Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

Graphical Assistant Grouped Network Autoregression Model: A Bayesian Nonparametric Recourse

Yimeng Ren, Xuening Zhu, Xiaoling Lu & Guanyu Hu

To cite this article: Yimeng Ren, Xuening Zhu, Xiaoling Lu & Guanyu Hu (2024) Graphical Assistant Grouped Network Autoregression Model: A Bayesian Nonparametric Recourse, Journal of Business & Economic Statistics, 42:1, 49-63, DOI: 10.1080/07350015.2022.2143784

To link to this article: https://doi.org/10.1080/07350015.2022.2143784

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2024, VOL. 42, NO. 1, 49–63
https://doi.org/10.1080/07350015.2022.2143784

Graphical Assistant Grouped Network Autoregression Model: A Bayesian Nonparametric Recourse

Yimeng Ren[a], Xuening Zhu[a], Xiaoling Lu[b], and Guanyu Hu[c]

[a]School of Data Science, Fudan University, Shanghai, China; [b]Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, China; [c]Department of Statistics, University of Missouri Columbia, Columbia, MO

ABSTRACT

Vector autoregression model is ubiquitous in classical time series data analysis. With the rapid advance of social network sites, time series data over latent graph is becoming increasingly popular. In this article, we develop a novel Bayesian grouped network autoregression model, which can simultaneously estimate group information (number of groups and group configurations) and group-wise parameters. Specifically, a graphically assisted Chinese restaurant process is incorporated under the framework of the network autoregression model to improve the statistical inference performance. An efficient Markov chain Monte Carlo sampling algorithm is used to sample from the posterior distribution. Extensive studies are conducted to evaluate the finite sample performance of our proposed methodology. Additionally, we analyze two real datasets as illustrations of the effectiveness of our approach.

ARTICLE HISTORY
Received October 2021
Accepted October 2022

KEYWORDS
Graphical assistant Chinese restaurant process; MCMC; Mixture models; Network data; Vector autoregression

1. Introduction

The network data is becoming increasingly popular, which has many important applications in various disciplines, including sociology (Simmel 1950; Wasserman and Faust 1994; Hanneman and Riddle 2005), genomics (Wu, Feng, and Stein 2010; Horvath 2011), psychology (Borgatti et al. 2009; Cramer et al. 2010; Borsboom and Cramer 2013), finance and economics (Zou et al. 2017; Leung et al. 2017; Zhu et al. 2019). In the field of sociology, the network data is always used to study the structure of relationships by linking social units and modeling interdependencies in behaviors related to configurations of social relations (O'Malley and Marsden 2008). In biological and genetic areas, the presence of an interaction between two genes or proteins indicates a biologically functional relationship (Von Mering et al. 2002), which makes the network data widely used in cancer and other disease data analysis with promising results (Chuang et al. 2007; Guda, Chittur, and Guda 2009). A continuous response observed on each node at equally spaced time points is very common in various applications such as biological studies, economics, and finance. As a result, building a network based model to recover the dynamics of the responses is necessary.

We consider a network with N nodes, which are indexed as $i = 1, \ldots, N$. To represent the network relationship among the network nodes, we employ an adjacency matrix $\mathcal{A} = (a_{ij}) \in \mathbb{R}^{N\times N}$, where $a_{ij} = 1$ indicates the ith node follows the jth node, otherwise $a_{ij} = 0$. Following the convention we let $a_{ii} = 0$ for $1 \le i \le N$. Denote Y_{it} as the continuous response collected from the node i during $1 \le t \le T$. Correspondingly, we collect a number of nodal covariates (e.g., user's age, gender, etc.), which is denoted by $V_i \in \mathbb{R}^p$. To model the dynamics of the response Y_{it}, Zhu et al. (2017) proposed a network vector autoregression (NAR) model, which can be expressed as follows,

$$Y_{it} = \beta_0 + \beta_1 n_i^{-1} \sum_j a_{ij} Y_{j(t-1)} + \beta_2 Y_{i(t-1)} + V_i^\top \gamma + \varepsilon_{it}, \tag{1.1}$$

where $n_i = \sum_j a_{ij}$ is the out-degree of node i, $(\beta_0, \beta_1, \beta_2, \gamma^\top)$ are the parameters to be estimated, and ε_{it} is the random noise term. The NAR model embeds the observed adjacency matrix $\mathcal{A}$ into the modeling framework, and similar modeling techniques are applied in a broad range of fields, including spatial data modeling (Lee and Yu 2009; Shi and Lee 2017), social studies (Sojourner 2013; Liu, Patacchini, and Rainone 2017; Zhu and Pan 2020), financial risk management (Härdle, Wang, and Yu 2016; Zou et al. 2017), and many others.

