.958	0.948	0.946
EXAMPLE 2: STOCHASTIC BLOCK NETWORK								
Bias	-0.002	0.003	0.001	0.000	-0.001	0.002	0.002	0.000
RMSE	0.031	0.028	0.042	0.026	0.041	0.046	0.053	0.054
CP	0.928	0.974	0.948	0.946	0.928	0.960	0.956	0.954
EXAMPLE 3: POWER LAW NETWORK								
Bias	0.004	-0.001	0.003	-0.002	0.002	-0.002	0.001	0.002
RMSE	0.042	0.028	0.048	0.034	0.044	0.045	0.055	0.054
CP	0.950	0.956	0.930	0.958	0.952	0.958	0.948	0.946

(gender = 1 for male and 0 for female). The histograms of the log-transformed registration time are provided in Figure 4. Moreover, 68.4% of users are VIP (VIP = 1 for VIP users) and 98.1% of users are male. All the continuous variables are standardized with mean 0 and variance 1.

Subsequently, for the network structure, we consider the adjacency matrix A constructed based on the user comments on merchants. If two users i_1, i_2 have reviewed the same merchant, $a_{i_1 i_2} = 1$ is recorded, and $a_{i_1 i_2} = 0$ otherwise. The histograms of the degrees are presented in Figure 4. The nodes with degrees higher than 500 are deleted. Moreover, 5534 users are involved in the real data analysis.

We treat the above $N = 5534$ nodes and corresponding edges as if they are the entire network. Furthermore, we regard the least-squares estimates on the entire network as true parameters. The estimation results on the entire

TABLE 5
Estimates on the entire network with 500 replications for three examples. The error term ε_i follows t(5) with mean $\mathbf{0}_2$ and covariance Σ_e.

$n_s = N$	$\boldsymbol{\alpha}_{11}$	$\boldsymbol{\alpha}_{21}$	$\boldsymbol{\alpha}_{12}$	$\boldsymbol{\alpha}_{22}$	B_{11}	B_{21}	B_{12}	B_{22}
EXAMPLE 1: DYAD INDEPENDENCE NETWORK								
Bias	-0.002	0.001	0.007	0.000	0.001	-0.001	0.000	-0.001
RMSE	0.072	0.06	0.098	0.061	0.021	0.022	0.024	0.025
CP	0.946	0.946	0.940	0.950	0.958	0.966	0.942	0.934
EXAMPLE 2: STOCHASTIC BLOCK NETWORK								
Bias	-0.001	0.002	-0.001	0.001	0.000	0.000	0.002	-0.002
RMSE	0.016	0.016	0.024	0.014	0.023	0.024	0.027	0.027
CP	0.942	0.938	0.934	0.954	0.950	0.964	0.932	0.940
EXAMPLE 3: POWER LAW NETWORK								
Bias	0.001	0.000	0.000	0.000	0.000	0.001	0.001	0.000
RMSE	0.022	0.017	0.026	0.019	0.022	0.022	0.025	0.027
CP	0.956	0.952	0.938	0.940	0.946	0.954	0.946	0.950

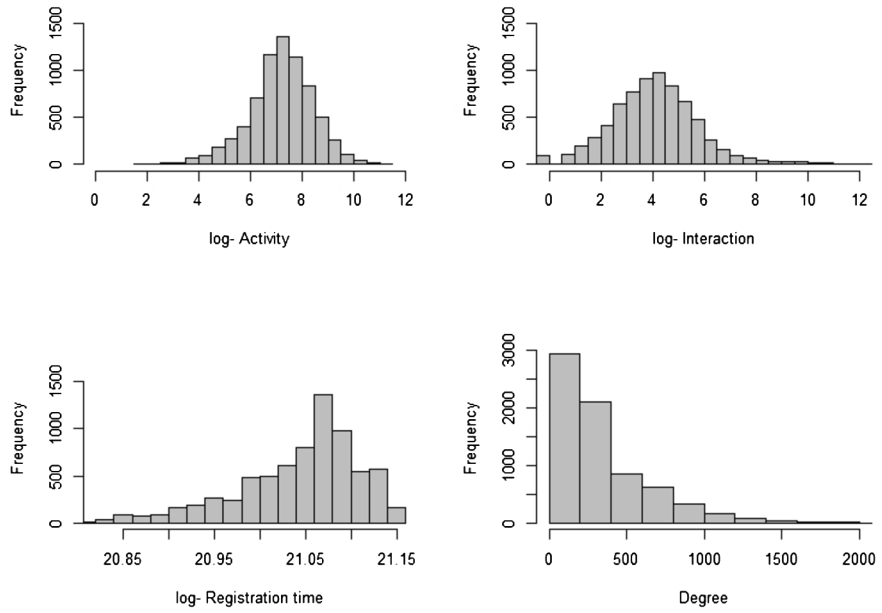

FIG 4. *Descriptive analysis regarding real data. Top left panels: histogram of log-activity; top right panels: histogram of log-interaction; bottom left panels: histogram of standardized log-registration time; bottom right panels: histogram of degrees.*

network are summarized in Table 6. The corresponding p-values are all smaller than 0.01. According to Table 6, the user *activity* and *interaction* have a positive *intra-activity effect* and *extra-activity effect*. For the exogenous nodal covariates, VIP users and male users exhibit significantly higher *activity* and *interaction*.

TABLE 6

CS estimates with 200 replications on real data. The activity, interaction, and registration times are abbreviated as ACT, INT, and RET, respectively. The "" indicates that the estimates on the entire network are significant under a level of 0.01. The represented estimates on the sampled real data are the mean of 200 replications. The empirical powers of the estimates are reported in brackets alongside the corresponding estimates. The estimates of the intercept term are omitted.*

	Whole network		$n_s = 300$		$n_s = 500$	
	ACT	INT	ACT	INT	ACT	INT
ACT ($W\mathbb{Y}_{\cdot 1}$)	0.605*	0.375*	0.606(1.00)	0.377(1.00)	0.608(1.00)	0.372(1.00)
INT ($W\mathbb{Y}_{\cdot 2}$)	0.259*	0.599*	0.259(0.99)	0.600(1.00)	0.262(1.00)	0.597(1.00)
RET	0.034*	−0.105*	0.034(0.99)	−0.105(1.00)	0.034(1.00)	−0.105(1.00)
VIP	1.076*	0.963*	1.076(1.00)	0.963(1.00)	1.076(1.00)	0.963(1.00)
GENDER	0.414*	0.456*	0.414(1.00)	0.457(1.00)	0.414(1.00)	0.456(1.00)

Moreover, it is found that users with an earlier registration time have higher *activity* but lower *interaction*.

Thereafter, we use only a small number of nodes to approximate the true parameters. We set the sample size to 300 and 500. For CS, we take the average of the estimates obtained by $R = 200$ replications. The estimation results are summarized in Table 6. We consider the bias, RMSE, and empirical rejection probability (ERP) as the performance measures. We set the size as $\alpha = 0.01$. The ERP for $\boldsymbol{\alpha}_{j_1 j_2}$ is computed as $R^{-1} \sum_{r=1}^{R} I(|\hat{\boldsymbol{\alpha}}_{j_1 j_2} / \widehat{SE}_{j_1 j_2}| > z_{\alpha/2})$. The ERP represents the empirical size or power, where $\hat{\boldsymbol{\alpha}}_{j_1 j_2}$ is 0 or not. Note that the three traditional sampling methods perform poorly on the bias and ERP. Thus, we focus on the CS results.

We compare the estimates obtained from the sampled data with the true parameters. In Table 6, we abbreviate the *activity*, *interaction*, and registration time as ACT, INT, and RET, respectively. The CS performs effectively on a finite sample size. First, we regard the estimates as unbiased because the maximum bias is 0.003. Moreover, as the sample size increases, the RMSE for all parameters decreases. Secondly, for significant true parameters, the ERP approaches 100% on the sampled data when $n_s = 500$. Consequently, CS can lead to a consistent estimation on the sampled data and the double-layer supplementary information from the sampled nodes is necessary for consistent estimation. Furthermore, compared to the neighbors' *interaction*, the neighbors' *activity* has a stronger impact on the users' *activity* and *interaction*. We suggest rewarding users and encouraging them to leave more reviews on merchants to make the website more active.

4. Concluding remarks

In this study, to approximate the estimation for MSAR on the entire network accurately, we have proposed the CS method for the LSE estimator in a general

framework. Together with the sampled nodal features and double-layer supplementary information, we constructed the least-squares objective function on the sampled data. The identification condition for the parameters on the sampled network was theoretically investigated. Furthermore, the sample size condition and asymptotic properties of the least-squares estimators were provided. Numerical results for the simulated data and real data were presented.

In conclusion, we consider three interesting topics for future studies. First, from a model-based viewpoint, we will focus on the sampling scheme for the MSAR model. Moreover, numerous other models, such as the exponential random graph model and network vector autoregressive model, could also be used to analyze the network data. Thus, to estimate the parameters in these models, extending the CS method could be an interesting research problem. Secondly, for the parameters in different network models and the properties of the network, the bootstrap estimator of the parameters deserves a separate study; see [35, 15, 8] for further discussions. Thirdly, we have regarded the network structure and nodal features as fixed during the observation period. However, in practice, the network structure, many responses, and exogenous covariates are observed in time series. Thus, a dynamic sampling mechanism and corresponding estimators could be designed.

Appendix A

In Appendix A, we present certain notations and verify the least-squares objective function (2.4) in Appendix A.1. Next, the proof of Proposition 1 is presented in Appendix A.2.

