Annals of the Institute of Statistical Mathematics (2022) 74:1043–1065
https://doi.org/10.1007/s10463-022-00822-w

A blockwise network autoregressive model with application for fraud detection

Bofei Xiao[1] · Bo Lei[1] · Wei Lan[1] · Bin Guo[1]

Received: 18 June 2021 / Revised: 3 December 2021 / Accepted: 7 March 2022 /
Published online: 4 May 2022

Abstract

This paper proposes a blockwise network autoregressive (BWNAR) model by grouping nodes in the network into nonoverlapping blocks to adapt networks with blockwise structures. Before modeling, we employ the pseudo likelihood ratio criterion (pseudo-LR) together with the standard spectral clustering approach and a binary segmentation method developed by Ma et al. (*Journal of Machine Learning Research*, **22**, 1–63, 2021) to estimate the number of blocks and their memberships, respectively. Then, we acquire the consistency and asymptotic normality of the estimator of influence parameters by the quasi-maximum likelihood estimation method without imposing any distribution assumptions. In addition, a novel likelihood ratio test statistic is proposed to verify the heterogeneity of the influencing parameters. The performance and usefulness of the model are assessed through simulations and an empirical example of the detection of fraud in financial transactions, respectively.

Keywords Blockwise network autoregressive model · Blockwise structure · Community detection · Likelihood ratio test · Quasi-maximum likelihood estimation

1 Introduction

The network autoregressive (NAR) model reflects the network interaction effect through the dependence between nodes to effectively solve complex network problems (see, e.g., Wang et al. 2012; Kass-Hout and Alhinnawi 2013). In recent years, due to the diversification of network data, the model has been extended by a series of academic researchers for improving its practicability and applicability (see, e.g., Moscone et al. 2017; Huang et al. 2020; Zhu et al. 2020; Zou et al. 2021). In addition, an increasing number of fields are using data possessing

✉ Bin Guo
guobin@swufe.edu.cn

[1] School of Statistics and Center of Statistical Research, Southwestern University of Finance and Economics, 555 LiuTai Road, Wenjiang, Chengdu 611130, Sichuan, China

network structure, and the model has gained great popularity in various fields (see, e.g., Lin and Weinberg 2014; Fracassi 2017; Chen et al. 2018; Cohen–Cole et al. 2018).

To explore the influence effect between the network nodes, a large-scale network is assumed with n nodes. The adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ represents the network structure, where $a_{ij} = 1$ if node i and node j are connected and $a_{ij} = 0$ otherwise. For completeness, we also define $a_{ii} = 0$ for $i = 1, \cdots, n$. Moreover, let y_i be the response collected from node i. Then, for illustration purpose, we introduce the pure NAR model as

$$y_i = \lambda \sum_{j=1}^{n} w_{ij} y_j + \varepsilon_i, \tag{1}$$

where $\lambda > 0$ is the influence parameter, $w_{ij} = a_{ij} / \sum_{j=1}^{n} a_{ij}$ and ε_i is the random error for $i, j = 1, \cdots, n$. Meanwhile, its matrix form is

$$Y = \lambda WY + \mathcal{E}, \tag{2}$$

where $Y = (y_1, \cdots, y_n)^\top \in \mathbb{R}^{n\times 1}$, $W = (w_{ij}) \in \mathbb{R}^{n\times n}$ and $\mathcal{E} = (\varepsilon_1, \cdots, \varepsilon_n)^\top \in \mathbb{R}^{n\times 1}$. Obviously, the pure NAR model takes the connected relationships between nodes into consideration and $\lambda w_{ij} y_j$ represents the influence of node j on node i in the network. Specifically, λ represents the common influence coefficient and w_{ij} reflects the strength of the connection between nodes j and i. Thus, nodes that are closer to one node are more influential than those that are farther away.

The NAR model is widely used, but it has the limitation of assuming that every node has the same influence in the network. In model (1), all nodes share a common influence parameter λ. However, in practice, different nodes may have different influences on a node in the network (see, e.g., Zhu et al. 2020; Zou et al. 2021). That is, λ becomes λ_j in the model (1) for $j = 1, \cdots, n$, which increases the number of influence parameters of the model from one to n. When n is large enough, it is inestimable, hence some structures are needed to impose on λ_js. A natural choice is the popularly assumed blockwise structure of the large network (see, e.g., Durlauf and Young 2001; Blume et al. 2015; Moscone et al. 2017). Specifically, the blockwise structure of the network refers to nodes partitioned into nonoverlapping blocks, where nodes have higher influence for others in the same block and have little or no effect on the nodes of other blocks. This is reasonable since the nodes of the network can always be grouped according to their attributes. We next provide some examples to reflect this fact. Individuals submitting similar social information when applying for a loan can be grouped together because they are more likely to know each other; individuals from one company can be grouped together because they have similar working experience; companies belonging to the same sector of economic activity and located within the same geographic area can be grouped together since they face similar opportunities and constraints. In other words, these examples represent many research-meaningful networks possessing blockwise structures in real life. This structure also exists in other networks, such as institutions (see, e.g., Moscone et al. 2017), neuroscience (see, e.g., Luo 2014) and biology (see, e.g., Hao et al. 2012). Therefore, it is particularly significant to extend model (1) to adapt a network with blockwise structure.

Motivated by this challenge, we propose a blockwise network autoregressive (BWNAR) model in (8) by grouping the network nodes into nonoverlapping k blocks. Corresponding to model (2), its pure form is

$$Y = W\Lambda Y + \mathcal{E}, \tag{3}$$

where $\Lambda =$ diag $\{\lambda_{g_1}, \cdots, \lambda_{g_n}\}$ is the influence parameter matrix with $g_i \in \{1, \cdots, k\}$ denoting the block label of node i for $i = 1, \cdots, n$ and the detailed definition of other terms is listed in Sect. 2.2. Hence, we noticeably observe that each node in the same block shares an equal λ_r for $r = 1, \cdots, k$, and different blocks are endowed with dissimilar network influence parameters. In addition, the BWNAR model becomes a NAR model when n nodes belong to a block. For the BWNAR model, we use a two-step method to estimate its parameters. First, we determine the number of blocks and their memberships using the pseudo-LR criterion together with the standard spectral clustering approach and a binary segmentation method, respectively, since Ma et al. (2021) has proven its consistency. Second, without imposing any distribution assumption on noise term, we adopt the quasi-maximum likelihood estimator (QMLE) to estimate the parameters of blocks and establish its asymptotic properties. Moreover, we provide a test statistic to assess the heterogeneity of the influence parameters λ_r of different blocks and demonstrate its validity.

Our contribution is twofold. First, a blockwise network autoregressive model is proposed and particularly exploited for networks with blockwise structure. In this model, different network influence coefficients are allocated for different blocks, and nodes belonging to the same block utilize the common network influence coefficient. Second, we construct a novel test statistic based on the likelihood ratio and prove its validity to assess the heterogeneity of the influence parameters λ_r.

The rest of the paper is organized as follows. In Sect. 2, we firstly employ the pseudo-LR criterion together with the standard spectral clustering approach and a binary segmentation method to determine the number of blocks and their memberships. Then, the BWNAR model is defined and the theoretical properties of QMLE are presented in this section. Finally, we give a test statistic based on the likelihood ratio to assess the heterogeneity of influence parameters. A Monte Carlo simulation and an empirical example of the detection of fraud in financial transactions are given in Sects. 3 and 4. Section 5 concludes this paper. The Supplementary Material contains the theoretical proofs.

2 Methodology

2.1 Community detection

Before introducing the BWNAR model, the number of blocks and their memberships need to be determined in advance. However, in practice, prior information on the real number of blocks and their memberships is usually unavailable. Hence, accurately estimating k and their memberships from network is of crucial importance. In this article, we employ the recent theoretical framework from Ma et al. (2021) to consistently estimate the number of blocks and their memberships in a network when the network's node degrees follow a power-law distribution (see, e.g., Kolaczyk 2009). To avoid causing any confusion, we next adopt the "community" term to replace "block" in this section.

To introduce the procedure of Ma et al. (2021), we first explain the degree-corrected stochastic block model (DCSBM) proposed by Karrer and Newman (2011). Let $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ be the adjacency matrix generated by a DCSBM with actual k communities. Specifically, let $C = (C_1, \cdots, C_k) \in \mathbb{R}^k$ be the community label and $Z = (Z_{ir}) \in \mathbb{R}^{n\times k}$ be a matrix reflecting the true community memberships of each node, where $Z_{ir} = 1$ if node i belongs to C_r for $r = 1, \cdots, k$, and $Z_{ir} = 0$ otherwise. Define $B = (B_{r_1 r_2}) \in \mathbb{R}^{k\times k}$ as symmetric block probability matrix where each entry $B_{r_1 r_2} \in (0, 1]$ means the probability of connection between communities r_1 and r_2 for $r_1, r_2 = 1, \cdots, k$. Let $\Theta = \mathrm{diag}\{\theta_1, \theta_2, \cdots, \theta_n\}$ be nonnegative degree parameters. Then, the probability matrix of edges $P = (P_{ij}) \in \mathbb{R}^{n\times n}$ is

$$P = \mathrm{E}(A) = \Theta Z B Z^\top \Theta^\top, \tag{4}$$

where $P_{ij} = \theta_i \theta_j B_{r_1 r_2}$ represents the probability of edge between nodes i and j belonging to communities C_{r_1} and C_{r_2}, respectively. That is, the edges between nodes i and j are chosen independently with probability depending on the communities to which nodes i and j belong.