Despite the usefulness of the NAR model, it still lacks flexibility in its model formulation and suffers from model misspecification. Specifically, it specifies a homogeneous network autoregression coefficient (i.e., β_1) for all network nodes, which is highly restrictive in practice. To enhance the model's flexibility, a popular way is to consider a group structure of model coefficients in panel data modeling (Ke, Fan, and Wu 2015; Su, Shi, and Phillips 2016; Ando and Bai 2016; Guðmundsson and Brownlees 2021). Specifically, to control the potential heterogeneity, the group panel data models consider group-wise slope coefficients in regression. The latent group structure can be estimated by the data information. For example, Su, Shi, and Phillips

CONTACT Xuening Zhu xueningzhu@fudan.edu.cn School of Data Science, Fudan University, Shanghai, China.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

(2016) propose a Classifier Lasso (C-Lasso) penalized procedure to shrink individual coefficients to the unknown group-specific coefficients, which simultaneously identifies groups and estimates parameters of panel data models. Bonhomme and Manresa (2015) and Bester and Hansen (2016) considered the group panel models with time-varying coefficients and interactive fixed effects, respectively. Liu et al. (2020) proposed a unified group parameter estimation framework under the condition that the number of groups is possibly over-specified. For the empirical studies, Gao, Xia, and Zhu (2020) studied the determinants of health care expenditure in OECD countries, where significant coefficients difference of five social economic variables are observed. Guðmundsson and Brownlees (2021) discovered the spillover group structure of the institutions in the financial system of the United States. See Gao, Xia, and Zhu (2020) and Li, Cui, and Lu (2020) for other group panel data extensions. With respect to the network data, Zhu and Pan (2020) considered a grouped network autoregression (GNAR) model, which introduces group heterogeneity to the NAR model (1.1). To be more specific, the nodes are clustered into several groups, and the nodes within the same group share a set of common parameters. As a consequence, the group labels of the nodes are totally determined by the nodal coefficients. In this model, both the group memberships and group-specific regression coefficients are to be estimated. Despite the usefulness of this type of approach, it does not incorporate the graphical information for estimating the group structure. To further motivate our study, we consider a real data example of stock return analysis, where the data details are given in Appendix E.2.1. We take the stock return as our response and construct a common shareholder network among the stocks. After applying the estimation procedure of the GNAR model, we obtain six groups. The within-group network density is given by $d_{\text{in}} = 0.0980$, which is larger than the between-group density $d_{\text{btw}} = 0.0932$. The calculation details are presented in Appendix E.2.2. This suggests the graphical information can be employed to model group structure among the nodes.

This is closely related to another line of research, which explores the group structure of the network data using the graphical information (i.e., $\mathcal{A}$) of the network nodes. In this regard, the group is typically referred to as community and the group labels are estimated by the community detection methods (Rohe, Chatterjee, and Yu 2011; Zhao, Levina, and Zhu 2012; Lei and Rinaldo 2015; Hu et al. 2020; Ma, Su, and Zhang 2021). A popular model for describing the community structure is the stochastic block model (SBM). In a SBM, the nodes within the same community are assumed to connect with higher probabilities, while nodes from different communities are less likely to connect. In Bayesian analysis, a similar popular choice to take advantage of the graphical information is to use nonparametric Bayesian methods for clustering nodes (Sewell and Chen 2017; van der Pas and van der Vaart 2018; Geng, Bhattacharya, and Pati 2019; Geng and Hu 2022). From the modeling aspects, both approaches assume that nodes should share denser connections within the same community. Particularly, we note that Geng and Hu (2022) directly embeds the adjacency matrix $\mathcal{A}$ in the prior information, which is more closely related to our modeling objective. However, although the community detection related methods have shown great usefulness in practical network data analysis, the related methods totally ignore the dynamics of $\{Y_{it}\}$.

To model the group heterogeneity pattern of $\{Y_{it}\}$ for network nodes, one needs to face the following three unsolved challenges. First, as we discussed before, the GNAR model does not incorporate the graphical information of the network nodes, while the community detection methods fail to consider the dynamics of $\{Y_{it}\}$. Consequently, the first challenge is to incorporate the graphical information into the estimation procedure of the GNAR model. It is noteworthy that the graphical information (i.e., a connectivity pattern among different nodes) will partially determine the group configurations. Particularly, we use a nonparametric Bayesian mixture model framework (Geng and Hu 2022) for exploiting the graphical information, under which nearby nodes are allowed to be grouped together with higher probability. Second, proposing a simultaneous inference procedure on both group configurations and group-wised parameters is another important challenge for the GNAR model. The uncertainty of parameter estimation is often neglected in existing literature, and it is difficult to get the probabilistic interpretations from existing frequentist approaches. Lastly, it is an important problem in grouping methods to determine the number of groups. Most existing methods (Bester and Hansen 2016; Zhu and Pan 2020; Liu et al. 2020) require specification of the number of groups first and then estimate group configurations. The number of groups is determined by information criteria (Ando and Bai 2016; Liu et al. 2020; Guðmundsson and Brownlees 2021) or some ad hoc procedures. As a consequence, the estimation may be sensitive to the tuning parameter specification. These procedures ignore the uncertainty of the estimation of the group number in the first stage and lack uncertainty quantification of them, which is also an important issue for GNAR model.