A.1. Notations and Verification of (2.4)

In this section, we first provide the notations for several useful matrices and vectors. The MSAR model on the entire network can be rewritten into the vector form as $\mathcal{Y} = \left(\boldsymbol{\alpha}^\top \otimes W\right)\mathcal{Y} + \widetilde{\mathbb{X}}\beta + \mathcal{E}$, where $\widetilde{\mathbb{X}} = I_p \otimes \mathbb{X}$ and $\mathcal{Y} = \mathrm{vec}(\mathbb{Y}) \in \mathbb{R}^{Np}$. The vector form can also be expressed as $\mathcal{Y} = S^{-1}(\widetilde{\mathbb{X}}\beta + \mathcal{E})$, where $S = I_{Np} - \boldsymbol{\alpha}^\top \otimes W$. Consider a matrix R with n_1 rows and n_2 columns. Here, R_{ij} represents the ith row and jth column element in matrix R, $R_{i,\cdot}$ is the i_1th row of matrix R, and $R_{i,\cdot}^\top \in \mathbb{R}^{n_2}$. Furthermore, $R_{\cdot,j} \in \mathbb{R}^{n_1}$ is the jth column of matrix R. We define $R_{i,(-i)}$ as the ith row of matrix R without the ith element, and $R_{i,(-i)}^\top \in \mathbb{R}^{n_2-1}$. Similarly, we define $R_{(-i),i} \in \mathbb{R}^{n_1-1}$ as the ith column vector of matrix R without the ith element, and $R_{(-i),(-i)} \in \mathbb{R}^{(n_1-1)\times(n_2-1)}$ as R without the ith column and ith row. Furthermore, with respect to $\boldsymbol{\alpha}_{j_1j_2}$ and $\boldsymbol{\beta}$, the first-order derivatives of R can be written as $R^{\boldsymbol{\alpha}}_{j_1j_2}$ and $R_{\boldsymbol{\beta}}$. We denote $R^{\boldsymbol{\alpha}}_{j_1j_2k_1k_2}$, $R^{\boldsymbol{\alpha}}_{j_1j_2\boldsymbol{\beta}}$, and $R_{\boldsymbol{\beta}\boldsymbol{\beta}}$ to represents the second-order derivatives $\partial^2 R/\partial\boldsymbol{\alpha}_{j_1j_2}\partial\boldsymbol{\alpha}_{k_1k_2}$, $\partial^2 R/\partial\boldsymbol{\alpha}_{j_1j_2}\partial\boldsymbol{\beta}$, and $\partial^2 R/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}$, respectively.

Next, we present the form of the BLP of the $\mathbb{Y}_{ij}$ element. When the error term $\mathbb{E}_{\cdot j}$ follows a normal distribution; the BLP of $\mathbb{Y}_{ij}$ given by $\mathbb{Y}_{-(i,j)}$ has the

same form as the conditional mean [31]. Let $n_{ij} = (j-1) \times N + i$; then, the BLP of $\mathbb{Y}_{ij}$ can be expressed as

$$F\left(\mathbb{Y}_{ij}|\mathbb{Y}_{-(i,j)}\right) = \boldsymbol{\mu}_{ij} - \Omega_{n_{ij},n_{ij}}^{-1}\Omega_{n_{ij},(-n_{ij})}(\mathcal{Y}_{(-n_{ij})} - E\{\mathcal{Y}_{(-n_{ij})}\}), \tag{A.1}$$

Moreover, the precision matrix of $\mathcal{Y}$ can be expressed as

$$\begin{aligned}\Omega = \Sigma^{-1} &= \begin{pmatrix} \Omega_{n_{ij},n_{ij}} & \Omega_{n_{ij},(-n_{ij})} \\ \Omega_{(-n_{ij}),n_{ij}} & \Omega_{(-n_{ij}),(-n_{ij})} \end{pmatrix} \\ &= (I_{Np} - \boldsymbol{\alpha}^\top \otimes W)^\top(\Omega_e \otimes I_N)(I_{Np} - \boldsymbol{\alpha}^\top \otimes W) \\ &= (\Omega_e \otimes I_N) - \mathcal{O}^\top - \mathcal{O} + \mathcal{Q}.\end{aligned}$$

where $\mathcal{O} = (\Omega_e\boldsymbol{\alpha}^\top) \otimes W$ and $\mathcal{Q} = (\boldsymbol{\alpha}\Omega_e\boldsymbol{\alpha}^\top) \otimes (W^\top W)$. Specifically, because $W_{ii} = 0$, we have $\Omega_{n_{ij},n_{ij}} = \Omega_{e,jj} + \mathcal{Q}_{n_{ij},n_{ij}} = m_{n_{ij},n_{ij}}^{-1}$, $\Omega_{n_{ij},(-n_{ij})} = -\mathcal{O}_{n_{ij},(-n_{ij})} - \mathcal{O}_{(-n_{ij}),n_{ij}}^\top + \mathcal{Q}_{n_{ij},(-n_{ij})}, m = \text{diag}^{-1}(\Omega) = [\text{diag}\left(\Sigma_e^{-1}\right) \otimes I_N + \text{diag}\left(\boldsymbol{\alpha}\Sigma_e^{-1}\boldsymbol{\alpha}^\top\right) \otimes \text{diag}\left(W^\top W\right)]^{-1}$. Next, we split the product in the form of vectors in (A.1) (i.e., $\Omega_{n_{ij},(-n_{ij})}(\mathcal{Y}_{(-n_{ij})} - E\{\mathcal{Y}_{(-n_{ij})}\})$) into the sum of the multipliers of their components. Thus, we have

$$\Omega_{n_{i_1j_1},(-n_{i_1j_1})}(\mathcal{Y}_{(-n_{i_1j_1})} - E\{\mathcal{Y}_{(-n_{i_1j_1})}\}) = \sum_{(i_2,j_2)\neq(i_1,j_1)}^{(N,p)} r_{i_1j_1i_2j_2}\left(\mathbb{Y}_{i_2j_2} - \boldsymbol{\mu}_{i_2j_2}\right)$$

where $r_{i_1j_1i_2j_2} = m_{n_{i_1j_1},n_{i_1j_1}}^{-1}\{\mathcal{O}_{n_{i_1j_1},n_{i_2j_2}} + \mathcal{O}_{n_{i_2j_2},n_{i_1j_1}} - \mathcal{Q}_{n_{i_1j_1},n_{i_2j_2}}\}$. We have $m_{n_{i_1j_1},n_{i_1j_1}} = \omega_{e,j_1j_1} + \left(\boldsymbol{\alpha}_{j_1\cdot}\Omega_e\boldsymbol{\alpha}_{j_1\cdot}^\top\right)\left(\sum_{i\in\mathcal{G}_y\bigcup\mathcal{G}_x} w_{ii_1}^2\right)$, $\mathcal{O}_{n_{i_1j_1},n_{i_2j_2}} = \boldsymbol{\alpha}_{j_1\cdot}\Omega_{e,\cdot j_2}$ $w_{i_2i_1}$, $\mathcal{O}_{n_{i_2j_2},n_{i_1j_1}} = \Omega_{e,j_2\cdot}\boldsymbol{\alpha}_{\cdot j_1}w_{i_1i_2}$, $\mathcal{Q}_{n_{i_1j_1},n_{i_2j_2}} = \left(\boldsymbol{\alpha}_{j_1\cdot}\Omega_e\boldsymbol{\alpha}_{j_2\cdot}^\top\right)\left(\sum_{i\in\mathcal{G}_y\bigcup\mathcal{G}_x} w_{ii_1}\right.$ $\left.w_{ii_2}\right)$. Thus, the calculation of each element in $r_{i_1j_1i_2j_2}$ only involve the three types of unsampled nodes in Figure 3. For any node $i_1 \in \mathcal{G}_y$, since only nodes $i_2 \in \mathcal{G}_y \bigcup \mathcal{G}_x$ are involved in (A.1), we can obtain the BLP of $\mathbb{Y}_{i_1j_1}$ in (2.2)-(2.3) as follows:

$$\begin{aligned}F\left(\mathbb{Y}_{i_1j_1}|\mathbb{Y}_{-(i_1,j_1)}\right) &= \boldsymbol{\mu}_{i_1j_1} + \sum_{(i_2,j_2)\neq(i_1,j_1)}^{(\mathcal{G}_y\bigcup\mathcal{G}_x,p)} r_{i_1j_1i_2j_2}\left(\mathbb{Y}_{i_2j_2} - \boldsymbol{\mu}_{i_2j_2}\right), \\ r_{i_1j_1i_2j_2} &= \frac{\boldsymbol{\alpha}_{j_1\cdot}\Omega_{e,\cdot j_2}w_{i_2i_1} + \Omega_{e,j_2\cdot}\boldsymbol{\alpha}_{\cdot j_1}w_{i_1i_2} - \left(\boldsymbol{\alpha}_{j_1\cdot}\Omega_e\boldsymbol{\alpha}_{j_2\cdot}^\top\right)\left(\sum_{i\in\mathcal{G}_y\bigcup\mathcal{G}_x} w_{ii_1}w_{ii_2}\right)}{\omega_{e,j_1j_1} + \left(\boldsymbol{\alpha}_{j_1\cdot}\Omega_e\boldsymbol{\alpha}_{j_1\cdot}^\top\right)\left(\sum_{i\in\mathcal{G}_y\bigcup\mathcal{G}_x} w_{ii_1}^2\right)}.\end{aligned} \tag{A.2}$$

We define $\boldsymbol{\mu}_c = \text{vec}\{F\left(\mathbb{Y}_{ij}|\mathbb{Y}_{-(i,j)}\right), 1 \le i,j \le N\} \in \mathbb{R}^N$, $\boldsymbol{\mu} = E(\mathcal{Y}) \in \mathbb{R}^N$, and we can write (A.2) in the vectorization form as $\boldsymbol{\mu}_c = \boldsymbol{\mu} - m(\Omega - m^{-1})(\mathcal{Y} - \boldsymbol{\mu})$. Thus, we have $\mathcal{Y} - \boldsymbol{\mu}_c = m\Omega(\mathcal{Y} - \boldsymbol{\mu})$.