Let $d_i = \sum_{j=1}^n a_{ij}$ denote the degree of node i and $\bar{d} = \sum_{i=1}^n d_i/n$ be the average degree. Denote $D = \mathrm{diag}(d_1 + \bar{d}, \ldots, d_n + \bar{d})$ as the diagonal matrix with diagonal elements $d_i + \bar{d}$ for $i = 1, \cdots, n$. As suggested in Ma et al. (2021), define the regularized graph Laplacian matrix $L = D^{-1/2} A D^{-1/2}$. Based on a standard spectral clustering approach of the first r eigenvectors of L and a binary segmentation (see, e.g., Wang and Su 2021) technique on its first $r+1$ eigenvectors, we first obtain the estimators of membership matrices $(\hat{Z}_r, \hat{Z}^b_{r+1})$ for each $r = 1, \cdots, k_{max}$. Here k_{max} denotes the pre-specified largest community number and the superscript "b" denotes the membership matrix that is estimated by binary segmentation. Then, we estimate the true number of communities, k, based on particular pseudo-LR that is to evaluate the deviance of goodness-of-fit of DCSBMs estimated with r and $r+1$ communities, utilizing the $(\hat{Z}_r, \hat{Z}^b_{r+1})$. Specifically, let

$$L_n(\hat{Z}^b_{r+1}, \hat{Z}_r) = \frac{1}{2} \sum_{i\neq j} \left(\frac{\hat{P}_{ij}(\hat{Z}^b_{r+1})}{\hat{P}_{ij}(\hat{Z}_r)} - 1 \right)^2, \tag{5}$$

and the pseudo-LR $\mathcal{R}(r)$ is proposed as

$$\mathcal{R}(r) = \begin{cases} \frac{L_n(\hat{Z}^b_{r+1}, \hat{Z}_r)}{\eta_n} & r = 1 \\ \frac{L_n(\hat{Z}^b_{r+1}, \hat{Z}_r)}{L_n(\hat{Z}^b_r, \hat{Z}_{r-1})} & r \geq 2 \end{cases}, \tag{6}$$

where $\hat{P}_{ij}(Z)$ is the estimator of P_{ij} for a given membership matrix Z, $\eta_n = c_\eta n^2$ and $c_\eta = 0.05$ as suggested in Ma et al. (2021). The estimated community number $\hat{k} = \min(\hat{r}_1, \hat{r}_2)$, where $\hat{r}_1 = \mathrm{argmin}_{1 \leq r \leq k_{max}} \mathcal{R}(r)$ and $\hat{r}_2$ is the smallest r such that $\mathcal{R}(r) \leq \bar{d}^{-1/2}$. The consistency of the estimated community number $\hat{k}$ and their memberships were established in Ma et al. (2021). Accordingly, in the remainder article, we assume the community number and their memberships are given prior.

2.2 Blockwise network autoregressive model

In this section, we introduce our Blockwise Network Autoregressive (BWNAR) model. We assume the n network nodes can be decomposed into k nonoverlapping blocks $\mathcal{C}_1, \cdots, \mathcal{C}_k$. For each node i, g_i denotes its block label, i.e., $g_i = r$ as long as $i \in \mathcal{C}_r$. $\mathcal{C}_1, \cdots, \mathcal{C}_k$ and $g_1, \cdots, g_n$ are all assumed given in this section. Accordingly, the BWNAR model is

$$y_i = \sum_{j=1}^{n} w_{ij} \lambda_{g_j} y_j + X_i^\top \beta + \varepsilon_i, \tag{7}$$

where $\lambda_{g_j} = \lambda_r$ is the influence parameter of block r for $1 \leq r \leq k$, $X_i = (X_{i1}, \cdots, X_{ip})^\top$ are p-dimensional covariates of node i, and $\beta = (\beta_1, \cdots, \beta_p)^\top \in \mathbb{R}^{p \times 1}$ are unknown influence parameters. Let $Y = (y_1, \cdots, y_n)^\top \in \mathbb{R}^{n \times 1}$ and $X = (x_{il}) \in \mathbb{R}^{n \times p}$ for $i = 1, \cdots, n$ and $l = 1, \cdots, p$ be the response and p-dimensional covariates, respectively. Then,

$$Y = W\Lambda Y + X\beta + \mathcal{E}, \tag{8}$$

where $\Lambda = \mathrm{diag}\{\lambda_{g_1}, \cdots, \lambda_{g_n}\} \in \mathbb{R}^{n \times n}$, and $\mathcal{E} = (\varepsilon_1, \cdots, \varepsilon_n)^\top \in \mathbb{R}^{n \times 1}$ are distributed with mean 0 and covariance $\sigma^2 I_n$, where $I_n \in \mathbb{R}^{n \times n}$ is the identity matrix of dimension n. The estimation of β and $\lambda_1, \cdots, \lambda_k$ for different blocks are given in next section. By utilizing this influence measures, we can identify what kind of blocks possessing higher network interaction effects.

2.3 Quasi-maximum likelihood estimation

Since we do not assume the specific distribution on the disturbance $\mathcal{E}$ in BWNAR model (8), we employ quasi-maximum likelihood estimation (QMLE, Lee 2004) to estimate the parameters in this section. Let $\lambda = (\lambda_1, \cdots, \lambda_k)^\top \in \mathbb{R}^{k \times 1}$, $S(\lambda) = I_n - W\Lambda$ and $\mathcal{E}(\lambda, \beta) = S(\lambda)Y - X\beta$. Then, the normal log-likelihood function of (8) is

$$\ln \ell(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \ln|S(\lambda)| - \frac{1}{2\sigma^2}(S(\lambda)Y - X\beta)^\top (S(\lambda)Y - X\beta), \tag{9}$$

where $\theta = (\beta^\top, \lambda^\top, \sigma^2)^\top \in \mathbb{R}^{p+k+1}$ is the vector form of parameters and its true value is denoted as $\theta_0 = (\beta_0^\top,\ \lambda_0^\top, \sigma_0^2)^\top$. We next adopt the concentrated quasi-likelihood approach by concentrating out β and σ^2. Given λ, the QMLE of β and σ^2 is

$$\begin{aligned}\hat{\beta}(\lambda) =&(X^\top X)^{-1}X^\top S(\lambda)Y,\\ \hat{\sigma}^2(\hat{\beta}(\lambda),\lambda) =&\frac{1}{n}\mathcal{E}^\top(\lambda,\hat{\beta}(\lambda))\mathcal{E}(\lambda,\hat{\beta}(\lambda)) = \frac{1}{n}Y^\top S^\top(\lambda)\mathrm{M}S(\lambda)Y,\end{aligned}$$

where $\mathrm{M} = I_n - X(X^\top X)^{-1}X^\top$. The concentrated log likelihood function of λ is

$$\ln \ell(\lambda) = -\frac{n}{2} - \frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln \hat{\sigma}^2(\hat{\beta}(\lambda),\lambda) + \ln|S(\lambda)|.$$

The quasi-maximum likelihood estimation of λ is given via $\hat{\lambda} = \mathrm{argmax}_\lambda \ln \ell(\lambda)$. Finally, we obtain the QMLE of β and σ^2, which are $\hat{\beta} = \hat{\beta}(\hat{\lambda})$ and $\hat{\sigma}^2 = \hat{\sigma}^2(\hat{\beta}(\hat{\lambda}),\hat{\lambda})$, respectively. In the following, we use generic notation $(\mathrm{g}_{t_1,t_2})_{T_1\times T_2}$ to denote a matrix that has dimensions $T_1 \times T_2$ and whose (t_1,t_2)-th element is g_{t_1,t_2} for $t_1 = 1,\cdots,T_1$ and $t_2 = 1,\cdots,T_2$. Before establishing the asymptotic distribution of $\hat{\theta}$, we first introduce some notations and equations. For $r_1, r_2 = 1,\cdots,k$, the Fisher information matrix of (9) is

$$\mathrm{I}_n(\theta_0) := -\mathrm{E}\left(\frac{1}{n}\frac{\partial^2 \ln \ell(\theta_0)}{\partial\theta\partial\theta^\top}\right) = \begin{pmatrix} (n\sigma_0^2)^{-1}X^\top X & \mathrm{I}_{\beta\lambda,n} & 0_{p\times 1}\\ \mathrm{I}_{\lambda\beta,n} & \mathrm{I}_{\lambda\lambda,n} & \mathrm{I}_{\lambda\sigma^2,n}\\ 0_{1\times p} & \mathrm{I}_{\sigma^2\lambda,n} & (2\sigma_0^4)^{-1}\end{pmatrix}, \tag{10}$$

where

$$\begin{aligned}\mathrm{I}_{\beta\lambda,n} &= (n\sigma_0^2)^{-1}\left\{X^\top W\Lambda_{\lambda_1}S^{-1}(\lambda_0)X\beta_0,\cdots,X^\top W\Lambda_{\lambda_k}S^{-1}(\lambda_0)X\beta_0\right\},\ \mathrm{I}_{\beta\lambda,n} = \mathrm{I}_{\lambda\beta,n}^\top,\\ \mathrm{I}_{\lambda\lambda,n} &= n^{-1}\Bigg(tr\left\{W\Lambda_{\lambda_{r_1}}S^{-1}(\lambda_0)W\Lambda_{\lambda_{r_2}}S^{-1}(\lambda_0)\right\}\\ &\quad + tr\left\{W\Lambda_{\lambda_{r_1}}S^{-1}(\lambda_0)(W\Lambda_{\lambda_{r_2}}S^{-1}(\lambda_0))^\top\right\}\\ &\quad + \sigma_0^{-2}(W\Lambda_{\lambda_{r_1}}S^{-1}(\lambda_0)\mathrm{X}\beta_0)^\top W\Lambda_{\lambda_{r_2}}S^{-1}(\lambda_0)\mathrm{X}\beta_0\Bigg)_{k\times k},\\ \mathrm{I}_{\lambda\sigma^2,n} &= (n\sigma_0^2)^{-1}\left\{tr(W\Lambda_{\lambda_1}S^{-1}(\lambda_0)),\cdots,tr(W\Lambda_{\lambda_k}S^{-1}(\lambda_0))\right\}^\top,\ \mathrm{I}_{\lambda\sigma^2,n} = \mathrm{I}_{\sigma^2\lambda,n}^\top.\end{aligned}$$