Our methodology development is directly motivated by solving the aforementioned challenges to fill the gap between nonparametric Bayesian methods and network autoregression model. In this article, we propose a novel graphical assistant grouped network autoregression (GAGNAR) model, considering both the heterogeneous network effects across nodes and the graph information in the grouping procedure under a novel nonparametric Bayesian mixture model framework. Specifically, a graphically assistant Chinese Restaurant Process (gaCRP) borrowed from Geng and Hu (2022) is introduced to the GNAR model framework, which incorporates the graphical information and uses the sampling procedure from the Chinese Restaurant Process (CRP) (Pitman 1995). The proposed prior will provide simultaneous inference on the number of groups and group configurations. Moreover, the gaCRP has a Pólya urn scheme similar to the traditional CRP. Due to this merit, we can develop a Gibbs sampler that enables efficient full Bayesian inference on the number of groups, mixture probabilities as well as group-specific parameters. We demonstrate its excellent numerical performance through simulations and analysis of two real datasets, compared with several benchmark methods.

Our proposed method is unique in the following aspects. First, the proposed method provides a useful model-based solution for the GNAR model that is able to leverage graphical information. In addition, a unified procedure is provided to simultaneously estimate the model parameters and determine

the group number. We show that the procedure is not sensitive to tuning parameter specification. Second, by adopting a full Bayesian framework, the grouping results yield useful probabilistic interpretation. Particularly, in the Bayesian estimation framework, our proposed method is able to quantify the uncertainty of the group estimation (including model parameter and group number estimation) with the posterior probabilities. Third, the developed posterior sampling scheme also renders efficient computation and convenient inference since our proposed collapsed Gibbs sampling algorithm by marginalizing over the number of groups could avoid complicated reversible jump Markov chain Monte Carlo (MCMC) algorithm (Green 1995) or allocation samplers.

The rest of this article is organized as follows. In Section 2, we briefly review the NAR and GNAR models, followed by a review of Bayesian mixture model and Dirichlet process. In addition, we describe the graphical assistant prior model and introduce our proposed GAGNAR model and its theoretical properties. The Bayesian inferences including the MCMC sampling algorithm, the post-MCMC estimation and the model selection criterion for tuning parameter are presented in Section 3. In Section 4, we study the finite sample performance of our proposed method via extensive simulation studies. A real data analysis of China fiscal revenue is illustrated by our proposed method in Section 5. Section 6 concludes and raises some in-depth discussions. All technique proofs, additional numerical results as well as another empirical application on stock return analysis are given in the supplementary materials.

2. Methodology

2.1. Grouped Network Autoregression Model

Consider a network with N nodes with the corresponding adjacency matrix denoted by $\mathcal{A}$. Assume the nodes are clustered into K latent groups. For each node i, suppose it carries a latent group membership $z_i \in \{1, \ldots, K\}$, where $z_i = k$ implies that the node i belongs to the kth group. Denote $\mathcal{F}_t$ as the σ-field generated by $\{Y_{it} : 1 \le i \le N, 1 \le t \le T\}$. Given $\mathcal{F}_{t-1}$, the observations at time point t are assumed to be independent and follow a Gaussian mixture distribution with K components. Given the number of groups K and the adjacency matrix $\mathcal{A}$, the GNAR model can be expressed as a mixture model, that is,

$$
\begin{aligned}
&p(Y_{it}|\boldsymbol{\theta}, \sigma^2) = \sum_k \pi_k \mathcal{N}(\mu_{it}, \sigma_k^2),\\
&\mathcal{N}(\mu_{it}, \sigma_k^2) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\Big\{ - \frac{(Y_{it} - \mu_{it})^2}{2\sigma_k^2} \Big\},\\
&\mu_{it} = \beta_{0z_i} + \beta_{1z_i} n_i^{-1} \sum_j a_{ij} Y_{j(t-1)} + \beta_{2z_i} Y_{i(t-1)} + V_i^\top \gamma_{z_i},
\end{aligned}
\tag{2.1}
$$

where $\mathcal{N}(\mu_{it}, \sigma_k^2)$ is the density function of normal distribution with mean μ_{it} and variance σ_k^2 in the kth group, and π_k is the mixing weight satisfying $\sum_k \pi_k = 1$. Here $\boldsymbol{\theta}_k = (\beta_{0k}, \beta_{1k}, \beta_{2k}, \gamma_k^\top)^\top$ collects the unknown parameters.