As a result, the objective function with all N nodes can be expressed as $L_w(\boldsymbol{\theta}) = ||\mathcal{Y} - \boldsymbol{\mu}_c||^2$. Instead of using all N nodes to establish the objective function, we construct the objective function L on the sampled data. The form of L in (2.4) can be written as

$$L(\boldsymbol{\theta}) = ||G(\mathcal{Y} - \boldsymbol{\mu}_c)||^2 = H^\top H, \tag{A.3}$$

where $H = G(\mathcal{Y} - \boldsymbol{\mu}_c) = mGS^\top(\Omega_e \otimes I_N)(S\mathcal{Y} - \widetilde{\mathbb{X}}\boldsymbol{\beta})$ and $G = I_p \otimes g$.

A.2. Proof of Proposition 1

Proof. We need to prove these two inequalities: $n_s^{-2}\operatorname{tr}(M_{k_1k_2}^{*\top}M_{j_1j_2}^{*}M_{j_1j_2}^{*\top}M_{k_1k_2}^{*}) \to 0$ and $n_s^{-2}V_{j_1j_2}^{*\top}M_{k_1k_2}^{*}M_{k_1k_2}^{*\top}V_{j_1j_2}^{*} \to 0$. The upper bounds of the elements of matrix $M_{j_1j_2}^{*}, V_{j_1j_2}^{*}$ need to be considered to prove these two conditions. Following Lemma 3 in [46], we define $W^* = W^\top g + W^\top W g + W^\top W g W^\top + gW^\top\{W(\sum_{i=1}^{K} W^i + \mathbf{1}_N\pi^\top)\}$, where K is a large finite integer [2]. Under condition (C2.2), we have $\lambda_{\max}\{g(W^\top + W)g\} = O(\log n_s)$. Then, it could be verified that $\lambda_{\max}\{g(W^\top W)g\} = O(\log n_s^2)$. Furthermore, according to Lemma 3 in [46], we have

$$\lambda_{\max}(W^{*\top}W^*) = \begin{cases} O\{(\log n_s)^6 N^{1/2-\delta}\}, & 1 < \delta < 1/2 \\ O\{(\log n_s)^6 (\log N)^{2(K+2)}\}, & \delta = 1/2 \end{cases} \tag{A.4}$$

In this case, $\mathbf{1}_N$ is an N-dimensional column vector with all elements equal to 1. Note that $|A|_e = (|A_{ij}|)$ takes the absolute value of each element in A. Moreover, $|A|_e \preceq B$ means that each $|A_{ij}|$ is smaller than the corresponding element B_{ij}. Furthermore, we have $|M_{j_1j_2}^{*}|_e \preceq c_{1M}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes W^*$ and $|V_{j_1j_2}^{*}|_e \preceq c_{1M}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes W^*|(I_p \otimes \mathbb{X})\boldsymbol{\beta}|_e$. Then, taking the upper bound of each element of the matrix as $n_s \to \infty$, we assume that there exists $\epsilon > 0$, $n_s = O(N^{1-2\delta+\epsilon})$. Thus, we have

$$\begin{aligned} n_s^{-2}\operatorname{tr}(M_{k_1k_2}^{*\top}M_{j_1j_2}^{*}M_{j_1j_2}^{*\top}M_{k_1k_2}^{*}) &\le n_s^{-2}c_{3m}\operatorname{tr}(W^{*\top}W^*W^{*\top}W^*) \\ &\le n_s^{-1}c_{3m}\lambda_{\max}^2(W^{*\top}W^*) \to 0, \end{aligned}$$

where c_{3m} is a finite constant. Similarly, for the second condition, we have

$$\begin{aligned} n_s^{-2}V_{j_1j_2}^{*\top}M_{k_1k_2}^{*}M_{k_1k_2}^{*\top}V_{j_1j_2}^{*} &\le n_s^{-2}c_u(|\widetilde{\mathbb{X}}\boldsymbol{\beta}|_e)^\top(\mathbf{1}_p\mathbf{1}_p^\top) \otimes (W^{*\top}W^*W^{*\top}W^*)(|\widetilde{\mathbb{X}}\boldsymbol{\beta}|_e) \\ &\le n_s^{-1}c_u\|\widetilde{\mathbb{X}}\boldsymbol{\beta}\|^2 n_s^{-1}\lambda_{\max}^2(W^{*\top}W^*), \end{aligned}$$

where $n_s^{-1}\lambda_{\max}^2(W^{*\top}W^*) \to 0$, $n_s^{-1}\|\widetilde{\mathbb{X}}\boldsymbol{\beta}\|^2 = O(1)$ under condition (C3), and c_u is a finite positive constant. As a result, we have $n_s^{-2}V_{j_1j_2}^{*\top}M_{k_1k_2}^{*}M_{k_1k_2}^{*\top}V_{j_1j_2}^{*} \to 0$. □

Appendix B

We present the proofs of Theorem 1, Theorem 2, and Theorem 3 in Appendix B.1, Appendix B.2, and Appendix B.3 respectively.

B.1. Proof of Theorem 1

Proof. We assume that the nodal in-degree is finite. Thus, for any sampled node i_1 in $\mathcal{G}_y$, $\sum_i w_{ii_1}^2$ and $\sum_i w_{ii_1}w_{ii_2}$ are summations of finite terms. The

computational complexity of each nonzero element in $\boldsymbol{\mu}_c$ is $O(1)$ according to (A.2). Then, there are only $n_s p$ nonzero elements in H. Similarly, the computational complexity of the objective function L and its derivatives (first–order and second–order) is linear in the sample size n_s. Further, we assume that the Newton–Raphson algorithm converges in a finite number of steps. As a result, the computational complexity is linear with the number of sampled nodes n_s. Thus, Theorem 1 is proven. □

B.2. Proof of Theorem 2

Proof. In this section, we investigate the identification condition on the sampled network. In the proof of the identification, $\boldsymbol{\theta} = \left(\mathrm{vec}(\boldsymbol{\alpha})^\top, \boldsymbol{\beta}^\top\right)^\top \in \mathbb{R}^{p^2+pq}$ represents the estimates obtained by calculation, and $\boldsymbol{\theta}_0 = \left(\mathrm{vec}(\boldsymbol{\alpha}_0)^\top, \boldsymbol{\beta}_0^\top\right)^\top$ represents the true parameter. Then, we define $\mathbb{V}_g = (I_p \otimes \mathbb{X}_g^*)(\boldsymbol{\theta}_0 - \boldsymbol{\theta}) \in \mathbb{R}^{n_s p}$ and $\mathbb{M}_g = (\Sigma_e^{-1} \otimes I_{n_s})(S_g m_g^2 S_g)(\Sigma_e^{-1} \otimes I_{n_s}) \in \mathbb{R}^{n_s p \times n_s p}$, where m_g retains the rows and columns corresponding to sampled nodes in m. Letting $\lambda_{\min}(\mathbb{M}_g) \geq c_M$, where c_M is a positive constant, we have

$$n_s^{-1}\mathbb{V}_g^\top \mathbb{M}_g \mathbb{V}_g \geq c_M \left(\boldsymbol{\theta}_0 - \boldsymbol{\theta}\right)^\top \left\{ I_p \otimes N^{-1} \left({\mathbb{X}_g^*}^\top \mathbb{X}_g^* \right) \right\} \left(\boldsymbol{\theta}_0 - \boldsymbol{\theta}\right) > 0.$$

As $n_s \to \infty$, under condition (C1), $n_s^{-1}\mathbb{V}_g^\top \mathbb{M}_g \mathbb{V}_g = 0$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ on the sampled network. We provide the details as follows.