Since the random error vector in model (8) is assumed to be distributed with mean 0 and covariance $\sigma^2 I_n$, the third and fourth moments, $\mu^s = \mathrm{E}(\varepsilon_i^s)$ for $s = 3, 4$, are needed and will be involved in the asymptotic distribution of $\hat{\theta}$. Let "$\circ$" be the Hadamard product of matrices, X_l be the l-th column of X and $e_n = (1,\cdots,1)^\top \in \mathbb{R}^n$. Then, the matrix $\Omega_n(\theta_0,\mu^3,\mu^4)$ is set to be

$$\Omega_n(\theta_0, \mu^3, \mu^4) = \begin{pmatrix} 0_{p\times p} & \Omega_{\beta\lambda,n} & 0_{p\times 1} \\ \Omega_{\lambda\beta,n} & \Omega_{\lambda\lambda,n} & \Omega_{\lambda\sigma^2,n} \\ 0_{1\times p} & \Omega_{\sigma^2\lambda,n} & \frac{\mu^4-3\sigma_0^4}{4\sigma_0^8} \end{pmatrix},$$

where

$$\Omega_{\beta\lambda,n} = \frac{\mu^3}{n\sigma_0^4}\left(tr\left[(X_l e_n^\top)\circ\left\{W\Lambda_{\lambda_r} S^{-1}(\lambda_0)\right\}\right]\right)_{p\times k}, \Omega_{\lambda\beta,n} = \Omega_{\beta\lambda,n}^\top,$$

$$\begin{aligned}\Omega_{\lambda\lambda,n} =& \frac{(\mu^4 - 3\sigma_0^4)}{n\sigma_0^4}\left(tr\left[\left\{W\Lambda_{\lambda_{r_1}} S^{-1}(\lambda_0)\right\}\circ\left\{W\Lambda_{\lambda_{r_2}} S^{-1}(\lambda_0)\right\}\right]\right)_{k\times k} \\ &+ \frac{\mu^3}{n\sigma_0^4}\left(tr\left[\left\{W\Lambda_{\lambda_{r_1}} S^{-1}(\lambda_0)X\beta_0 e_n^\top\right\}\circ\left\{W\Lambda_{\lambda_{r_2}} S^{-1}(\lambda_0)\right\}\right]\right)_{k\times k} \\ &+ \frac{\mu^3}{n\sigma_0^4}\left(tr\left[\left\{W\Lambda_{\lambda_{r_2}} S^{-1}(\lambda_0)X\beta_0 e_n^\top\right\}\circ\left\{W\Lambda_{\lambda_{r_1}} S^{-1}(\lambda_0)\right\}\right]\right)_{k\times k},\end{aligned}$$

$$\Omega_{\lambda\sigma^2,n} = \frac{1}{2n\sigma_0^6}\left[\mu^3 e_n^\top W\Lambda_{\lambda_r} S^{-1}(\lambda_0)X\beta_0 + (\mu^4 - 3\sigma_0^4)tr(\Lambda_{\lambda_r} S^{-1}(\lambda_0))\right]_{k\times 1}, \ \Omega_{\sigma^2\lambda,n} = \Omega_{\lambda\sigma^2,n}^\top.$$

Then, the covariance matrix of $\frac{1}{\sqrt{n}}\frac{\partial \ln\ell(\theta_0)}{\partial\theta}$ is

$$\mathrm{cov}\left(\frac{1}{\sqrt{n}}\frac{\partial \ln\ell(\theta_0)}{\partial\theta}\right) = \mathrm{I}_n(\theta_0) + \Omega_n(\theta_0, \mu^3, \mu^4). \tag{11}$$

The asymptotic distribution of $\hat{\theta}$ is given in the following theorem.

Theorem 1 *Under Conditions* $(C1)$–$(C8)$ *in Supplementary Material S.2, as* $n \to \infty$, *we obtain that*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathrm{I}^{-1}(\theta_0) + \mathrm{I}^{-1}(\theta_0)\Omega(\theta_0, \mu^3, \mu^4)\mathrm{I}^{-1}(\theta_0)),$$

where $\mathrm{I}(\theta_0) = \lim_{n\to\infty} \mathrm{I}_n(\theta_0)$ *and* $\Omega(\theta_0, \mu^3, \mu^4) = \lim_{n\to\infty} \Omega_n(\theta_0, \mu^3, \mu^4)$, $\mathrm{I}(\theta_0)$ *and* $\Omega(\theta_0, \mu^3, \mu^4)$ *are positive definite matrices. If* $\epsilon_i^{'}s$ *are normally distributed, then* $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathrm{I}^{-1}(\theta_0))$.

Since both $\mathrm{I}(\theta_0)$ and $\Omega(\theta_0, \mu^3, \mu^4)$ are unknown, we then need to seek consistent estimators to make Theorem 1 available. By $\mathrm{I}(\theta_0) = \lim_{n\to\infty} \mathrm{I}_n(\theta_0)$ and $\Omega(\theta_0, \mu^3, \mu^4) = \lim_{n\to\infty} \Omega_n(\theta_0, \mu^3, \mu^4)$, $\mathrm{I}_n(\hat{\theta})$ and $\Omega_n(\hat{\theta}, \hat{\mu}^3, \hat{\mu}^4)$ can be used consistently as the estimators of $\mathrm{I}(\theta_0)$ and $\Omega(\theta_0, \mu^3, \mu^4)$, respectively, where $\hat{\mu}^s = n^{-1}\sum_{i=1}^n \hat{\varepsilon}_i^s$ for $s = 3, 4$. In reality, for more accurate estimation of the covariance of $\hat{\theta}$ with finite sample, we propose a spatial bootstrap procedure in BWNAR model. Similar procedure was adopted for the inference of spatial autoregressive (SAR) models (see, e.g., Anselin 1988, 1990). Let $\hat{\Lambda}$ and $\hat{\beta}$ be the QMLE of Λ and β, $\hat{\mathcal{E}}$ be the estimated residuals via $\hat{\mathcal{E}} = (I_n - W\hat{\Lambda})Y - X\hat{\beta}$. A bootstrap replication of Y is then constructed from a set of randomly sampled residuals (with replacement) in conjunction with the estimated parameters:

$$Y^* = (I_n - W\hat{\Lambda})^{-1}(X\hat{\beta} + \hat{\mathcal{E}}^*), \tag{12}$$

where $\hat{\mathcal{E}}^*$ is a vector of re-sampled residuals from $\hat{\mathcal{E}}$. The bootstrap resample of $\hat{\theta}$ can be obtained from QMLE by using the re-sampled response Y^*, weight matrix W and covariates X. With a total of Q replications, we can estimate the covariance of $\hat{\theta}$ via $\sum_{q=1}^{Q}(\hat{\theta}^{*q} - \bar{\theta}^*)(\hat{\theta}^{*q} - \bar{\theta}^*)^\top/(Q-1)$, where $\hat{\theta}^{*q}$ is the q-th bootstrap resample of $\hat{\theta}$ and $\bar{\theta}^* = \sum_{q=1}^{Q}\hat{\theta}^{*q}/Q$.

2.4 Test for network coefficients

Given the parameter estimator of θ and its asymptotic property, we next give the hypothesis test to examine the effect of different influence coefficients of blocks. To this end, we consider the following null and alternative hypotheses:

$$\begin{aligned} &\mathrm{H}_0 : \lambda_1 = \cdots = \lambda_k = \lambda_c \\ &\mathrm{H}_1 : \lambda_{r_1} \neq \lambda_{r_2} \;\text{ for some } r_1, r_2 = 1, 2, \cdots, k, \text{ and } r_1 \neq r_2. \end{aligned} \tag{13}$$

Obviously, failure to reject the null hypothesis suggests that the NAR model and its associated estimators and properties can be considered. There are three classic large sample tests (Wald, Lagrange multiplier, and likelihood ratio test) under the QMLE framework. We consider the likelihood ratio test (LR) here. Given $\hat{\theta} = (\hat{\beta}^\top, \hat{\lambda}^\top, \hat{\sigma}^2)^\top$, we obtain the estimated quasi-log-likelihood function $\ln\ell(\hat{\theta}) = \ln\ell(\hat{\beta}^\top, \hat{\lambda}^\top, \hat{\sigma}^2)$. Under the null hypothesis of $\mathrm{H}_0 : \lambda_1 = \cdots = \lambda_k = \lambda_c$, we also obtain the constrained QMLE, $\tilde{\theta} = (\tilde{\beta}^\top, \tilde{\lambda}^\top, \tilde{\sigma}^2)^\top$, and its associated quasi-log-likelihood function $\ln\ell(\tilde{\theta})$. Then, the LR test statistic compares the performance of the constrained and unconstrained specifications based on the likelihood ratio,

$$\mathcal{LR} = -2\left[\ln\ell(\tilde{\theta}) - \ln\ell(\hat{\theta})\right].$$

Before showing the theoretical property of $\mathcal{LR}$, we introduce additional notations and equations as below. We slightly arrange the notation $\theta = \left(\beta^\top, \lambda^\top, \sigma^2\right)^\top$ to $\theta = \left(\sigma^2, \beta^\top, \lambda^\top\right)^\top = \left(\theta_1^\top, \theta_2^\top\right)^\top$, where $\theta_1 = \left(\sigma^2, \beta^\top\right)^\top$ and $\theta_2 = \lambda = \left(\lambda_1, \cdots, \lambda_k\right)^\top$. Let $R = \left(0_{(k-1)\times(p+1)}, R_1\right)$, where

$$R_1 = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -1 & & 0 & 0 & 0 \\ & \vdots & & \ddots & & \vdots & \\ & 0\ 0\ 0 & & \cdots & 0 & 1 & -1 \end{pmatrix}_{(k-1)\times k},$$

$r_c = 0_{(k-1)\times 1} \in \mathbb{R}^{(k-1)\times 1}$. Then, the null hypothesis is equivalent to $\mathrm{H}_0 : R(\theta) = R\theta = r_c$, the asymptotic distribution of $\mathcal{LR}$ is given below.