The GNAR model (2.1) characterizes the conditional mean of the response Y_{it} by a combination of four components. The first one is the *group baseline effect* β_{0k}, which is a constant but varying among different groups. The second one is the network component $(\beta_{1k} n_i^{-1} \sum_j a_{ij} Y_{j(t-1)})$, which reflects the average impact of following nodes at the previous time point. Therefore, we refer to β_{1k} as the *group network effect*. The third one is the momentum component $(\beta_{2k} Y_{i(t-1)})$. It characterizes how the focal node is influenced by its historical behaviors. The corresponding parameter β_{2k} is then referred to as *group momentum effect*, quantifying the node's self-dependence. The last one is the nodal covariate component $(V_i^\top \gamma_k)$, which includes the node specific characteristics and it is invariant over time. As implied by the GNAR model (2.1), the nodes within the same group share the same set of coefficients, therefore, they exhibit similar dynamic patterns. Compared to the NAR model (Zhu et al. 2017), the GNAR model is able to capture the nodes' heterogeneous pattern in the group level. To estimate the model parameters simultaneously with the group memberships, an EM algorithm is designed (Zhu and Pan 2020).

Despite the usefulness, the model has three main limitations. First, the estimation can be unreliable when the observed time length is short. That decreases the model's stability and robustness in practice. Second, the network structure information is not fully exploited for estimating the group memberships. Third, the number of groups K cannot be automatically estimated but needs to be specified in advance for the GNAR model. Although in existing panel data models the group number can be selected with some designed information criterion (Ando and Bai 2016; Liu et al. 2020; Guðmundsson and Brownlees 2021), the procedure is still sensitive to tuning parameter specification. That increases the possibility that the model is misspecified. As a consequence, we are motivated to consider the problem in a Bayesian framework and introduce a graph assisted prior, where the network structure information is employed when forming the groups. The details are presented in the following section.

2.2. Bayesian Graph Assisted Prior

Given $\{z_i\}$ and $\{\boldsymbol{\theta}_k, \sigma_k^2 : 1 \le k \le K\}$, we have a Gaussian mixture model (Marin, Mengersen, and Robert 2005) as follows,

$$
(Y_{it}|z_i, \boldsymbol{\theta}_{z_i}, \sigma_{z_i}^2) \sim \mathcal{N}(\mu_{it}, \sigma_{z_i}^2), \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,
$$

where $\mu_{it} = \beta_{0z_i} + \beta_{1z_i} n_i^{-1} \sum_j a_{ij} Y_{j(t-1)} + \beta_{2z_i} Y_{i(t-1)} + V_i^\top \gamma_{z_i}$. Now we introduce a prior for modeling $\{z_i\}$ by embedding the graph information.

Under the Bayesian hierarchical model, the priors for the unknown group memberships are given as

$$
\begin{aligned}
&z_i \sim \text{Cat}(\pi_1, \ldots, \pi_K), \quad i = 1, \ldots, N,\\
&\pi_1, \ldots, \pi_K \sim \text{Dirichlet}_K(\alpha, \ldots, \alpha),\\
&\boldsymbol{\theta}_j \sim \psi(\boldsymbol{\theta}), \quad j = 1, \ldots, K,
\end{aligned}
\tag{2.2}
$$

where z_i follows a categorical distribution with parameters $\pi = \{\pi_1, \ldots \pi_K\}$, π follows a Dirichlet distribution with parameter α, and $\psi(\boldsymbol{\theta})$ is the joint prior for all group-wise parameters. When K goes to infinity, the model in (2.2) becomes the Dirichlet process mixture model (DPMM; Ishwaran and Zarepour 2002b). The DPMM is very flexible and efficient for density estimation. However, it suffers from several problems in estimating the group structure. First, the partitions sampled from the DPMM

posterior tend to have multiple small transient groups, which decreases the interpretability of the model. Second, the DPMM is inconsistent on the estimation of the number of groups (Miller and Harrison 2018). Third, the DPMM does not take advantage of the network structure information: the connected nodes do not have higher probability being in the same group.

In order to address the above problems, we introduce a Bayesian graph assisted prior. Specifically, we assume that the group membership $z_i \in \{1, \ldots, K\}$ is generated from the weighted Dirichlet process. Define group membership vector $\mathbb{Z} = (z_1, \ldots, z_N)^\top \in \mathbb{R}^N$, and further denote $\mathbb{Z}_{(-i)} = \{z_j : j \neq i\}$. By incorporating the graph information, the conditional distribution of $z_i|\mathbb{Z}_{(-i)}$ is expressed as

$$P(z_i = k|\mathbb{Z}_{(-i)}) \propto \begin{cases} p_k(\mathbf{W}), & 1 \leq k \leq K' \\ \alpha, & k = K' + 1 \end{cases} \tag{2.3}$$

where $p_k(\mathbf{W}) = \sum_{j \neq i} w_{ij} I(z_j = k)$ and $\mathbf{W} = (w_{ij})$ is a weighting matrix calculated based on $\mathcal{A}$. The w_{ij} in the weight matrix denotes the connection strength between node i and j. We use K' to denote the current group number, and the parameter α controls the probability for introducing a new group. As a result, (2.3) implies that a node is more likely to fit in the same group with its connected friends.