First, we can obtain $\mathbb{V}_g = (I_p \otimes M_g)\{\mathrm{vec}(\boldsymbol{\alpha}_0) - \mathrm{vec}(\boldsymbol{\alpha})\} + (I_p \otimes \mathbb{X}_g)(\boldsymbol{\beta}_0 - \boldsymbol{\beta})$. Moreover, we define $v \in \mathbb{R}^{Np}$, $\Sigma_e^* = \Sigma_e / \mathrm{tr}(\Sigma_e)$, $c_m = \max\{m\}$. We denote $v_g^\top \in \mathbb{R}^{Np}$ with only $n_s p$ nonzero elements, and the elements corresponding to the unsampled nodes in v_g are all equal to 0. According to Theorem 2, even if $\lambda_{\min}\left(SS^\top\right) \geq c_s$, we also have

$$\begin{aligned}
\lambda_{\min}\left(S_g S_g^\top\right) &= \min_{||v_g||=1} v_g^\top (I_{Np} - D^\top \otimes W) G (I_{Np} - D \otimes W^\top) v_g \\
&\geq \min_{||v||=1} v^\top (I_{Np} - D^\top \otimes W)(I_{Np} - D \otimes W^\top) v \\
&\geq c_s > 0.
\end{aligned}$$

Moreover, there exists a finite constant c_e, and $\lambda_{\max}\left(\Sigma_e^*\right) \leq c_e$.

$$\begin{aligned}
\lambda_{\min}(\mathbb{M}_g) = \min_{\|v_n\|=1} v_n^\top \mathbb{M}_n v_n &\geq c_m^2 \lambda_{\min}\{S_g S_g^\top\} \lambda^{-2}(\Sigma_e^*) \\
&\geq c_m^2 c_s c_e^{-2} > 0.
\end{aligned}$$

As a result, letting $c_M = c_m^2 c_S c_e^{-2}$, we can conclude that $\lambda_{\min}(\mathbb{M}_g) \geq c_M$. Next, recall that $M_p = (V_p \otimes I_{n_s}) S_{0g}^{-1} [I_p \otimes (W_g \mathbb{X}_g)]\boldsymbol{\beta}$, $M_g = (M_1, ..., M_p) \in \mathbb{R}^{n_s p}$. Then, $\mathbb{X}_g^* = (M_g, \mathbb{X}_g) \in \mathbb{R}^{n_s \times (p+q)}$. Subsequently, $\lim_{n_s \to \infty} n_s^{-1} \left({\mathbb{X}_g^*}^\top \mathbb{X}_g^* \right)$ exists and is nonsingular. Finally, as a result, we have $\lim_{n_s \to \infty} n_s^{-1} \mathbb{V}_g^\top \mathbb{M}_g \mathbb{V}_g = 0$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ on the sampled network. This completes the proof. □

B.3. Proof of Theorem 3

Proof. In this section, we investigate the asymptotic properties of the least-squares estimators, defined as $\hat{\boldsymbol{\theta}}$, in two parts. In **PART 1**, $\hat{\boldsymbol{\theta}}$ is proven to be $\sqrt{n_s}$-consistent. In **PART 2**, the asymptotic normality of $\hat{\boldsymbol{\theta}}$ is proven.

PART 1: Consistency. To establish the consistency result, following [13], we demonstrate that, for any $\epsilon > 0$, there exists a constant $C_s > 0$ such that

$$\lim_{n_s \to \infty} P\{\inf_{||u||=C_s} L(\boldsymbol{\theta} + n_s^{-1/2}u) > L(\boldsymbol{\theta})\} \geq 1 - \epsilon. \tag{B.1}$$

(B.1) implies that there exists a local minimizer, written as $\hat{\boldsymbol{\theta}}$, in the ball $\{\boldsymbol{\theta} + n_s^{-1/2}uC_s : ||u|| \leq 1\}$ with probability at least $1 - \epsilon$. For (B.1), using Taylor expansion, we obtain

$$\inf_{\|u\|=1} \left\{ L\left(\boldsymbol{\theta} + n_s^{-1/2}u\right) - L(\boldsymbol{\theta}) \right\} = C_s n_s^{-1/2} \dot{L}(\boldsymbol{\theta})^\top u + 2^{-1} C_s^2 n_s^{-1} u^\top \ddot{L}(\boldsymbol{\theta}) u + o_p(1)$$

$$\geq 2^{-1} C_s^2 \lambda_{\min} \left\{ n_s^{-1} \ddot{L}(\boldsymbol{\theta}) \right\} - C_s n_s^{-1/2} \|\dot{L}(\boldsymbol{\theta})\| + o_p(1), \tag{B.2}$$

where $\dot{L}(\boldsymbol{\theta}) = \partial L(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \in \mathbb{R}^{p^2+pq}$,$\ddot{L}(\boldsymbol{\theta}) = \partial^2 L(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top \in \mathbb{R}^{(p^2+pq)\times(p^2+pq)}$. Next, we discuss $\dot{L}(\boldsymbol{\theta})$ and $\ddot{L}(\boldsymbol{\theta})$ in turn.

First, we need to prove that $\lim_{n_s \to \infty} n_s^{-1} \text{cov}\{\dot{L}(\boldsymbol{\theta}), \dot{L}(\boldsymbol{\theta})\} = \Sigma_{1s}$. We treat the right side of (B.2) as a quadratic function of C_s. Then, the linear term coefficient $n_s^{-1/2}\|\dot{L}(\boldsymbol{\theta})\|$ is $O_p(1)$. Next, based on the law of large numbers, we can prove that $n_s^{-1}\ddot{L}(\boldsymbol{\theta}) \to_p \Sigma_{2s}$. Thereafter, we can determine that $\lambda_{\min}\{n_s^{-1}\ddot{L}(\boldsymbol{\theta})\} \to_p \lambda_{\min}(\Sigma_{2s}) > 0$ since Σ_{2s} is a real symmetric matrix; i.e., the quadratic term coefficient is asymptotically positive. As a result, there exists a sufficiently large C_s, and (B.2) is positive with probability 1 when n_s goes to ∞. Furthermore, (B.1) holds, and the consistency of $\hat{\boldsymbol{\theta}}$ can be obtained.

Next, we provide the details. In **PART 1 − step 1**, we present the proof of $\lim_{n_s \to \infty} n_s^{-1} \text{cov}\{\dot{L}(\boldsymbol{\theta}), \dot{L}(\boldsymbol{\theta})\} = \Sigma_{1s}^L$. In **PART 1 − step 2**, we present the proof of $n_s^{-1}\ddot{L}(\boldsymbol{\theta}) \to_p \Sigma_{2s}^L$. Under condition (C6), we assume that the following six limits exist: $\lim_{n_s \to \infty} n_s^{-1} \text{cov}(L_{\boldsymbol{\alpha}}, L_{\boldsymbol{\alpha}}) = \Sigma_{1\boldsymbol{\alpha}}$, $\lim_{n_s \to \infty} n_s^{-1} \text{cov}(L_{\boldsymbol{\alpha}}, L_{\boldsymbol{\beta}}) = \Sigma_{1\boldsymbol{\alpha\beta}}$, $\lim_{n_s \to \infty} n_s^{-1} \text{cov}(L_{\boldsymbol{\beta}}, L_{\boldsymbol{\beta}}) = \Sigma_{1\boldsymbol{\beta}}$, $\lim_{n_s \to \infty} n_s^{-1} E(L_{\boldsymbol{\alpha\alpha}}) = \Sigma_{2\boldsymbol{\alpha}}$, $\lim_{n_s \to \infty} n_s^{-1} E(L_{\boldsymbol{\alpha\beta}}) = \Sigma_{2\boldsymbol{\alpha\beta}}$, and $\lim_{n_s \to \infty} n_s^{-1} E(L_{\boldsymbol{\beta\beta}}) = \Sigma_{2\boldsymbol{\beta}}$. The detailed expressions are concluded in (2.6). We denote $L_{\boldsymbol{\alpha}}, L_{\boldsymbol{\beta}}$ as the first-order derivatives of L with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Similarly, we denote $L_{\boldsymbol{\alpha\beta}}, L_{\boldsymbol{\alpha\beta}}, L_{\boldsymbol{\beta\beta}}$ as the second-order derivatives of L.

PART 1 − step 1: In this step, we prove that $\lim_{n_s \to \infty} n_s^{-1} \text{cov}(\dot{L}(\boldsymbol{\theta}), \dot{L}(\boldsymbol{\theta})) = \Sigma_{1s}$ and that the limit is a matrix with finite values; i.e., we need to prove that $\dot{L}(\boldsymbol{\theta})$ meets the requirements of Lemma 2 in [46]. For $1 \leq j_1, j_2 \leq p$, the derivative of the objective function can be expressed as $\dot{L}(\boldsymbol{\theta}) = \text{vec}(L_{\boldsymbol{\alpha}}, L_{\boldsymbol{\beta}}) \in \mathbb{R}^{p^2+pq}$, where $L_{\boldsymbol{\alpha}} = \{L_{j_1 j_2}^{\boldsymbol{\alpha}}\}$ and

$$L_{j_1 j_2}^{\boldsymbol{\alpha}} = 2\widetilde{\mathcal{E}}^\top M_{j_1 j_2} \widetilde{\mathcal{E}} + 2\widetilde{\mathcal{E}}^\top V_{j_1 j_2},$$
$$L_{\boldsymbol{\beta}} = 2H_{\boldsymbol{\beta}}^\top H^* \widetilde{\mathcal{E}}.$$

The detailed notations are provided in section 2.3, where $\mathcal{E} = S\mathcal{Y} - \widetilde{\mathbb{X}}\boldsymbol{\beta}, \widetilde{\mathcal{E}} = (\widetilde{\mathcal{E}}_1^\top, \widetilde{\mathcal{E}}_2^\top, \cdots, \widetilde{\mathcal{E}}_p^\top)^\top = \left(\Sigma_e^{-1/2} \otimes I_N\right)\mathcal{E}$, $\mathcal{Q}_{j_1j_2}^{\boldsymbol{\alpha}} = \left(I_{j_1j_2}\Omega_e\boldsymbol{\alpha}^\top + \boldsymbol{\alpha}\Omega_e I_{j_2j_1}\right) \otimes (W^\top W)$, $S_{j_1j_2}^{\boldsymbol{\alpha}} = -I_{j_2j_1} \otimes W, \widetilde{S}_{j_1j_2}^{\boldsymbol{\alpha}} = -\left(\Omega_e I_{j_2j_1}\right) \otimes W$, $\Omega_{j_1j_2}^{\boldsymbol{\alpha}} = -\left(I_{j_1j_2}\Omega_e\right) \otimes W^\top - \left(\Omega_e I_{j_2j_1}\right) \otimes W + \mathcal{Q}_{j_1j_2}^{\boldsymbol{\alpha}}, m_{j_1j_2}^{\boldsymbol{\alpha}} = -m^2\,\mathrm{diag}\left(\mathcal{Q}_{j_1j_2}^{\boldsymbol{\alpha}}\right)$. Then, the form of the terms in $\dot{L}(\boldsymbol{\theta})$ (i.e., $L_{j_1j_2}^{\boldsymbol{\alpha}}$, $L_{\boldsymbol{\beta}}$) meet the requirements of Lemma 2 in [46]. As a result, we have $\lim_{n_s\to\infty} n_s^{-1}\mathrm{cov}(\dot{L}(\boldsymbol{\theta}), \dot{L}(\boldsymbol{\theta})) = \Sigma_{1s}$ and the detailed expressions are represented in (2.6) to (2.8).