Theorem 2 *Assume that Conditions* (*C*1)–(*C*8) *in Supplementary Material S.2 hold. Under the null hypothesis* H_0, *the quasi-likelihood ratio test statistic* $\mathcal{LR}$ *is asymptotically* *distributed as* $\sum_{r=1}^{k-1}\lambda_r(\theta_0,\mu^3,\mu^4)\mathcal{X}_r^2(1)$ *as* $n\to\infty$, *where* $\lambda_r(\theta_0,\mu^3,\mu^4)$ *is the r-th largest eigenvalue of the matrix* $\Pi^2=I(\theta_0)^{-1/2}R'(\theta_0)^\top\left[R'(\theta_0)\Sigma^{-1}R'(\theta_0)^\top\right]^{-1}R'(\theta_0)I(\theta_0)^{-1/2}$, $\Sigma=I(\theta_0)+\Omega(\theta_0,\mu^3,\mu^4)$ *and* $\mathcal{X}_r^2(1)$ *are independent Chi-squared random variables with degree of freedom* 1 *for* $r=1,\cdots,(k-1)$. *Furthermore, under the normal assumption of* $\mathcal{E}$, $\mathcal{LR}\xrightarrow{d}\chi^2(k-1)$.

Since $\lambda_r(\theta_0,\mu^3,\mu^4)$ is unknown, we can estimate it by $\lambda_{n,r}(\tilde\theta,\tilde\mu^3,\tilde\mu^4)$, where $\lambda_{n,r}(\tilde\theta,\tilde\mu^3,\tilde\mu^4)$ is the r-th largest eigenvalue of the matrix $\tilde\Pi^2$. Note tlhat $\tilde\Pi^2=(-H_n(\tilde\theta))^{-1/2}R'(\tilde\theta)^\top[R'(\tilde\theta)\tilde\Sigma^{-1}R'(\tilde\theta)^\top]^{-1}R'(\tilde\theta)(-H_n(\tilde\theta))^{-1/2}$, where $H_n(\tilde\theta)=\frac{1}{n}\frac{\partial^2\ln L(\tilde\theta)}{\partial\theta\partial\theta^\top}$, $\tilde\Sigma=\mathrm{I}_n(\tilde\theta)+\Omega_n(\tilde\theta,\tilde\mu^3,\tilde\mu^4)$ and $\tilde\mu^s=n^{-1}\sum_{i=1}^n\tilde\varepsilon_i^s$ for $s=3,4$. In addition, by above equation and Condition (C8), $-H_n(\tilde\theta)$, $\Omega_n(\tilde\theta,\tilde\mu^3,\tilde\mu^4)$ and $R'(\tilde\theta)$ are consistent estimators of $I(\theta_0)$, $\Omega(\theta_0,\mu^3,\mu^4)$ and $R'(\theta_0)$, respectively. However, the convergence of $\mathcal{LR}$ to its asymptotic distribution may be slow, which may cause size distortion in reality. We can also adopt the bootstrap procedure for better approximation of the distribution $\mathcal{LR}$ under the null hypothesis H_0 of (13). Similar to the bootstrap procedure outlined in Sect. 2.3 for covariance estimation, let $\tilde\Lambda$ and $\tilde\beta$ be the restricted QMLE of Λ and β under H_0, $\tilde{\mathcal{E}}$ be the restricted estimated residuals. The q-th bootstrap replication of response Y under H_0, denotes as $\tilde Y^{*q}$ that is generated via (12) with $\hat\Lambda$, $\hat\beta$ and $\hat{\mathcal{E}}$ replaced by $\tilde\Lambda$, $\tilde\beta$ and $\tilde{\mathcal{E}}$. Then the q-th bootstrap resample of $\mathcal{LR}$, denoted as $\mathcal{LR}^{*q}$, can be obtained by using $\tilde Y^{*q}$, W and X. The process is replicated Q times and we can obtain the bootstrap estimate of the null distribution of $\mathcal{LR}$ via the empirical distribution of $\{\mathcal{LR}^{*q}\}_{q=1}^Q$ and the p-value $\sum_{q=1}^Q I\{\mathcal{LR}^{*q}>\mathcal{LR}\}/Q$, where $I\{\cdot\}$ is the indicator function. We reject the null hypothesis if the p-value is smaller than a significant level.

If we reject the null hypothesis of H_0 that all λ_rs for $r=1,\cdots,k$ are equal, we can conduct the following procedure for making pairwise comparisons at the significance level α. First, we sort the λ_rs based on their estimators $\hat\lambda_r$s. For illustration purpose, we assume $\hat\lambda_1>\hat\lambda_2>\cdots>\hat\lambda_k$. We then test the equality of λ_1 versus λ_j for $j=2,\cdots,k$ sequentially, via the test statistic $T_{1j}=n^{1/2}(\hat\lambda_1-\hat\lambda_j)/(e_{1j}^\top W(\hat\theta)e_{1j})^{1/2}$, where $W(\theta_0)=\mathrm{I}^{-1}(\theta_0)+\mathrm{I}^{-1}(\theta_0)\Omega(\theta_0,\mu^3,\mu^4)\mathrm{I}^{-1}(\theta_0)$ is the asymptotic covariance of $\hat\theta$ that defined in Theorem 1, and e_{ij} is a $p+k+1$-dimensional vector that the i-th element is 1, the j-th element is -1 and other elements are 0. Under the null hypothesis of $\lambda_1=\lambda_j$, by Theorem 1, we can show that T_{1j} converges to a standard normal distribution, and we reject the null hypothesis if $T_{1j}>z_{1-\alpha}$, where $z_{1-\alpha}$ represents the α-th upper quantile of a standard normal distribution. Let g_1 be the smallest index that we reject the null hypothesis of $\lambda_1=\lambda_j$. Then, $g_1=\mathrm{argmin}_j(T_{1j}>z_{1-\alpha})$ and we know that $\lambda_1,\cdots,\lambda_{g_1-1}$ are all equal. Subsequently, we focus on testing λ_{g_1} versus λ_j for $j=g_1+1,\cdots,k$ sequentially via the same procedure.

3 Simulation studies

In this section, we conduct simulations to evaluate the finite sample performance of our proposed method. The network adjacent matrix $A=(a_{ij})\in\mathbb{R}^{n\times n}$ is generated by the R package *randnet* to implement the DCSBM. Specifically, the average degree is set as 20, the ratio of cross-block edges over within-block edges is 0, and the node degrees follow the power law distribution. Then, the weighting matrix W is set to be

$w_{ij} = a_{ij} / \sum_{j=1}^{n} a_{ij}$ for $i, j = 1, \cdots, n$. The corresponding block influence coefficients $\lambda = (\lambda_1, \cdots, \lambda_k)^\top$ are generated from the uniform distribution $U(0, 1)$. In addition, in the covariate vector $X_i = (x_{i1}, x_{i2})^\top \in \mathbb{R}^2$, let $x_{i1} \equiv 1$ and x_{i2} be independent and identically generated from the standard normal distribution $N(0, 1)$ for $i = 1, \cdots, n$. Its corresponding regression parameters are $\beta = (2, 1)^\top$. The random errors are independent and identically generated by two distributions: standard normal distribution $N(0, 1)$ and mixture distribution $0.9N(0, 5/9) + 0.1N(0, 5)$. It is worth noting that the above model settings satisfy our technical Conditions (C1)–(C8) in Supplementary Material S.2. Finally, the response vector Y is generated from model (8).

For each setting, we consider three sample sizes and four block numbers: $n = 200, 500, 1000$ and $k = 1, 2, 3, 4$, respectively. Meanwhile, all simulations are conducted via 1000 realizations. In addition, to assess the performance of parameter estimation, we define the vector estimator of θ_0 as $\hat{\theta}^{(h)}$ in the h-th realization. For each component of $\hat{\theta}^{(h)}$, the averaged bias of $\hat{\theta}_j^{(h)}$ is $\text{BIAS} = 1000^{-1} \sum_h (\hat{\theta}_j^{(h)} - \theta_{0,j})$, the true standard deviation of $\hat{\theta}_j^{(h)}$ is $\text{SD} = \{1000^{-1} \sum_h (\hat{\theta}_j^{(h)} - 1000^{-1} \sum_h \hat{\theta}_j^{(h)})^2\}^{1/2}$ and the root mean squared error is $\text{RMSE} = \sqrt{\text{SD}^2 + \text{BIAS}^2}$. Let $\text{SE}^{(h)}$ be the estimated standard error of $\hat{\theta}_j^{(h)}$ calculated with its asymptotic distribution of Theorem 1, then, the average of the estimated standard error is $\text{SE} = 1000^{-1} \sum_h \text{SE}^{(h)}$. Let $\text{SE}^{(Qh)}$ be the estimated standard error of $\hat{\theta}_j^{(h)}$ by bootstrap calibration (Bootstrap) procedure outlined in Sect. 2.3, then, the average of the estimated standard error is $\text{SE}^{\text{Q}} = 1000^{-1} \sum_h \text{SE}^{(Qh)}$.

Table 1 reports the BIAS, SD and RMSE of the QMLE via 1000 realizations when the distribution of error is a standard normal distribution. According to Table 1, we find that the values of BIAS and SD generally decrease for all parameter estimators and all four block numbers when the sample size n increases. It is not surprising that RMSE shows the same pattern. The above findings support our theoretical result that the QMLE is consistent and asymptotically normal. To demonstrate the performance of bootstrap procedure in covariance matrix estimation, Table 1 also provides the estimation of the standard errors using both the asymptotic distribution (SE) and bootstrap calibration (SE^{Q}). It is observed that both the procedures generate similar results, which echos the effectiveness of the bootstrap calibration in reality. Similarly, the QMLE also works well for mixture distribution of error depicted in Table 2.