One should note that the generation process guarantees the exchangeability of $\{z_i\}$ since the specification of $p_k(\mathbf{W})$ is not related to the ordering of nodes. As one can see, the graph information is directly embedded in the model of the conditional probability $P(z_i = k|\mathbb{Z}_{(-i)})$. When the graph is fully connected (i.e., $w_{ij} = 1$ for all $i \neq j$), the model reduces to the DPMM considered by Ishwaran and Zarepour (2002a). Similar approaches have been considered in recent literature. For instance, Lu, Li, and Dunson (2018) implemented powered CRP to penalize over-clustering in traditional CRP. However, the powered CRP neglected the graph information when conducting the clustering task. For network data analysis, Blei and Frazier (2011) proposed distance dependent CRP similar to (2.3) and model the networked documents. Chen et al. (2016) introduced node attributes and Bayesian priors to the Newman's mixture model (Newman and Leicht 2007), and they explore structural regularities in networks with node attributes. Similarly, Geng and Hu (2022) considered the gaCRP for group discovery of graph vertices. Although the above research considers the graph information for nodes clustering problem, they ignore the dynamics of the valuable response information.

Remark 1. Simultaneous estimation of the group number and configurations is typically achieved by complicated searching algorithms (e.g., the reversible jump MCMC (Green 1995)) under the probabilistic Bayesian framework. However, it suffers from high computational complexity. In our framework, the specification of (2.3) allows the nodes' memberships to be generated without setting the group number in advance. According to (2.3), a new group will formulate once z_i is assigned with a new group label $K' + 1$. Once we generate all the group memberships $\{z_i\}$, we know immediately the number of groups K.

Let G_0 be a continuous probability measure on $\mathbb{R}^p \times \mathbb{R}^+$. We define the full conditional distribution of $\boldsymbol{\theta}_i$ given $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \ldots, \boldsymbol{\theta}_N)$ as

$$f(\boldsymbol{\theta}_i|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \ldots, \boldsymbol{\theta}_N) \propto \sum_{r=1}^{K^*} \sum_{j \neq i} w_{ij} I(\boldsymbol{\theta}_i = \boldsymbol{\theta}_r^*) f_{\boldsymbol{\theta}_r^*}(\boldsymbol{\theta}_i) + \alpha G_0(\boldsymbol{\theta}_i), \tag{2.4}$$

where $f(\cdot)$ is the density function, K^* denotes the number of groups excluding the ith observation, $\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_{K^*}^*$ are K^* distinguished values of $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \ldots, \boldsymbol{\theta}_N\}$ and $I(\cdot)$ is the identity function. In practice we set the base distribution G_0 to be the normal inverse gamma distribution.

Inspired by the graph assisted Chinese restaurant process (gaCRP) (Geng and Hu 2022) for survival model, we define the graph distance d_{ij} between node i and j as the shortest path length between node i and j. If node i and j cannot be connected with a finite length of path, then $d_{ij} = \infty$. The weight is then defined as

$$w_{ij} = \begin{cases} 1, & \text{if } d_{ij} \leq 1 \\ \exp(-d_{ij} \times h), & \text{if } d_{ij} > 1 \end{cases}. \tag{2.5}$$

Therefore, the weight w_{ij} characterizes the closeness of the node i and j in the network. The parameter h denotes the smoothing scale of the weighting matrix. A larger value of h indicates the weight is more concentrating on the local scale. For simplicity, we refer to gaCRP introduced above as gaCRP(α, h). Although the idea of using a weighted version of CRP has been explored in Geng and Hu (2022), their model is based on Cox regression for survival data analysis, which significantly differs from our model, since it lacks discussion on dynamic process of Y_{it}.

Remark 2. Implied by (2.5), if $h = 0$, then the gaCRP is the same as the traditional CRP and may tend to over cluster the nodes. If $h \to \infty$, the probability for a new node to choose the existing groups tends to be small, which may also lead to the over-clustering problem. Hence, an appropriate selection of h is critical for the model performance. We discuss in details about the selection of h in Section 3.3.

In summary, we express the proposed model hierarchically as

$$\begin{aligned} &\text{Data model}: Y_{it}|z_i, \boldsymbol{\theta}_{z_i}, \sigma_{z_i}^2 \sim \mathcal{N}(\mu_{it}, \sigma_{z_i}^2), \\ &\text{Prior model}: p(z_i|\mathbb{Z}_{(-i)}) \sim \text{gaCRP}(\alpha, h), \\ &\text{Parameter model}: (\boldsymbol{\theta}_{z_i}, \sigma_{z_i}^2) \sim \text{NIG}(\boldsymbol{\tau}_0, \Sigma_0, a_0, b_0), \end{aligned} \tag{2.6}$$

where $\boldsymbol{\tau}_0$, Σ_0, a_0, and b_0 are hyperparameters for the base distribution of $(\boldsymbol{\theta}_k, \sigma_k^2)$. As one can see in (2.6), the true group for generating Y_{it} is still defined by the GNAR model. It means that the nodes from the same group should share a common set of parameters. We additionally add the graph information in the prior for modeling $\{z_i\}$ considering that nodes in the same group may share denser connections. This prior will help the estimation if the true group structure embeds certain graph information that nodes within the same group have denser connections. Hence, the group in our context is different from the definition of community in the SBM (and many other community detection methods), since their communities are totally determined by the graph information. Specifically, the communities are defined by a set of nodes with denser within-community connections and sparser between-community connections. However, in our setting, the groups are defined by the

model coefficients as in the GNAR model, while we embed the graph information in the prior.