PART 1–step 2: In this step, we prove that $n_s^{-1}\ddot{L}(\boldsymbol{\theta}) \to_p \Sigma_{2s}$ by the law of large numbers. The second-order derivative of the objective function can be expressed as

$$\ddot{L}(\boldsymbol{\theta}) = \begin{pmatrix} L_{\boldsymbol{\alpha}\boldsymbol{\alpha}} & L_{\boldsymbol{\alpha}\boldsymbol{\beta}} \\ L_{\boldsymbol{\alpha}\boldsymbol{\beta}}^\top & L_{\boldsymbol{\beta}\boldsymbol{\beta}} \end{pmatrix},$$

where $i_1 = (j_1 - 1)p + j_2$, $i_2 = (k_1 - 1)p + k_2$, $L_{\boldsymbol{\alpha}\boldsymbol{\alpha}} = \{L_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(i_1,i_2)}\} \in \mathbb{R}^{p^2\times p^2}$, $L_{\boldsymbol{\alpha}\boldsymbol{\beta}} = \{L_{\boldsymbol{\alpha}\boldsymbol{\beta}}^{(i_1,\cdot)}\} \in \mathbb{R}^{p^2\times pq}$, $L_{\boldsymbol{\beta}\boldsymbol{\beta}} \in \mathbb{R}^{pq\times pq}$. Moreover, $L_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(i_1,i_2)} = 2H_{j_1j_2}^{\boldsymbol{\alpha}\top}H_{k_1k_2}^{\boldsymbol{\alpha}} + 2H^\top H_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}}, L_{\boldsymbol{\alpha}\boldsymbol{\beta}}^{(i_1,\cdot)} = 2H_{j_1j_2}^{\boldsymbol{\alpha}\top}H_{\boldsymbol{\beta}} + H^\top H_{j_1j_2\boldsymbol{\beta}}^{\boldsymbol{\alpha}}$, $L_{\boldsymbol{\beta}\boldsymbol{\beta}} = 2H_{\boldsymbol{\beta}}^\top H_{\boldsymbol{\beta}} + H_{\boldsymbol{\beta}\boldsymbol{\beta}}$. For the details, we have $H_{\boldsymbol{\beta}\boldsymbol{\beta}} = 0$ and

$$\begin{aligned} H_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}} &= m_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}} G\{\widetilde{S}^\top(S\mathcal{Y} - \widetilde{\mathbb{X}}\boldsymbol{\beta})\} + m_{j_1j_2}^{\boldsymbol{\alpha}} G\{\widetilde{S}^\top(S\mathcal{Y} - \widetilde{\mathbb{X}}\boldsymbol{\beta})\}_{k_1k_2}^{\boldsymbol{\alpha}} \\ &\quad + m_{k_1k_2}^{\boldsymbol{\alpha}} G\{\widetilde{S}^\top(S\mathcal{Y} - \widetilde{\mathbb{X}}\boldsymbol{\beta})\}_{j_1j_2}^{\boldsymbol{\alpha}} + mG\{\widetilde{S}^\top(S\mathcal{Y} - \widetilde{\mathbb{X}}\boldsymbol{\beta})\}_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}}, \end{aligned}$$

$$\begin{aligned} H_{j_1j_2\boldsymbol{\beta}}^{\boldsymbol{\alpha}} &= \{m_{j_1j_2}^{\boldsymbol{\alpha}} G\widetilde{S}^\top(S\mathcal{Y} - \widetilde{\mathbb{X}}\boldsymbol{\beta})\}_{\boldsymbol{\beta}} + m\{G\widetilde{S}^\top(S\mathcal{Y} - \widetilde{\mathbb{X}}\boldsymbol{\beta})\}_{\boldsymbol{\alpha}_{j_1j_2}\boldsymbol{\beta}} \\ &= m_{j_1j_2}^{\boldsymbol{\alpha}} G\widetilde{S}^\top(-\widetilde{\mathbb{X}}) + m\{-G(I_{j_1j_2} \otimes W^\top)(\Omega_e \otimes I_N)(-\widetilde{\mathbb{X}})\}, \end{aligned}$$

where $\mathcal{Q}_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}} = (I_{j_1j_2}\Omega_e I_{k_2k_1} + I_{k_1k_2}\Omega_e I_{j_2j_1}) \otimes \left(W^\top W\right)$, $m_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}} = 2m^3\,\mathrm{diag}(\mathcal{Q}_{j_1j_2}^{\boldsymbol{\alpha}})\mathrm{diag}(\mathcal{Q}_{k_1k_2}^{\boldsymbol{\alpha}}) - m^2\mathcal{Q}_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}}$, $\Omega_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}} = \mathcal{Q}_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}}$. Now, we express the mean and variance of each component of $\lim_{n_s\to\infty} n_s^{-1}\ddot{L}(\boldsymbol{\theta})$ (i.e., $L_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$, $L_{\boldsymbol{\alpha}\boldsymbol{\beta}}$, $L_{\boldsymbol{\beta}\boldsymbol{\beta}}$).

Thereafter, we provide the mean of $L_{\boldsymbol{\alpha}\boldsymbol{\alpha}} \in \mathbb{R}^{p^2\times p^2}$. Consider that $L_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(i_1,i_2)}$ represents the second-order derivative $\partial^2 L(\boldsymbol{\theta})/\partial\boldsymbol{\alpha}_{j_1j_2}\partial\boldsymbol{\alpha}_{k_1k_2}$. Then, we have

$$E(L_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(i_1,i_2)}) = 2E(H_{j_1j_2}^{\boldsymbol{\alpha}\top}H_{k_1k_2}^{\boldsymbol{\alpha}}) + 2E(H^\top H_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}}),$$

where $H^\top H_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}} = \mathbb{I}_1 + \mathbb{I}_2 + \mathbb{I}_3 + \mathbb{I}_4$, in which $\mathbb{I}_1 = \mathcal{E}^\top\widetilde{S}mGm_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}}G\widetilde{S}^\top\mathcal{E}$, $\mathbb{I}_2 = \mathcal{E}^\top\widetilde{S}mGm_{j_1j_2}^{\boldsymbol{\alpha}}G\left(\widetilde{S}_{k_1k_2}^{\boldsymbol{\alpha}\top}\mathcal{E} + \widetilde{S}^\top S_{k_1k_2}^{\boldsymbol{\alpha}}\mathcal{Y}\right)$, $\mathbb{I}_3 = \mathcal{E}^\top\widetilde{S}mGm_{k_1k_2}^{\boldsymbol{\alpha}}G(\widetilde{S}_{j_1j_2}^{\boldsymbol{\alpha}\top}\mathcal{E} + \widetilde{S}^\top S_{j_1j_2}^{\boldsymbol{\alpha}}\mathcal{Y})$, and $\mathbb{I}_4 = \mathcal{E}^\top\widetilde{S}mGmG\Omega_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}}\mathcal{Y}$. Using the property that $E(\widetilde{\mathcal{E}}) = 0, E(\widetilde{\mathcal{E}}^2) = 1, \mathcal{Y} = S^{-1}(\mathcal{E} + \mathbb{X}\boldsymbol{\beta}), \mathrm{tr}(AB) = \mathrm{tr}(BA)$, we obtain $E(\mathbb{I}_1) = \mathrm{tr}(\widetilde{S}mGm_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}}G\widetilde{S}^\top)$, $E(\mathbb{I}_2) = \mathrm{tr}(mGm_{j_1j_2}^{\boldsymbol{\alpha}}GS_{k_1k_2}^{\boldsymbol{\alpha}\top}\widetilde{S}) + \mathrm{tr}(mGm_{j_1j_2}^{\boldsymbol{\alpha}}G\widetilde{S}^\top S_{k_1k_2}^{\boldsymbol{\alpha}})$, $E(\mathbb{I}_3) = \mathrm{tr}\left(mGm_{k_1k_2}^{\boldsymbol{\alpha}}GS_{j_1j_2}^{\boldsymbol{\alpha}\top}\widetilde{S}\right) + \mathrm{tr}\left(mm_{k_1k_2}^{\boldsymbol{\alpha}}\,\widetilde{S}^\top S_{j_1j_2}^{\boldsymbol{\alpha}}\right), E\left(\mathbb{I}_4\right) = \mathrm{tr}(m^2G\mathcal{Q}_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}})$. Thus, we can determine that $E\left(\mathbb{I}_1 + \mathbb{I}_4\right) = -E\left(\mathbb{I}_2 + \mathbb{I}_3\right)$. As a result, we have $E(H^{\boldsymbol{\alpha}\top}H_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}}) = 0$. As $H_{j_1j_2}^{\boldsymbol{\alpha}} = M_{j_1j_2}^*\widetilde{\mathcal{E}} + V_{j_1j_2}^*$, the mean of each element in

$L_{\boldsymbol{\alpha}\boldsymbol{\alpha}}$ can be expressed as $E(L_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(i_1,i_2)}) = 2E(H_{k_1k_2}^{\boldsymbol{\alpha}\top} H_{j_1j_2}^{\boldsymbol{\alpha}}) = 2\,\mathrm{tr}(M_{j_1j_2}^{*\top} M_{k_1k_2}^{*}) + 2V_{k_1k_2}^{*\top} V_{j_1j_2}^{*}$.