We next assess the finite sample performance of the quasi-likelihood ratio test. Theoretically, the quasi-likelihood ratio test statistic $\mathcal{LR}$ is asymptotically weighted Chi-squared distribution with the weights $\lambda_r(\theta_0, \mu^3, \mu^4)$ under the null hypothesis. In order to conduct the test, we firstly generate the independent and identical random variables $\mathcal{X}_{r,h}^2$ from the Chi-squared distribution with 1 degree of freedom for $r = 1, \cdots, (k-1)$ and $h = 1, \cdots, 10{,}000$. Then, we can compute the p-value of the quasi-likelihood ratio test approximately by $p\text{-value} = 10000^{-1} \sum_h I\{\mathcal{LR} < \sum_{r=1}^{k-1} \lambda_{n,r}(\tilde{\theta}, \tilde{\mu}^3, \tilde{\mu}^4) \mathcal{X}_{r,h}^2\}$, where $\lambda_{n,r}(\tilde{\theta}, \tilde{\mu}^3, \tilde{\mu}^4)$ is a consistent estimator of $\lambda_r(\theta_0, \mu^3, \mu^4)$ under the null hypothesis stated Theorem 2. However, the convergence of $\mathcal{LR}$ to weighted Chi-squared distribution can be slow when the sample size is small, we also employ the bootstrap testing procedure outlined at the end of Sect. 2.4.

Furthermore, we evaluate the empirical sizes of the quasi-likelihood ratio test with the significance levels ranging from 0.01 to 0.30 and examine its empirical powers with the significance level 0.05. The empirical size and power are the percentages of rejections under H_0 and H_1, respectively, via the hypothesis test (13) with 1000 realizations. Specifically, the empirical size is the percentage of rejections under the setting of $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$, while the empirical power is the percentage of rejections under the settings of $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (\lambda_1\rho, 2\lambda_1\rho, 3\lambda_1\rho, 4\lambda_1\rho)$, where the signal strength $\rho > 0$. Figure 1 displays the empirical sizes of the proposed test using both the asymptotic distribution (QMLE) and bootstrap calibration (Bootstrap) procedures when the sample size $n = 200$. We observe that the likelihood ratio test based on asymptotic distribution exhibits size distortion while the bootstrap calibration performs consistently well. In addition, under the mixture normal distribution, Fig. 2 depicts the powers of the likelihood ratio test with bootstrap calibration when the sample size $n = 200$, 500, and 1000 and $k = 4$. We note that, as the increase of the signal strength ρ and sample size n, the powers of the test will increase to 100%. The testing results under the other two random error distributions, yield similar findings, so we do not present them here.

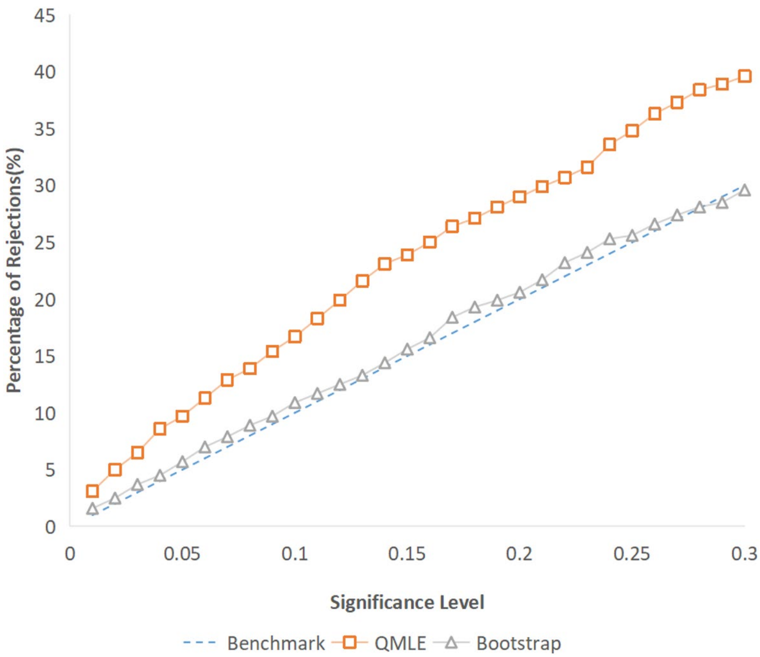

Fig. 1 The empirical sizes of the quasi-likelihood ratio test via asymptotic distribution (QMLE) and bootstrap calibration (Bootstrap) for the significance levels ranging from 0.01 to 0.30 when the sample size $n = 200$, under the setting of $k = 4$. The Benchmark represents the ideal case when the percentage of rejections from 1000 realizations is equal to the significance level. The independent and identically distributed random errors are simulated from mixture normal distribution $0.9N(0, 5/9) + 0.1N(0, 5)$

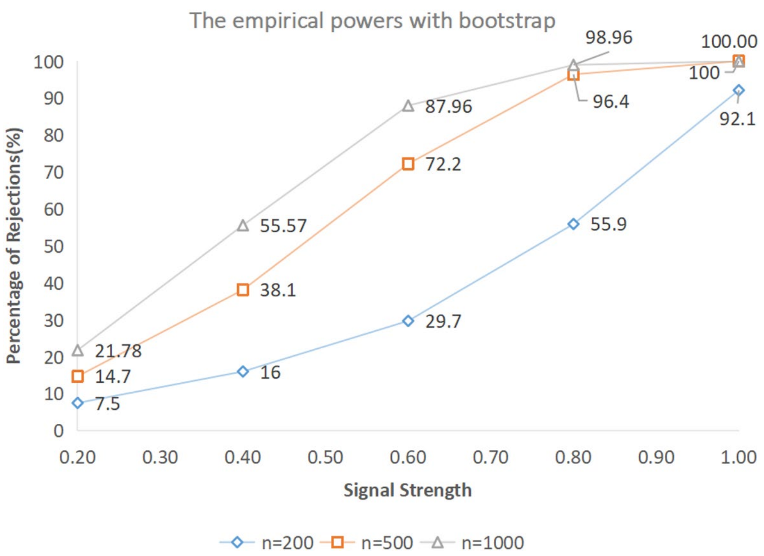

Fig. 2 The empirical powers of the quasi-likelihood ratio test by using bootstrap calibration at a nominal level of 0.05 under the setting of $k = 4$ with the sample size $n = 200, 500$ and 1000, respectively. The signal strengths $\rho = 0.2$, 0.4, 0.6, 0.8 and 1.0, which correspond to the settings of $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (\lambda_1\rho, 2\lambda_1\rho, 3\lambda_1\rho, 4\lambda_1\rho)$, respectively. The independent and identically distributed random errors are simulated from mixture normal distribution $0.9N(0, 5/9) + 0.1N(0, 5)$

4 Real data analysis

4.1 Data description

To study the effectiveness of the BWNAR model, we consider an application for detecting financial frauds, such as loan fraud, credit card fraud and insurance fraud, which cause serious consequences in the financial sector. Historically, a large number of studies had focused on the influence of the attributes of borrowers on fraud (see, e.g., Kirkos et al. 2005; Gao and Ye 2007; Panigrahi et al. 2009; Ravisankar et al. 2011; Xu et al. 2015; Dai et al. 2016; Malini and Pushpa 2017). Recently, some studies revealed that the performance of identifying fraud is improved by combining machine learning and complex networks (see, e.g., Zanin et al. 2018; Sadgali et al. 2019). Unlike the above research, we investigate loan fraud by simultaneously considering the network and covariate attributes of borrowers. Specifically, we assume that loan fraud tends to occur within professional fraud groups. That is, there is a block structure in the network.

In this dataset, information on fraud loans is collected from an internet lending platform in China, where borrowers with incomplete information and those who have not applied for loans are omitted. This results in a total of $n = 5083$ borrowers. Next, to efficiently identify the groups of fraud borrowers in the network, we construct an adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ via defining $a_{ij} = 1$ if borrowers i and j have the same register phone number or bank card, or have called or sent messages to each other, and $a_{ij} = 0$ otherwise. Hence, there are 31, 774 edges and the

Table 1 The performance of QMLE of parameters $\theta_0 = (\beta_1, \beta_2, \lambda_1, \cdots, \lambda_k, \sigma^2)^\top$ for $k \in \{1, 2, 3, 4\}$. The independent and identically distributed random errors are simulated from the normal distribution $N(0, 1)$. The influence coefficients λ_r are generated from the uniform distribution $U(0, 1)$ for $r = 1, \cdots, k$, and the adjacent matrix obeys DCSBM. Five measures are considered: the averaged bias of the estimators (BIAS), the standard deviation of the estimators (SD), the average of the estimated standard error via Theorem 1 (SE), the average of the estimated standard error via Bootstrap (SE^Q), and the root mean squared error of the estimators (RMSE)

Blocks	Para.	$n = 200$					$n = 500$					$n = 1000$				
		BIAS	SD	SE	SE^Q	RMSE	BIAS	SD	SE	SE^Q	RMSE	BIAS	SD	SE	SE^Q	RMSE
$k = 1$	$\hat{\beta}_1$	−0.0039	0.0784	0.0777	0.0775	0.0785	−0.0010	0.0497	0.0473	0.0468	0.0497	0.0003	0.0321	0.0337	0.0334	0.0321
	$\hat{\beta}_2$	0.0013	0.0745	0.0724	0.0717	0.0745	−0.0023	0.0494	0.0476	0.0469	0.0495	0.0035	0.0322	0.0307	0.0304	0.0324
	$\hat{\lambda}_1$	−0.0138	0.0270	0.0276	0.0260	0.0303	−0.0118	0.0243	0.0263	0.0218	0.0270	−0.0136	0.0276	0.0331	0.0276	0.0308
	$\hat{\sigma}^2$	−0.0050	0.1073	0.0996	0.0962	0.1075	0.0002	0.0677	0.0633	0.0623	0.0677	0.0003	0.0459	0.0448	0.0446	0.0459
$k = 2$	$\hat{\beta}_1$	−0.0065	0.0786	0.0775	0.0773	0.0788	−0.0016	0.0498	0.0472	0.0467	0.0498	0.0001	0.0323	0.0337	0.0334	0.0323
	$\hat{\beta}_2$	0.0029	0.0756	0.0730	0.0722	0.0757	−0.0036	0.0493	0.0475	0.0468	0.0494	0.0029	0.0323	0.0307	0.0304	0.0325
	$\hat{\lambda}_1$	−0.0327	0.0622	0.0642	0.0643	0.0702	−0.0184	0.0325	0.0324	0.0321	0.0374	−0.0190	0.0326	0.0370	0.0355	0.0377
	$\hat{\lambda}_2$	0.0311	0.1566	0.1676	0.1442	0.1597	−0.0044	0.1126	0.1160	0.1081	0.1127	0.0056	0.0712	0.0693	0.0738	0.0715
	$\hat{\sigma}^2$	−0.0099	0.1069	0.0992	0.0956	0.1074	−0.0023	0.0678	0.0632	0.0620	0.0679	−0.0011	0.0458	0.0447	0.0445	0.0458
$k = 3$	$\hat{\beta}_1$	−0.0077	0.0782	0.0776	0.0774	0.0786	−0.0022	0.0497	0.0472	0.0467	0.0498	−0.0004	0.0321	0.0337	0.0334	0.0321
	$\hat{\beta}_2$	0.0005	0.0742	0.0722	0.0716	0.0742	−0.0046	0.0492	0.0475	0.0467	0.0494	0.0027	0.0322	0.0307	0.0303	0.0323
	$\hat{\lambda}_1$	−0.0389	0.0839	0.0860	0.0834	0.0925	−0.0109	0.0231	0.0258	0.0266	0.0256	−0.0159	0.0304	0.0305	0.0323	0.0343
	$\hat{\lambda}_2$	−0.0190	0.1243	0.1441	0.1103	0.1258	−0.0129	0.1164	0.1223	0.1076	0.1171	−0.0001	0.0885	0.0878	0.0839	0.0885
	$\hat{\lambda}_3$	−0.0197	0.1206	0.1131	0.1208	0.1222	−0.0163	0.1174	0.1229	0.1214	0.1186	−0.0087	0.0784	0.0787	0.0788	0.0788
	$\hat{\sigma}^2$	−0.0136	0.1063	0.0988	0.0949	0.1071	−0.0045	0.0679	0.0631	0.0619	0.0681	−0.0027	0.0459	0.0447	0.0444	0.0460