Here the prior distribution of $(\boldsymbol{\theta}_{z_i}, \sigma^2_{z_i})$ is set to be normal inverse gamma (NIG) distribution with parameters $(\boldsymbol{\tau}_0, \Sigma_0, a_0, b_0)$. Without loss of generality, we choose $\boldsymbol{\tau}_0 = \mathbf{0}$, $\Sigma_0 = 100\mathbf{I}$, $a_0 = 0.01$, $b_0 = 0.01$ in our simulation study and real data applications. We refer to the model (2.6) as graph assisted group network autoregression (GAGNAR) model. The proposed process in (2.6) is a modification of the DPMM, which is a nonparametric mechanism to model the data distribution. It replaces the finite dimensional model (i.e., finite K) with an infinite dimensional model (i.e., diverging K) (Ishwaran and Zarepour 2002a). Our proposed model also starts with a finite parametric model with NIG distribution and taking the limit as the number of groups possibly goes to infinity. Our proposed model can automatically infer an adequate number of groups from the data, without resorting to Bayesian model comparison, which is an important characteristic of Bayesian nonparametric methods.

3. Bayesian Estimation

In this section, we discuss the model estimation for the GAGNAR model. To estimate the model parameters simultaneously with the latent group memberships, we employ the Markov chain Monte Carlo (MCMC) algorithm.

3.1. MCMC Algorithm

To conduct Bayesian estimation, we first derive the posterior distribution of the unknown parameters and group memberships. Define $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_K^2)$. In addition, denote $\mathbb{V} = (V_1, \ldots, V_N)^\top \in \mathbb{R}^{N\times p}$, $\mathbb{Y}_t = (Y_{1t}, \ldots, Y_{Nt})^\top \in \mathbb{R}^N$, and $\mathbb{Y} = (\mathbb{Y}_1, \ldots, \mathbb{Y}_T) \in \mathbb{R}^{N\times T}$. Given the data $\{\mathbb{Y}, \mathbb{V}\}$, the joint posterior distribution of latent membership $\mathbb{Z}$ and unknown parameters $\{\boldsymbol{\theta}, \boldsymbol{\sigma}^2\}$ is

$$\psi(\mathbb{Z}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2 | \mathbb{Y}, \mathbb{V}) \propto \psi(\mathbb{Z})\psi(\boldsymbol{\theta})\psi(\boldsymbol{\sigma}^2) f(\mathbb{Y}|\mathbb{Z}, \mathbb{V}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2), \quad (3.1)$$

where $\propto$ denotes "proportional to", and $\psi(\mathbb{Z})$, $\psi(\boldsymbol{\theta})$, $\psi(\boldsymbol{\sigma}^2)$ stand for the prior distributions for $\mathbb{Z}$, $\boldsymbol{\theta}$, and $\boldsymbol{\sigma}^2$, respectively. For convenience, let $\mathbb{X}_t \overset{\text{def}}{=} (X_{1t}, \ldots, X_{Nt})^\top = (\mathbf{1}_N, \mathbf{W}\mathbb{Y}_{t-1}, \mathbb{Y}_{t-1}, \mathbf{1}_N V_i^\top) \in \mathbb{R}^{N\times(p+3)}$, then given $\boldsymbol{\theta}_k, \sigma_k^2$ we have

$$(\mathbb{Y}_t | \boldsymbol{\theta}_k, \sigma_k^2) \sim \mathcal{N}(\boldsymbol{\mu}_k, \sigma_k^2 \mathbf{I}), \text{ where } \boldsymbol{\mu}_k = \mathbb{X}_t \boldsymbol{\theta}_k.$$

The prior distribution for $\mathbb{Z}$ is specified as gaCRP(α, h) as stated in Section 2.2. In addition, we specify the prior distribution for $\boldsymbol{\theta}_k$ as multivariate normal distribution $\mathcal{N}(\boldsymbol{\tau}_0, \sigma_k^2 \Sigma_0)$ and we set the prior for σ_k^2 as the commonly used scaled inverse gamma distribution IG(a_0, b_0), where $\boldsymbol{\tau}_0$, Σ_0, a_0l and b_0 are the hyperparameters.