For the variance of $L_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(i_1,i_2)}$, as $n_s \to \infty$, we wish to verify that $n_s^{-2}\mathrm{var}(H_{k_1k_2}^{\boldsymbol{\alpha}\top} H_{j_1j_2}^{\boldsymbol{\alpha}}) \to 0$ and $n_s^{-2}\mathrm{var}(H^{\top} H_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}}) \to 0$. We first prove that $n_s^{-2}\mathrm{var}(H_{k_1k_2}^{\boldsymbol{\alpha}\top} H_{j_1j_2}^{\boldsymbol{\alpha}}) \to 0$. We have

$$H_{k_1k_2}^{\boldsymbol{\alpha}\top} H_{j_1j_2}^{\boldsymbol{\alpha}} = \widetilde{\mathcal{E}}^{\top} M_{k_1k_2}^{*\top} M_{j_1j_2}^{*} \widetilde{\mathcal{E}} + \widetilde{\mathcal{E}}^{\top} M_{k_1k_2}^{*\top} V_{j_1j_2}^{*} + V_{k_1k_2}^{*\top} M_{j_1j_2}^{*} \widetilde{\mathcal{E}} + V_{j_1j_2}^{*\top} V_{k_1k_2}^{*}.$$

Further, to prove that $n_s^{-2}\mathrm{var}(H_{k_1k_2}^{\boldsymbol{\alpha}\top} H_{j_1j_2}^{\boldsymbol{\alpha}}) \to 0$, we could follow the requirement of Lemma 2 in [46]. As $n_s = O(N^{1-2\delta+\epsilon})$, in Proposition 1 in Appendix A.2, we obtain that $n_s^{-2}\,\mathrm{tr}(M_{k_1k_2}^{*\top} M_{j_1j_2}^{*} M_{j_1j_2}^{*\top} M_{k_1k_2}^{*}) \to 0$ and $n_s^{-2} V_{j_1j_2}^{*\top} M_{k_1k_2}^{*} M_{k_1k_2}^{*\top} V_{j_1j_2}^{*} \to 0$. Thus, with $n_s \to \infty$, we obtain $n_s^{-2}\mathrm{var}(H_{k_1k_2}^{\boldsymbol{\alpha}\top} H_{j_1j_2}^{\boldsymbol{\alpha}}) \to 0$.

Next, we need to prove that $n_s^{-2}\mathrm{var}(H^{\top} H_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}}) \to 0$. We split $H^{\top} H_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}}$ into the form of quadratic terms and linear terms of $\widetilde{\mathcal{E}}$. Thus,

$$H^{\top} H_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}} = \widetilde{\mathcal{E}}^{\top} T_1 \widetilde{\mathcal{E}} + \widetilde{\mathcal{E}}^{\top} T_2 \widetilde{\mathcal{E}} + \widetilde{\mathcal{E}}^{\top} T_3 \widetilde{\mathcal{E}} + \widetilde{\mathcal{E}}^{\top} O_1 + \widetilde{\mathcal{E}}^{\top} T_4 \widetilde{\mathcal{E}} + \widetilde{\mathcal{E}}^{\top} O_2 + \widetilde{\mathcal{E}}^{\top} T_5 \widetilde{\mathcal{E}} + \widetilde{\mathcal{E}}^{\top} O_3,$$

where

$$\begin{aligned}
T_1 &= (\Sigma_e^{1/2} \otimes I_N)\widetilde{S} mGm_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}} G\widetilde{S}^{\top}(\Sigma_e^{1/2} \otimes I_N),\\
T_2 &= (\Sigma_e^{1/2} \otimes I_N)\widetilde{S} mGm_{j_1j_2}^{\boldsymbol{\alpha}} G\widetilde{S}_{k_1k_2}^{\boldsymbol{\alpha}\top}(\Sigma_e^{1/2} \otimes I_N),\\
T_3 &= (\Sigma_e^{1/2} \otimes I_N)\widetilde{S} mGm_{j_1j_2}^{\boldsymbol{\alpha}} G\widetilde{S}^{\top} S_{k_1k_2}^{\boldsymbol{\alpha}} S^{-1}(\Sigma_e^{1/2} \otimes I_N),\\
O_1 &= (\Sigma_e^{1/2} \otimes I_N)\widetilde{S} mGm_{j_1j_2}^{\boldsymbol{\alpha}} G\widetilde{S}^{\top} S_{k_1k_2}^{\boldsymbol{\alpha}} S^{-1}\widetilde{\mathbb{X}}\boldsymbol{\beta},\\
T_4 &= (\Sigma_e^{1/2} \otimes I_N)\widetilde{S} mGm_{k_1k_2}^{\boldsymbol{\alpha}} G\widetilde{S}^{\top} S_{j_1j_2}^{\boldsymbol{\alpha}} S^{-1}(\Sigma_e^{1/2} \otimes I_N)\widetilde{\mathcal{E}},\\
O_2 &= (\Sigma_e^{1/2} \otimes I_N)\widetilde{S} mGm_{k_1k_2}^{\boldsymbol{\alpha}} G\widetilde{S}^{\top} S_{j_1j_2}^{\boldsymbol{\alpha}} S^{-1}(\widetilde{\mathbb{X}}\boldsymbol{\beta}),\\
T_5 &= (\Sigma_e^{1/2} \otimes I_N)\widetilde{S} mGmG\Omega_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}} S^{-1}(\Sigma_e^{1/2} \otimes I_N),\\
O_3 &= (\Sigma_e^{1/2} \otimes I_N)\widetilde{S} mGmG\Omega_{j_1j_2k_1k_2}^{\boldsymbol{\alpha}} S^{-1}(\widetilde{\mathbb{X}}\boldsymbol{\beta}).
\end{aligned}$$

Next, similar to the proof of Proposition 1 in Appendix A.2, we need to prove that $\forall i, j = 1, 2.., 5$,

$$\begin{aligned}
\mathrm{tr}(T_iT_j)/n_s^2 \to 0, \mathrm{tr}(T_iT_j^{\top})/n_s^2 \to 0,\\
(O_1 + O_2 + O_3)^{\top}(O_1 + O_2 + O_3)/n_s^2 \to 0,
\end{aligned}$$

where $|T_1|_e \preceq c_{HH1}(\mathbf{1}_p\mathbf{1}_p^{\top}) \otimes W_{HH_1}$, $|T_2|_e \preceq c_{HH2}(\mathbf{1}_p\mathbf{1}_p^{\top}) \otimes W_{HH2}$, $|T_3|_e \preceq c_{HH3}(\mathbf{1}_p\mathbf{1}_p^{\top}) \otimes W_{HH3}$, $|T_4|_e \preceq c_{HH4}(\mathbf{1}_p\mathbf{1}_p^{\top}) \otimes W_{HH3}$, and $|T_5|_e \preceq c_{HH5}(\mathbf{1}_p\mathbf{1}_p^{\top}) \otimes W_{HH5}$, in which $W_{HH1} = (I_N + W)g\{W^{\top}W + (W^{\top}W)^2\}g(I_N + W)$, $W_{HH2} = (I_N + W)g\{W^{\top}W\}gW$, $W_{HH3} = (I_N + W)g\{W^{\top}W\}g(I_N + W)WW^*$, and $W_{HH5} = (I_N + W)g\{W^{\top}W\}W^*$. Furthermore, $c_{HH1}, c_{HH2}, c_{HH3}, c_{HH4}$, and c_{HH5} are finite positive constants. Similar to the proof of $n_s^{-2}\mathrm{var}(H_{k_1k_2}^{\boldsymbol{\alpha}\top} H_{j_1j_2}^{\boldsymbol{\alpha}}) \to 0$, we have $\forall i, j = 1, 2.., 5$, $\mathrm{tr}(T_iT_j)/n_s^2 \to 0$, $\mathrm{tr}(T_iT_j^{\top})/n_s^2 \to 0$, and we can obtain $|O_1|_e \preceq c_{O1}\left\{(\mathbf{1}_p\mathbf{1}_p^{\top}) \otimes W_{HH3}\right\}|\widetilde{\mathbb{X}}\boldsymbol{\beta}|_e$, $|O_2|_e \preceq c_{O2}\left\{(\mathbf{1}_p\mathbf{1}_p^{\top}) \otimes W_{HH3}\right\}|\widetilde{\mathbb{X}}\boldsymbol{\beta}|_e$,

$|O_3|_e \preceq c_{O3}\left\{(\mathbf{1}_p\mathbf{1}_p^\top)\otimes W_{HH5}\right\}|\widetilde{\mathbb{X}}\boldsymbol{\beta}|_e$, where c_{O1}, c_{O2}, and c_{O3} are finite positive constants. Therefore, define $O_{123}=(O_1+O_2+O_3)$,we have

$$\begin{aligned}O_{123}^\top O_{123}/n_s^2 &\le n_s^{-2}c_{OO}(|\widetilde{\mathbb{X}}\boldsymbol{\beta}|_e)^\top(\mathbf{1}_p\mathbf{1}_p^\top)\otimes(W_{HH3}W_{HH3}^\top)|\widetilde{\mathbb{X}}\boldsymbol{\beta}|_e\\ &\le n_s^{-1}(c_{OO}||\widetilde{\mathbb{X}}\boldsymbol{\beta}||^2)n_s^{-1}\{\lambda_{\max}^2(W_{HH3}W_{HH3}^\top)\},\end{aligned}$$

where c_{OO} is a finite positive constant. Thus, we have $n_s^{-1}(c_{OO}\|\widetilde{\mathbb{X}}\boldsymbol{\beta}\|^2)=O(1)$, $n_s^{-1}\lambda_{\max}(W_{oij})\to 0$. As a result, we have $n_s^{-2}\mathrm{var}(H^\top H^{\boldsymbol{\alpha}}_{j_1j_2k_1k_2})\to 0$. Finally, as a consequence, we obtain $L^{(i_1,i_2)}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\to_p \Sigma^{(i_1,i_2)}_{2\boldsymbol{\alpha}\boldsymbol{\alpha}}$, where $1\le i_1,i_2\le q^2$.