Table 1 (continued)

Blocks	Para.	$n = 200$					$n = 500$					$n = 1000$				
		BIAS	SD	SE	SE^Q	RMSE	BIAS	SD	SE	SE^Q	RMSE	BIAS	SD	SE	SE^Q	RMSE
$k = 4$	$\hat{\beta}_1$	−0.0036	0.0781	0.0776	0.0775	0.0782	−0.0028	0.0500	0.0472	0.0467	0.0500	−0.0009	0.0323	0.0336	0.0333	0.0323
	$\hat{\beta}_2$	−0.0013	0.0750	0.0723	0.0716	0.0750	−0.0034	0.0492	0.0475	0.0468	0.0493	0.0025	0.0323	0.0308	0.0305	0.0324
	$\hat{\lambda}_1$	−0.0293	0.0733	0.0812	0.0784	0.0789	−0.0261	0.0491	0.0503	0.0519	0.0556	−0.0027	0.0088	0.0080	0.0096	0.0092
	$\hat{\lambda}_2$	0.0042	0.1550	0.1978	0.1449	0.1550	−0.0043	0.1207	0.1385	0.1165	0.1207	−0.0065	0.0964	0.0988	0.0918	0.0966
	$\hat{\lambda}_3$	−0.0405	0.0988	0.0970	0.1037	0.1068	−0.0230	0.1268	0.1250	0.1248	0.1289	−0.0110	0.0690	0.0677	0.0701	0.0698
	$\hat{\lambda}_4$	−0.0293	0.1843	0.2044	0.1770	0.1866	−0.0161	0.0910	0.0858	0.0906	0.0924	0.0072	0.1025	0.0989	0.0984	0.1027
	$\hat{\sigma}^2$	−0.0180	0.1067	0.0984	0.0944	0.1082	−0.0054	0.0680	0.0630	0.0618	0.0682	−0.0044	0.0458	0.0446	0.0443	0.0460

Table 2 The performance of QMLE of parameters $\theta_0 = (\beta_1, \beta_2, \lambda_1, \cdots, \lambda_k, \sigma^2)^\top$ for $k \in \{1, 2, 3, 4\}$. The independent and identically distributed random errors are simulated from the mixture normal distribution $0.9N(0, 5/9) + 0.1N(0, 5)$. The influence coefficients λ_r are generated from the uniform distribution $U(0, 1)$ for $r = 1, \cdots, k$, and the adjacent matrix obeys DCSBM. Five measures are considered: the averaged bias of the estimators (BIAS), the standard deviation of the estimators (SD), the average of the estimated standard error via Theorem 1 (SE), the average of the estimated standard error via Bootstrap (SE^Q), and the root mean squared error of the estimators (RMSE)

Blocks	Para.	$n = 200$					$n = 500$					$n = 1000$				
		BIAS	SD	SE	SE^Q	RMSE	BIAS	SD	SE	SE^Q	RMSE	BIAS	SD	SE	SE^Q	RMSE
$k = 1$	$\hat{\beta}_1$	−0.0044	0.0795	0.0780	0.0781	0.0796	−0.0010	0.0478	0.0474	0.0467	0.0478	0.0029	0.0341	0.0337	0.0334	0.0342
	$\hat{\beta}_2$	0.0011	0.0742	0.0726	0.0722	0.0742	−0.0011	0.0495	0.0477	0.0468	0.0495	0.0024	0.0315	0.0307	0.0304	0.0316
	$\hat{\lambda}_1$	−0.0127	0.0300	0.0273	0.0259	0.0326	−0.0116	0.0240	0.0262	0.0219	0.0267	−0.0136	0.0271	0.0271	0.0277	0.0304
	$\hat{\sigma}^2$	0.0073	0.1915	0.1807	0.1760	0.1916	0.0048	0.1273	0.1185	0.1168	0.1274	0.0015	0.0837	0.0841	0.0835	0.0837
$k = 2$	$\hat{\beta}_1$	−0.0068	0.0788	0.0778	0.0777	0.0791	−0.0016	0.0480	0.0473	0.0466	0.0481	0.0028	0.0345	0.0337	0.0334	0.0346
	$\hat{\beta}_2$	0.0024	0.0745	0.0732	0.0727	0.0745	−0.0024	0.0495	0.0475	0.0466	0.0496	0.0019	0.0316	0.0307	0.0304	0.0317
	$\hat{\lambda}_1$	−0.0331	0.0662	0.0642	0.0627	0.0740	−0.0179	0.0348	0.0302	0.0321	0.0391	−0.0211	0.0368	0.0374	0.0360	0.0425
	$\hat{\lambda}_2$	0.0205	0.1532	0.1681	0.1429	0.1546	−0.0007	0.1117	0.1160	0.1074	0.1117	0.0042	0.0751	0.0693	0.0739	0.0752
	$\hat{\sigma}^2$	0.0027	0.1903	0.1797	0.1740	0.1903	0.0024	0.1268	0.1181	0.1161	0.1268	0.0000	0.0834	0.0838	0.0832	0.0834
$k = 3$	$\hat{\beta}_1$	−0.0071	0.0788	0.0779	0.0779	0.0792	−0.0021	0.0480	0.0472	0.0465	0.0480	0.0022	0.0341	0.0337	0.0334	0.0342
	$\hat{\beta}_2$	0.0010	0.0741	0.0724	0.0720	0.0741	−0.0034	0.0492	0.0475	0.0465	0.0494	0.0017	0.0315	0.0307	0.0304	0.0316
	$\hat{\lambda}_1$	−0.0307	0.0627	0.0734	0.0807	0.0699	−0.0109	0.0251	0.0258	0.0274	0.0274	−0.0162	0.0338	0.0306	0.0321	0.0374
	$\hat{\lambda}_2$	−0.0182	0.1233	0.1443	0.1095	0.1246	−0.0101	0.1117	0.1222	0.1078	0.1122	−0.0020	0.0861	0.0878	0.0836	0.0862
	$\hat{\lambda}_3$	−0.0321	0.1197	0.1158	0.1217	0.1239	−0.0175	0.1265	0.1230	0.1208	0.1277	−0.0062	0.0779	0.0789	0.0792	0.0781
	$\hat{\sigma}^2$	−0.0002	0.1899	0.1785	0.1726	0.1899	−0.0002	0.1263	0.1176	0.1153	0.1263	−0.0015	0.0833	0.0837	0.0830	0.0833

Table 2 (continued)

Blocks	Para.	$n = 200$					$n = 500$					$n = 1000$				
		BIAS	SD	SE	SE^{Q}	RMSE	BIAS	SD	SE	SE^{Q}	RMSE	BIAS	SD	SE	SE^{Q}	RMSE
$k = 4$	$\hat{\beta}_1$	−0.0039	0.0789	0.0778	0.0779	0.0790	−0.0025	0.0480	0.0473	0.0466	0.0480	0.0017	0.0347	0.0336	0.0333	0.0347
	$\hat{\beta}_2$	−0.0014	0.0753	0.0725	0.0720	0.0754	−0.0021	0.0495	0.0475	0.0466	0.0496	0.0016	0.0320	0.0308	0.0305	0.0320
	$\hat{\lambda}_1$	−0.0282	0.0663	0.0612	0.0779	0.0721	−0.0250	0.0472	0.0401	0.0518	0.0535	−0.0031	0.0094	0.0080	0.0097	0.0099
	$\hat{\lambda}_2$	0.0039	0.1580	0.1980	0.1437	0.1581	−0.0083	0.1230	0.1386	0.1151	0.1233	−0.0061	0.1006	0.0987	0.0916	0.1007
	$\hat{\lambda}_3$	−0.0386	0.1030	0.0868	0.1031	0.1100	−0.0245	0.1253	0.1256	0.1238	0.1277	−0.0091	0.0684	0.0677	0.0699	0.0690
	$\hat{\lambda}_4$	−0.0399	0.1760	0.2077	0.1751	0.1804	−0.0216	0.0901	0.0865	0.0906	0.0926	−0.0007	0.1017	0.0996	0.0988	0.1017
	$\hat{\sigma}^2$	−0.0062	0.1902	0.1773	0.1705	0.1903	−0.0008	0.1270	0.1175	0.1152	0.1270	−0.0033	0.0831	0.0835	0.0827	0.0831

average degree is 12.5, where the degree of borrower i means the number of borrowers j associated with borrower i. Furthermore, Fig. 3 shows the degree distribution of this network, which is right skewed and it approximately follows a power-law distribution. Finally, we consider the probability of default (PD) caused by fraud as the response, which is measured by the probability that the borrowers failed to repay the debt that has been overdue for more than 3 months. There are 845 borrowers with a high probability of default in this dataset. In addition, the 16 covariates are depicted in detail in Table 3 and are divided into 4 categories, and we segment variables with the weight of evidence (WOE) method.