Given the joint posterior distribution of $\{\mathbb{Z}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2\}$ in (3.1), we can obtain the conditional posterior distribution for each latent group membership and model parameters, respectively. This enables us to conduct the MCMC algorithm for model estimation. We first give the conditional posterior distribution for z_i $(i = 1, \ldots, N)$ as follows. For convenience we denote $\widetilde{p}(z_i = k) = p(z_i = k | \mathbb{Z}_{(-i)}, \mathbb{Y}, \mathbb{V}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2)$. Further define $\mathbb{Y}_{(i)} = (Y_{i1}, \ldots, Y_{iT})^\top \in \mathbb{R}^T$ and $\mathbb{X}_{(i)} = (X_{i1}, \ldots, X_{iT})^\top \in \mathbb{R}^{T\times(p+3)}$. Then we have

$$\begin{aligned}\widetilde{p}(z_i = k) &\propto f(\mathbb{Z}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2 | \mathbb{Y}, \mathbb{V}) \\ &\propto f(\mathbb{Y}_{(i)} | z_i = k, \mathbb{V}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2) p(z_i = k | \mathbb{Z}_{(-i)}).\end{aligned}$$

The analytical form of $\widetilde{p}(z_i = k)$ is given in Proposition 1.

Proposition 1. Suppose currently the nodes are formed into K' groups. Then according to (2.6) we have

$$\widetilde{p}(z_i = k) \propto \begin{cases} \kappa_k f(\mathbb{Y}_{(i)}; \mathbb{X}_{(i)}, \boldsymbol{\theta}_k, \sigma_k^2), & \text{for } 1 \le k \le K' \\ \alpha g(\mathbb{Y}_{(i)}; \boldsymbol{\tau}_0, \boldsymbol{\tau}, \Sigma_0, \Sigma, a_0, b_0), & \text{for } k = K'+1 \end{cases},$$

where $\kappa_k = \sum_{j\neq i} w_{ij} I(z_j = k)$ and

$$\begin{aligned}&f(\mathbb{Y}_{(i)}; \mathbb{X}_{(i)}, \boldsymbol{\theta}_k, \sigma_k^2) \\ &\quad = (2\pi\sigma_k^2)^{-(T-1)/2} \exp\Big\{ -\frac{1}{2\sigma_k^2} \|\mathbb{Y}_{(i)} - \mathbb{X}_{(i)}\boldsymbol{\theta}_k\|^2 \Big\}, \\ &g(\mathbb{Y}_{(i)}; \boldsymbol{\tau}_0, \boldsymbol{\tau}, \Sigma_0, \Sigma, a_0, b_0) \\ &\quad = \varphi(\Sigma_0, \Sigma, a_0, b_0) \\ &\qquad \Big\{ b_0 + \frac{1}{2}(\boldsymbol{\tau}_0^\top \Sigma_0^{-1} \boldsymbol{\tau}_0 + \mathbb{Y}_{(i)}^\top \mathbb{Y}_{(i)} - \boldsymbol{\tau}^\top \Sigma^{-1} \boldsymbol{\tau}) \Big\}^{-(a_0 + \frac{T-1}{2})}, \\ &\varphi(\Sigma_0, \Sigma, a_0, b_0) = \frac{b_0^{a_0} \Gamma(a_0 + \frac{T-1}{2}) |\Sigma|^{1/2}}{(2\pi)^{(T-1)/2} \Gamma(a_0) |\Sigma_0|^{1/2}}, \\ &\Sigma = (\Sigma_0^{-1} + \mathbb{X}_{(i)}^\top \mathbb{X}_{(i)})^{-1}, \quad \boldsymbol{\tau} = \Sigma(\Sigma_0^{-1} \boldsymbol{\tau}_0 + \mathbb{X}_{(i)}^\top \mathbb{Y}_{(i)}).\end{aligned}$$

By the formulation of $\widetilde{p}(z_i = k)$, we can observe that the probability z_i belonging to an existing group k $(1 \le k \le K')$ is closely related to the network weighting matrix $\mathbf{W}$. Particularly, κ_k denotes a weighted average ratio of its connected friends belonging to the group k, which characterizes a *stickiness* to the group k of ith node's neighbourhood. If the stickiness level is higher, then the conditional probability of $\{z_i = k\}$ will be higher. That is how the network topology information is involved in the sampling procedure. Next, the node is allowed to fit into a new group with probability proportional to $g(\mathbb{Y}_{(i)}; \boldsymbol{\tau}_0, \boldsymbol{\tau}, \Sigma_0, \Sigma, a_0, b_0)$, which relates to both the prior and the data information.

Next, we investigate the full conditional distribution for model parameters $\{\boldsymbol{\theta}, \boldsymbol{\sigma}^2\}$. Denote the posterior distribution of $(\boldsymbol{\theta}_k, \sigma_k^2)$ as $\widetilde{f}(\boldsymbol{\theta}_k, \sigma_k^2) = f(\boldsymbol{\theta}_k, \sigma_k^2 | \mathbb{Z}, \mathbb{Y}, \mathbb{V})$, then we derive $\widetilde{f}(\boldsymbol{\theta}_k, \sigma_k^2)$ in the following Proposition 2.