Next, we provide the expression for the mean and variance of $L_{\boldsymbol{\alpha}\boldsymbol{\beta}}=\{L^{(i_1,\cdot)}_{\boldsymbol{\alpha}\boldsymbol{\beta}}\}\in\mathbb{R}^{p^2\times(pq)}$. For $1\le i_1\le p^2$, consider that each $L^{(i_1\cdot)}_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ is the i_1th row of $L_{\boldsymbol{\alpha}\boldsymbol{\beta}}$. Equally, we need to prove that $E(L^{(i_1\cdot)}_{\boldsymbol{\alpha}\boldsymbol{\beta}})=V^{*\top}_{j_1j_2}H_{\boldsymbol{\beta}}$, $\mathrm{var}(L^{(i_1\cdot)}_{\boldsymbol{\alpha}\boldsymbol{\beta}}/n_s)\to 0$. We write $L^{(i_1\cdot)}_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ as

$$\begin{aligned}L^{(i_1\cdot)}_{\boldsymbol{\alpha}\boldsymbol{\beta}} &= 2H^{\boldsymbol{\alpha}\top}_{j_1j_2}H_{\boldsymbol{\beta}}+H^\top H^{\boldsymbol{\alpha}}_{j_1j_2\boldsymbol{\beta}}\\ &= 2\widetilde{\mathcal{E}}^\top O_{\boldsymbol{\alpha}\boldsymbol{\beta}1}+2C_{\boldsymbol{\alpha}\boldsymbol{\beta}1}+\widetilde{\mathcal{E}}^\top O_{\boldsymbol{\alpha}\boldsymbol{\beta}2},\end{aligned}$$

where $O_{\boldsymbol{\alpha}\boldsymbol{\beta}1}=M^{*\top}_{j_1j_2}H_{\boldsymbol{\beta}}$, $C_{\boldsymbol{\alpha}\boldsymbol{\beta}1}=V^{*\top}_{j_1j_2}H_{\boldsymbol{\beta}}$ and $O_{\boldsymbol{\alpha}\boldsymbol{\beta}2}=H^{*\top}H^{\boldsymbol{\alpha}}_{j_1j_2\boldsymbol{\beta}}$. As a result, we have $E(L^{(i_1\cdot)}_{\boldsymbol{\alpha}\boldsymbol{\beta}})=2C_{\boldsymbol{\alpha}\boldsymbol{\beta}1}=2V^{*\top}_{j_1j_2}H_{\boldsymbol{\beta}}$. Next, we write the variance of $L^{(i_1\cdot)}_{\boldsymbol{\alpha}\boldsymbol{\beta}}$ as

$$\begin{aligned}\mathrm{var}(L^{(i_1\cdot)}_{\boldsymbol{\alpha}\boldsymbol{\beta}}/n_s) &= (O_{\boldsymbol{\alpha}\boldsymbol{\beta}1}+O_{\boldsymbol{\alpha}\boldsymbol{\beta}2})^\top(O_{\boldsymbol{\alpha}\boldsymbol{\beta}1}+O_{\boldsymbol{\alpha}\boldsymbol{\beta}2})/n_s^2\\ &\le \frac{1}{n_s}(C_{\boldsymbol{\alpha}\boldsymbol{\beta}}|\widetilde{\mathbb{X}}|_e)\frac{1}{n_s}\lambda_{\max}^2[(\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}1}+\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}2})^\top(\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}1}+\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}2})],\end{aligned}$$

where $\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}1}=M^{*\top}_{j_1j_2}[-mG\widetilde{S}^\top]$ and $\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}2}=H^{*\top}[m^{\boldsymbol{\alpha}}_{j_1j_2}G\widetilde{S}^\top+m\{-G(I_{j_1j_2}\otimes W^\top)(\Omega_e\otimes I_N)\}]$. The upper bounds of $\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}1}$ and $\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}2}$ can be expressed as $|\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}1}|_e\preceq c_{\boldsymbol{\alpha}\boldsymbol{\beta}1}(\mathbf{1}_p\mathbf{1}_p^\top)\otimes W^*G(\mathbf{1}_p\mathbf{1}_p^\top)\otimes(I_N+W)$ and $|\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}2}|_e\preceq c_{\boldsymbol{\alpha}\boldsymbol{\beta}2}(\mathbf{1}_p\mathbf{1}_p^\top)\otimes W^*\{G(\mathbf{1}_p\mathbf{1}_p^\top)\otimes(W^\top W)G(\mathbf{1}_p\mathbf{1}_p^\top)\otimes(I_N+W)+G(\mathbf{1}_p\mathbf{1}_p^\top)\otimes W$, respectively, where $c_{\boldsymbol{\alpha}\boldsymbol{\beta}1}$ and $c_{\boldsymbol{\alpha}\boldsymbol{\beta}2}$ are finite constants. Thus, we have $n_s^{-1}\lambda_{\max}[(\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}1}+\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}2})^\top(\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}1}+\widetilde{O}_{\boldsymbol{\alpha}\boldsymbol{\beta}2})]\to 0$. As a result, we obtain $\mathrm{var}(L^{(i_1\cdot)}_{\boldsymbol{\alpha}\boldsymbol{\beta}}/n_s)\to 0$. Furthermore, according to the law of large numbers, we can determine that $n_s^{-1}L^{(i_1\cdot)}_{\boldsymbol{\alpha}\boldsymbol{\beta}}\to_p n_s^{-1}H_{\boldsymbol{\beta}}^\top V^*_{j_1j_2}=\Sigma^{(i_1\cdot)}_{2\boldsymbol{\alpha}\boldsymbol{\beta}}$.

Finally, we provide the form of the mean and variance of $L_{\boldsymbol{\beta}\boldsymbol{\beta}}$. We have $n_s^{-1}L_{\boldsymbol{\beta}\boldsymbol{\beta}}\to_p 2H_{\boldsymbol{\beta}}^\top H_{\boldsymbol{\beta}}/n_s$ because $H_{\boldsymbol{\beta}}=-mG\widetilde{S}\widetilde{\mathbb{X}}$ contains no random terms. As a result, we can write the mean item as $E(L_{\boldsymbol{\beta}\boldsymbol{\beta}})=2H_{\boldsymbol{\beta}}^\top H_{\boldsymbol{\beta}}$. Moreover, the variance item is represented by $\mathrm{var}(n_s^{-1}L_{\boldsymbol{\beta}\boldsymbol{\beta}})=0$. As a result, we use the law of large numbers for $L_{\boldsymbol{\alpha}\boldsymbol{\alpha}}, L_{\boldsymbol{\alpha}\boldsymbol{\beta}}, L_{\boldsymbol{\beta}\boldsymbol{\beta}}$. Then, we obtain $n_s^{-1}\ddot{L}(\boldsymbol{\theta})\to_p\Sigma_{2s}^L$.

PART 2: Asymptotic Normality. It has been proven that $\hat{\boldsymbol{\theta}}$ is $\sqrt{n_s}$-consistent. Then, we need to prove the asymptotic normality of $\hat{\boldsymbol{\theta}}$. By applying Taylor expansion for $\dot{L}(\boldsymbol{\theta})$ on the point of the true parameter $\boldsymbol{\theta}$, we obtain

$$\sqrt{n_s}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})=\{n_s^{-1}\ddot{L}(\boldsymbol{\theta}^*)\}^{-1}\{n_s^{-1/2}\dot{L}(\boldsymbol{\theta})\},$$

where $\boldsymbol{\theta}^*$ is between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$. We have proven that $n_s^{-1}\ddot{L}(\boldsymbol{\theta}^*) \to_p \Sigma_{2s}$ in **PART 1 – step 2**. Next, we prove that $n_s^{-1}\dot{L}(\boldsymbol{\theta}) \to_d N(\mathbf{0}_{p^2+pq}, \Sigma_{1s})$. Equally, it suffices to prove that, for any $u = (u_{\boldsymbol{\alpha}}^\top, u_{\boldsymbol{\beta}}^\top)^\top \in \mathbb{R}^{p^2+pq}$, we have $n_s^{-1/2}u^\top\dot{L}(\boldsymbol{\theta}) \to_d N(\mathbf{0}_{p^2+pq}, u^\top\Sigma_{1s}u)$, where $u_{\boldsymbol{\alpha}} = (u_{\boldsymbol{\alpha},1}, \cdots, u_{\boldsymbol{\alpha},p^2})^\top \in \mathbb{R}^{p^2}$ and $u_{\boldsymbol{\beta}} = (u_{\boldsymbol{\beta},1}, \cdots, u_{\boldsymbol{\beta},pq})^\top \in \mathbb{R}^{pq}$. According to [46], we subsequently obtain

$$u^\top\dot{L}(\boldsymbol{\theta}) = \widetilde{\mathcal{E}}^\top M_{1u}\widetilde{\mathcal{E}} + \widetilde{\mathcal{E}}^\top M_{2u}(\widetilde{\mathbb{X}}\boldsymbol{\beta}) + \widetilde{\mathcal{E}}^\top M_{3u}\left(\widetilde{\mathbb{X}}u_{\boldsymbol{\beta}}\right),$$

where $M_{1u} = 2\sum_{j_1,j_2=1}^{p} u_{\boldsymbol{\alpha},(j_1-1)p+j_2}M_{j_1j_2}$, $M_{2u} = 2\sum_{j_1,j_2=1}^{p} u_{\boldsymbol{\alpha},(j_1-1)p+j_2}H^{*\top}(mG\widetilde{S}^\top S_{j_1j_2}^{\boldsymbol{\alpha}}S^{-1})$, $M_{3u} = -2H^{*\top}m\widetilde{S}^\top$.