4.2 Empirical results

Before we adopt our BWNAR model, we first employ the community detection method mentioned in Sect. 2.1 to determine the optimal number of blocks and their memberships of each block. The outcome is $\hat{k} = 44$, which means 44 blocks exist in the dataset. Then, we fit the data by model (8) with the 44 blocks, and Tables 4 and 5 report the results of parameter estimators, their associated standard errors and t-statistics as well as the p-values. The standard errors of the estimators are obtained by using both the asymptotic distribution (QMLE) and bootstrap calibration (Bootstrap) outlined in Sect. 2.3. We observe from Tables 4 and 5 that both procedures produce similar standard errors and p-values, which indicates that the Conditions (C1)–(C8) in Supplementary Material S.2 are reasonable for the data. In addition, the pairwise comparisons between $\hat{\lambda}_r$s for $r = 1, \cdots, 44$ at the significance level 0.05 show that all $\hat{\lambda}_r$s are significantly different.

From the results, we notice that, there are 8 significant covariates at 5% significance level (Table 4). We discover that the top 3 factors, which are positively and significantly related to fraud risk, are the ID number of borrower in the loan blacklist, the borrowers' loan application count at nonbank institutes in the last

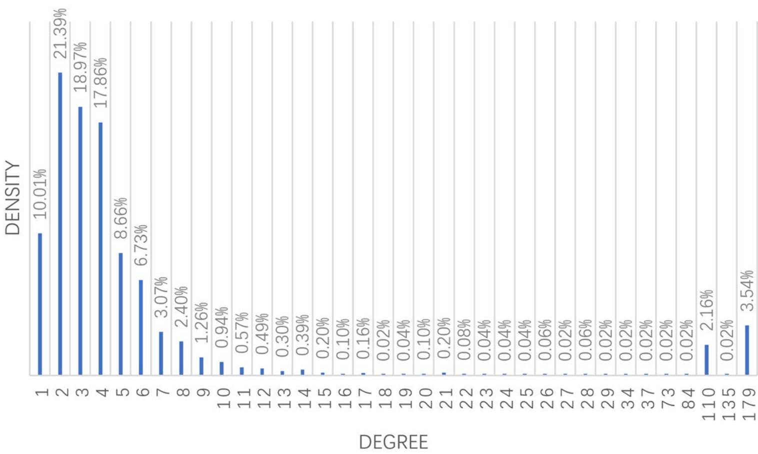

Fig. 3 Degree distribution of loan applicant network

Table 3 The 16 covariates and their descriptions in real data analysis. The segmented categories are based on the weight of evidence (WOE) method

Categories	Covariates	Description
Asset and Debt	AD_Bankcard_withdraw_amt_m12	Borrower's bank card withdraw amount in recent 12 months
	AD_Income	Borrower's income
	AD_Phone_bill_m5	Borrower's phone bills in recent 5 months
Credit History	CH_hit_p2p_bad_loan_list	Dummy variable equals 1 if the ID number of borrower is in the loan blacklist
	CH_loan_apply_cnt_m3	Borrower's loan application count in all institute in last 3 months
	CH_loan_apply_cnt_nonbank_m12	Borrower's loan application count in non-bank institute in last 12 months
	CH_loan_apply_cnt_nonbank_m3	Borrower's loan application count in non-bank institute in last 3 months
	CH_loan_apply_org_cnt_nonbank_m12	Borrower's loan application institute count of non-bank in last 12 months
Social interaction	SI_phone_contact_no	Borrower's phone contact numbers
	SI_phone_contact_workmate_no_m3	Borrower's phone contact workmate number in last 1 month
	SI_phone_nonactive_time	Borrower's phone nonactive time
	SI_phone_used_time	Borrower's phone used time
Terms of Loan	LN_loan_amount	The total loan amount
	LN_merchant_type	The type of merchants
	LN_merchent_grade	The grade of merchants
	LN_pct_down_payment	The percent of down payment to loan amount

Table 4 The estimation of the coefficient β, its standard error (SE), t-statistic and p-value in BWNAR model (8) for loan dataset. The standard errors are obtained via both asymptotic distribution (QMLE) and bootstrap calibration (Bootstrap)

Covariates	QMLE				Bootstrap		
	Estimate	SE	t-statistic	p-value	SE	t-statistic	p-value
AD_Bankcard_withdraw_amt_m12	0.0328	0.0386	0.8486	0.1981	0.0353	0.9281	0.1767
AD_Income	−0.0020	0.0219	−0.0899	0.4642	0.0187	−0.1055	0.4580
AD_Phone_bill_m5	−0.3892	0.1381	−2.8192	0.0024	0.1162	−3.3493	0.0004
CH_hit_p2p_bad_loan_list	0.7554	0.3586	2.1069	0.0176	0.3342	2.2608	0.0119
CH_loan_apply_cnt_m3	−0.0568	0.0235	−2.4162	0.0079	0.0205	−2.7732	0.0028
CH_loan_apply_cnt_nonbank_m12	0.1195	0.0238	5.0198	0.0000	0.0226	5.3008	0.0000
CH_loan_apply_cnt_nonbank_m3	0.0060	0.0359	0.1671	0.4337	0.0319	0.1879	0.4255
CH_loan_apply_org_cnt_non-bank_m12	0.0465	0.0327	1.4201	0.0778	0.0296	1.5689	0.0584
SI_phone_contact_no	0.0612	0.0213	2.8805	0.0020	0.0185	3.3134	0.0005
SI_phone_contact_workmate_no_m3	0.0624	0.0262	2.3791	0.0087	0.0228	2.7408	0.0031
SI_phone_nonactive_time	0.0291	0.0215	1.3560	0.0876	0.0184	1.5799	0.0571
SI_phone_used_time	−0.0160	0.0159	−1.0065	0.1571	0.0142	−1.1303	0.1292
LN_loan_amount	0.0130	0.0122	1.0675	0.1429	0.0089	1.4536	0.0731
LN_merchant_type	0.0190	0.0162	1.1684	0.1214	0.0143	1.3251	0.0926
LN_merchent_grade	0.0672	0.0172	3.9046	0.0000	0.0152	4.4158	0.0000
LN_pct_down_payment	0.0432	0.0095	4.5450	0.0000	0.0085	5.1095	0.0000

12 months and the grade of merchants. This implies that the borrowers have high repayment and default risk because they have applied to many financial companies. These findings are consistent with the intuition in real business scenarios.

For different blocks, we sort $\hat{\lambda}_r$s for $r = 1, \cdots, 44$ and obtain $\hat{\lambda}_{(1)} \geq \cdots \geq \hat{\lambda}_{(44)}$. Figure 4 and Table 5 depict the sorted influences. Furthermore, Table 5 shows interesting findings, that the influential power of blocks is positive and significantly related to the fraud risk. Specifically, the top 10 influential blocks have an average 24% fraud risk, while the last 10 have an average 9%. These findings are not surprising since a block has a greater influence, indicating that relationships between members in the block is complex, resulting in a higher fraud risk of the block.

To visualize the influential power, we explore the network structure in block 10 representing the top 10 blocks that have high influential power and high risk and block 36 representing the last 10 blocks that have low influential power and low risk. Each node in Fig. 5 is a borrower. The left panel of Fig. 5 reveals the entire network structure with 129 borrowers with 6, 267 connections. It is obvious the block consists of two groups. The membership of each group has a larger degree, resulting in the node having a greater number of connections; that is, the borrower is more trustworthy. All borrowers may be attracted by good loans and internal

Table 5 The estimated influence coefficients and its standard errors (SE), t-statistics and p-values for the top and last 10 influential blocks. The numbers of nodes (No. of nodes), edges (No. of edges) and the proportion of risky borrowers (Pct. of risk) in each block are shown in the right panel

	No.	QMLE				Bootstrap			No. of nodes	No. of edges	Pct. of risk
		$\hat{\lambda}$	SE	t-statistic	p-value	SE	t-statistic	p-value			
Top 10 influential blocks	1	0.41	0.08	5.42	0.0000	0.07	6.00	0.0000	152	410	22%
	2	0.38	0.07	5.73	0.0000	0.06	6.36	0.0000	175	339	25%
	3	0.38	0.09	4.31	0.0000	0.08	4.99	0.0001	114	277	21%
	4	0.38	0.08	4.55	0.0000	0.07	5.44	0.0000	147	379	24%
	5	0.37	0.07	5.10	0.0000	0.08	4.95	0.0001	149	283	20%
	6	0.37	0.13	2.82	0.0024	0.13	2.93	0.0052	42	68	21%
	7	0.36	0.08	4.72	0.0000	0.07	4.88	0.0001	127	221	24%
	8	0.36	0.07	5.13	0.0000	0.07	5.51	0.0000	152	275	26%
	9	0.32	0.08	4.14	0.0000	0.06	5.16	0.0001	146	273	29%
	10	0.32	0.09	3.38	0.0004	0.09	3.55	0.0015	97	184	30%
Avg.									130	271	24%
Last 10 influential blocks	35	0.10	0.11	0.93	0.1775	0.10	0.95	0.1788	93	161	9%
	36	0.10	0.25	0.39	0.3489	0.17	0.58	0.2863	129	6267	2%
	37	0.06	0.11	0.56	0.2890	0.10	0.62	0.2726	95	196	18%
	38	0.04	0.15	0.30	0.3836	0.11	0.39	0.3518	56	114	13%
	39	0.03	0.11	0.24	0.4051	0.08	0.33	0.3738	85	149	8%
	40	0.03	0.11	0.24	0.4058	0.09	0.28	0.3908	99	190	13%
	41	0.02	0.11	0.19	0.4263	0.09	0.24	0.4068	83	123	8%
	42	0.02	0.11	0.18	0.4272	0.09	0.22	0.4126	101	229	12%
	43	0.02	0.13	0.15	0.4398	0.16	0.12	0.4534	69	101	3%
	44	0.01	0.10	0.13	0.4497	0.09	0.14	0.4435	114	175	4%
Avg.									92	771	9%

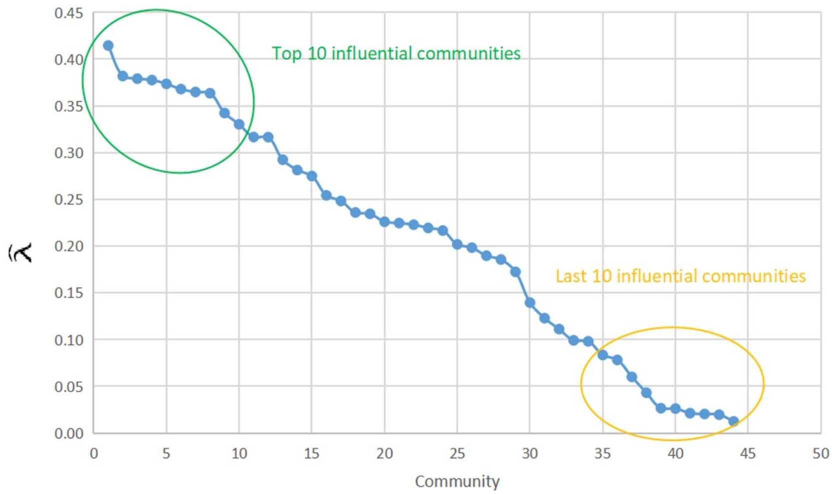

Fig. 4 The scatter plots of network influence coefficients for the 44 blocks

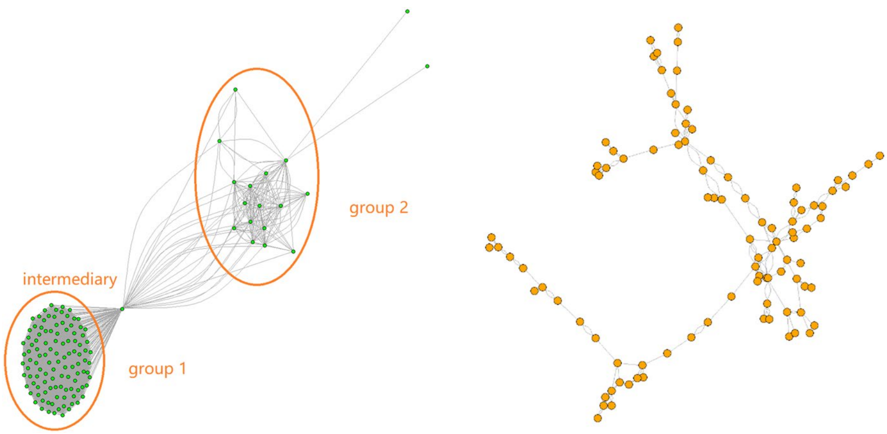

Fig. 5 The network structures of community 36 (left) and community 10 (right)

recommendations. The connections presented in the right panel reveal that the borrowers are from a large block but do not know each other.

In short, the best strategy is to lend the loan to borrowers who are connected with the most influential blocks rather than noninfluential blocks. That is, we effectively identify fraud borrowers by our BWNAR model.

5 Conclusion

In this work, we proposed the BWNAR model, which divides nodes into nonoverlapping blocks to analyze a network with blockwise structure. The number of blocks and their memberships in the network were first determined. Then, QMLE was employed to estimate the parameters of BWNAR model, and the asymptotic properties of estimators were investigated. Third, to confirm the heterogeneity of the influence parameters, a novel test statistic was proposed. Finally, we illustrated the performance of the BWNAR model via simulation studies and an application of the detection of fraud in financial transactions. It was found that fraudulent groups have a risk relationship, but ordinary loan applicants do not, which is consistent with reality.

There are five interesting extensions of this work for future research. The first extension is to analyze a blockwise network simultaneously combining the influence of cross-block and within-block nodes. For the assumption of non-overlapping grouping, the statement that the nodes are grouped into overlapping blocks is second extension. Developing a fast algorithm with theoretical justification is the third extension, for getting rid of the predicament of slow calculation of QMLE in the larger network. Fourth, a whole procedure to simultaneously estimate the group numbers together with its memberships and the model parameters is a further extension for the suboptimal but simple two-step method. The last extension is to generalize the BWNAR model to a dynamic blockwise network. We believe that these efforts are valuable and worth further investigation in order to broaden the application of the BWNAR model.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-022-00822-w.

Acknowledgements Wei Lan's research was supported by the National Natural Science Foundation of China (NSFC, 71991472, 12171395, 11931014), the Joint Lab of Data Science and Business Intelligence at Southwestern University of Finance and Economics, and the Fundamental Research Funds for the Central Universities (JBK1806002). Bin Guo's research was supported by the National Natural Science Foundation of China (NSFC, 11971390) and the Fundamental Research Funds for the Central Universities (JBK1806002). Bo Lei's research was supported by the Fundamental Research Funds for the Central Universities (JBK2207075). The authors are grateful to the Editor, the AE, and two anonymous referees for their insightful comments and constructive suggestions. Xiao and Lei are co-first authors.

References

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Anselin, L. (1990). Some robust approaches to testing and estimation in spatial econometrics. *Regional Science and Urban Economics, 20*, 141–163.

Blume, L. E., Brock, W. A., Durlauf, S. N., Jayaraman, R. (2015). Linear social interactions models. *Journal of Political Economy, 123*, 444–496.

Chen, Y., Trimborn, S., Zhang, J. (2018). Discover regional and size effects in global bitcoin blockchain via sparse-group network autoregressive modeling. Avaliable online: https://papers.ssrn.com/sol3/papers.

Cohen-Cole, E., Liu, X., Zenou, Y. (2018). Multivariate choices and identification of social interactions. *Journal of Applied Econometrics, 33*, 165–178.

Dai, Y., Yan, J., Tang, X., Zhao, H., Guo, M. (2016) Online credit card fraud detection: A hybrid framework with big data technologies. *In 2016 IEEE Trustcom/BigDataSE/ISPA* (pp. 1644–1651).

Durlauf, S. N., Young, H. P. (2001). The new social economics. In S. N. Durlauf & H. P. Young (Eds.), *Social Dynamics*. Cambridge, MA: MIT Press.

Fracassi, C. (2017). Corporate finance policies and social networks. *Management Science, 63*, 2420–2438.
Gao, Z., Ye, M. (2007). A framework for data mining-based anti-money laundering research. *Journal of Money Laundering Control, 10*, 170–179.
Hao, D., Ren, C., Li, C. (2012). Revisiting the variation of clustering coefficient of biological networks suggests new modular structure. *BMC Systems Biology, 6*, 1–10.
Huang, D., Wang, F., Zhu, X., Wang, H. (2020). Two-mode network autoregressive model for large-scale networks. *Journal of Econometrics, 216*, 203–219.
Karrer, B., Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E, 83*, 016107.
Kass-Hout, T. A., Alhinnawi, H. (2013). Social media in public health. *British Medical Bulletin, 108*, 5–24.
Kirkos, E., Spathis, C., Manolopoulos, Y. (2005). "Detection of fraudulent financial statements through the use of data mining techniques", In *Proceedings of the 2nd International Conference on Enterprise Systems and Accounting,* Thessaloniki, Greece, 310–325.
Kolaczyk, E. D. (2009). *Statistical analysis of network data*: *Methods and models*. Springer.
Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica, 72*, 1899–1925.
Lin, X., Weinberg, B. A. (2014). Unrequited friendship? How reciprocity mediates adolescent peer effects. *Regional Science and Urban Economics, 48*, 144–153.
Luo, X. (2014) A hierarchical graphical model for big inverse covariance estimation with an application to fMRI. *arXiv preprint:1403.4698*.
Ma, S., Su, L., Zhang, Y. (2021). Determining the number of communities in degree-corrected stochastic block models. *Journal of Machine Learning Research, 22*, 1–63.
Malini, N., Pushpa, M. (2017). Analysis on credit card fraud identification techniques based on KNN and outlier detection. *2017 Third International Conference on Advances in Electrical, Electronics Information, Communication and Bio-Informatics (AEEICB)*, 255–258.
Moscone, F., Tosetti, E., Vinciotti, V. (2017). Sparse estimation of huge networks with a blockwise structure. *The Econometrics Journal, 20*, S61–S85.
Panigrahi, S., Kundu, A., Sural, S., Majumdar, A. K. (2009). Credit card fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning. *Information Fusion, 10*, 354–363.
Ravisankar, P., Ravi, V., Rao, G. R., Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems, 50*, 491–500.
Sadgali, I., Sael, N., Benabbou, F. (2019). Performance of machine learning techniques in the detection of financial frauds. *Procedia Computer Science, 148*, 45–54.
Wang, X., Yu, C., Wei, Y. (2012). Social media peer communication and impacts on purchase intentions: A consumer socialization framework. *Journal of Interactive Marketing, 26*, 198–208.
Wang, W., Su, L. (2021). Identifying latent group structures in nonlinear panels. *Journal of Econometrics, 220*, 272–295.
Xu, J. J., Lu, Y., Chau, M. (2015). *P2P lending fraud detection: A big data approach. Pacific-Asia workshop on intelligence and security informatics*, 71–81. Cham: Springer.
Zanin, M., Romance, M., Moral, S., Criado, R. (2018). Credit card fraud detection through parenclitic network analysis. *Complexity, 2018*, 1–9.
Zhu, X., Huang, D., Pan, R., Wang, H. (2020). Multivariate spatial autoregressive model for large scale social networks. *Journal of Econometrics, 215*, 591–606.
Zou, T., Luo, R., Lan, W., Tsai, C. L. (2021). Network influence analysis. *Statistica Sinica, 31*, 1727–1748.