Thus, with condition (C4), for arbitrary $\boldsymbol{\beta}_r \in \mathbb{R}^{pq}$ and $R_g \in \mathbb{R}^{n_sp\times n_sp}$, it can be verified that $|n_s^{-1}\left(\widetilde{\mathbb{X}}_g\boldsymbol{\beta}_r\right)^\top R_g\left(\widetilde{\mathbb{X}}_g\boldsymbol{\beta}_r\right)| \le n_s^{-1}c_{\boldsymbol{\beta}}\operatorname{tr}(R_g)$. Similar to the proof in Appendix A.2 and **PART 1 – step 2**, take the upper bound of each linear and quadric term coefficients of $\widetilde{\mathcal{E}}$; it can be verified that

$$n_s^{-2}\operatorname{tr}\left(|M_{1u}|_e\,|M_{1u}|_e^\top\,|M_{1u}|_e\,|M_{1u}|_e^\top\right) \to 0$$
$$n_s^{-1}\lambda_{\max}^2\left(|M_{2u}|_e\,|M_{2u}|_e^\top\right) \to 0, \quad n_s^{-1}\lambda_{\max}^2\left(|M_{3u}|_e\,|M_{3u}|_e^\top\right) \to 0.$$

As a result, according to Lemma 4 in [46], as $n_s \to \infty$, we have $n_s^{-1/2}\dot{L}(\boldsymbol{\theta}) \to_d N(\mathbf{0}_{p^2+pq}, \Sigma_{1s})$. Finally, by combining **PART 1** and **PART 2**, we complete the proof. □

References

[1] Anselin, L. (2013). *Spatial econometrics: methods and models* **4**. Springer Science & Business Media.

[2] Anselin, L. and Smirnov, O. (2001). Fast Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models. *Computational Statistics & Data Analysis* **35** 301–319. MR1819042

[3] Barry, R. and Pace, R. (1999). A Monte Carlo Estimator of the Log Determinant of Large Sparse Matrices. *Linear Algebra and Its Applications* **289** 41–54. MR1670972

[4] Bhattacharyya, S. and Bickel, P. J. (2015). Subsampling bootstrap of count features of networks. *The Annals of Statistics* **43** 2384–2411. MR3405598

[5] Bruggeman, J. (2013). *Social networks: An introduction*. Routledge. MR3203208

[6] Chandrasekaran, V. and Jordan, M. I. (2013). Computational and statistical tradeoffs via convex relaxation. *Proceedings of the National Academy of Sciences* **110** E1181–E1190. MR3047651

[7] Chen, X., Chen, Y. and Xiao, P. (2013). The impact of sampling and network topology on the estimation of social intercorrelations. *Journal of Marketing Research* **50** 95–110.

[8] Chen, Y., Gel, Y. R., Lyubchich, V. and Nezafati, K. (2019). Snowboot: bootstrap methods for network inference. *arXiv preprint arXiv:1902.09029*.
[9] Christensen, K., Podolskij, M., Thamrongrat, N. and Veliyev, B. (2017). Inference from high-frequency data: A subsampling approach. *Journal of Econometrics* **197** 245–272. MR3614200
[10] Clauset, A., Shalizi, C. R. and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review* **51** 661–703. MR2563829
[11] Cohen-Cole, E., Liu, X. and Zenou, Y. (2018). Multivariate choices and identification of social interactions. *Journal of Applied Econometrics* **33** 165–178. MR3782756
[12] Dhillon, P., Lu, Y., Foster, D. P. and Ungar, L. (2013). New subsampling algorithms for fast least squares regression. In *Advances in Neural Information Processing Systems* 360–368.
[13] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96** 1348–1360. MR1946581
[14] Fiacco, A. V. and McCormick, G. P. (1990). *Nonlinear programming: sequential unconstrained minimization techniques*. SIAM. MR1058438
[15] Gel, Y. R., Lyubchich, V. and Ramirez, L. L. R. (2017). Bootstrap quantification of estimation uncertainties in network degree distributions. *Scientific Reports* **7** 1–12.
[16] Goodman, L. A. (1961). Snowball sampling. *The Annals of Mathematical Statistics* 148–170. MR0124140
[17] Hall, P., Jing, B.-Y. and Lahiri, S. N. (1998). On the sampling window method for long-range dependent data. *Statistica Sinica* 1189–1204. MR1666245
[18] Handcock, M. S. and Gile, K. J. (2010). Modeling social networks from sampled data. *The Annals of Applied Statistics* **4** 5. MR2758082
[19] Holland, P. W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical association* **76** 33–50. MR0608176
[20] Huang, D., Lan, W., Zhang, H. and Wang, H. (2019). Least squares estimation of spatial autoregressive models for large-scale social networks. *Electronic Journal of Statistics* **13** 1135–1165. MR3935846
[21] Huang, D., Wang, F., Zhu, X. and Wang, H. (2020). Two-mode network autoregressive model for large-scale networks. *Journal of Econometrics* **216** 203–219. MR4077390
[22] Leenders, R. T. A. (2002). Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks* **24** 21–47.
[23] LeSage, J. and Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC. MR2730939
[24] Li, R., Lin, D. K. and Li, B. (2013). Statistical inference in massive data sets. *Applied Stochastic Models in Business and Industry* **29** 399–409. MR3117826
[25] Ma, P., Mahoney, M. W. and Yu, B. (2015). A statistical perspective

on algorithmic leveraging. *The Journal of Machine Learning Research* **16** 861–911. MR3361306

[26] Newman, M. E., Barabási, A.-L. E. and Watts, D. J. (2006). *The structure and dynamics of networks*. Princeton University Press. MR2352222

[27] Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association* **96** 1077–1087. MR1947255

[28] Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling*. Springer Science & Business Media. MR1707286

[29] Rustagi, J. S. (2014). *Optimization Techniques in Statistics*. Elsevier. MR1274047

[30] Shalizi, C. R. and Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *Annals of Statistics* **41** 508. MR3099112

[31] Shumway, R. H. and Stoffer, D. S. (2017). *Time series analysis and its applications: with R examples*. Springer. MR3642322

[32] Stuti, K. and Srivastava, A. (2016). Performance Analysis and Comparison of Sampling Algorithms in Online Social Network. *International Journal of Computer Applications* **133** 30–35.

[33] Sun, Z. and Wang, H. (2021). Network imputation for a spatial autoregression model with incomplete data. *Statistica Sinica, in press* 10.5705/ss.202017.0366. MR4257535

[34] Thompson, S. K. and Frank, O. (2000). Model-based estimation with link-tracing sampling designs. *Survey Methodology* **26** 87–98.

[35] Thompson, M. E., Ramirez Ramirez, L. L., Lyubchich, V. and Gel, Y. R. (2016). Using the bootstrap for statistical inference on random graphs. *Canadian Journal of Statistics* **44** 3–24. MR3474218

[36] Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science* **47**. Cambridge University Press. MR3837109

[37] Wang, H. (2019). More Efficient Estimation for Logistic Regression with Optimal Subsamples. *Journal of Machine Learning Research* **20** 1–59. MR4002886

[38] Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association* **82** 8–19. MR0883333

[39] Wang, H., Yang, M. and Stufken, J. (2019). Information-based optimal subdata selection for big data linear regression. *Journal of the American Statistical Association* **114** 393–405. MR3941263

[40] Wang, H., Zhu, R. and Ma, P. (2018). Optimal subsampling for large sample logistic regression. *Journal of the American Statistical Association* **113** 829–844. MR3832230

[41] Wang, S., Liu, B., Zhou, J. and Li, G. (2019). Unbiased Sampling Method Analysis on Online Social Network. In *3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019)*. Atlantis Press.

[42] Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods*

and applications **8**. Cambridge University Press.

[43] Yang, K. and Lee, L.-f. (2017). Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. *Journal of Econometrics* **196** 196–214. MR3572822

[44] Yao, Y. and Wang, H. (2019). Optimal subsampling for softmax regression. *Statistical Papers* **60** 235–249. MR3969047

[45] Zhou, J., Tu, Y., Chen, Y. and Wang, H. (2017). Estimating spatial autocorrelation with sampled network data. *Journal of Business & Economic Statistics* **35** 130–138. MR3591541

[46] Zhu, X., Huang, D., Pan, R. and Wang, H. (2020). Multivariate spatial autoregressive model for large scale social networks. *Journal of Econometrics* **215** 591–606. MR4076946