Proposition 2. The full conditional distribution $\widetilde{f}(\boldsymbol{\theta}_k, \sigma_k^2)$ $(k = 1, 2, \ldots, K)$ is given as

$$\begin{aligned}\widetilde{f}(\boldsymbol{\theta}_k, \sigma_k^2) &\propto \psi(\boldsymbol{\theta}_k, \sigma_k^2; \boldsymbol{\tau}_0, \Sigma_0, a_0, b_0) \prod_{z_i = k} f(\mathbb{Y}_{(i)}; \mathbb{X}_{(i)}, \boldsymbol{\theta}_k, \sigma_k^2) \\ &\propto f(\boldsymbol{\theta}_k, \sigma_k^2; \boldsymbol{\tau}^*, \Sigma^*, a^*, b^*),\end{aligned}$$

where

$$\begin{aligned}&f(\boldsymbol{\theta}_k, \sigma_k^2; \boldsymbol{\tau}_0, \Sigma_0, a_0, b_0) \\ &\quad = \frac{b_0^{a_0}}{(2\pi)^{d/2} |\Sigma_0|^{1/2} \Gamma(a_0)} \Big(\frac{1}{\sigma_k^2}\Big)^{a_0 + \frac{d}{2} + 1} \\ &\qquad \exp\Big[-\frac{1}{\sigma_k^2} \Big\{ b_0 + \frac{1}{2}(\boldsymbol{\theta}_k - \boldsymbol{\tau}_0)^\top \Sigma_0^{-1} (\boldsymbol{\theta}_k - \boldsymbol{\tau}_0) \Big\} \Big],\end{aligned}$$

$$\boldsymbol{\tau}^* = (\Sigma_0^{-1} + \sum_{z_i=k} \mathbb{X}_{(i)}^\top \mathbb{X}_{(i)})^{-1}(\Sigma_0^{-1}\boldsymbol{\tau}_0 + \sum_{z_i=k} \mathbb{X}_{(i)}^\top \mathbb{Y}_{(i)}),$$

$$\Sigma^* = (\Sigma_0^{-1} + \sum_{z_i=k} \mathbb{X}_{(i)}^\top \mathbb{X}_{(i)})^{-1},$$

$$a^* = a_0 + (T-1)N_k/2,$$

$$b^* = b_0 + \frac{1}{2}\Big(\boldsymbol{\tau}_0^\top \Sigma_0^{-1}\boldsymbol{\tau}_0 + \sum_{z_i=k} \mathbb{Y}_{(i)}^\top \mathbb{Y}_{(i)} - \boldsymbol{\tau}^{*\top}\Sigma^{*-1}\boldsymbol{\tau}^*\Big),$$

$$\text{and } N_k = \sum_i I(z_i = k).$$

The proofs of the two propositions are given in supplementary materials. As shown in Proposition 2, the conditional distribution of $\{\boldsymbol{\theta}_k, \sigma_k^2\}$ follows normal inverse gamma distribution (NIG) with density function $f(\boldsymbol{\tau}^*, \Sigma^*, a^*, b^*)$. With the expressions of $(\boldsymbol{\tau}^*, \Sigma^*, a^*, b^*)$, we see that it unifies the prior information $\{\boldsymbol{\tau}_0, \Sigma_0, a_0, b_0\}$ and the data information from the kth group, that is, $\{\mathbb{X}_{(i)}, \mathbb{Y}_{(i)} : z_i = k\}$. Then, based on Propositions 1 and 2, we could efficiently cycle through the full conditional distribution $\widetilde{p}(z_i = k)$ for $i = 1, 2, \ldots, N$ and $\widetilde{f}(\boldsymbol{\theta}_k, \sigma_k^2)$ for $k = 1, 2, \ldots, K$ to conduct the Gibbs sampling procedure. One should note that in the process of node memberships generation, K is marginalized over, which could avoid complicated reversible jump MCMC algorithms or allocation samplers. We summarize the collapsed Gibbs sampling procedure for the GAGNAR model in Algorithm 1.

Algorithm 1 Collapsed Gibbs sampler for GAGNAR

1: Initialize the number of groups K, the group assignment z_i and the parameters $(\boldsymbol{\theta}_{z_i}, \sigma_{z_i}^2)$ for each $\mathbb{Y}_{(i)}$
2: **for** *iter* from 1 to M **do**
3: **for** i from 1 to N **do**
4: Update z_i conditional on $\boldsymbol{\theta}, \sigma^2$ for each node $i \in (1, \ldots, N)$ by the conditional distribution $\widetilde{p}(z_i = k)$ in Proposition 1.
5: If updated z_i belongs to existing groups, then K does not change; otherwise, if updated z_i forms a new group, then $K := K + 1$.
6: **end for**
7: **for** c from 1 to K **do**
8: Update $(\boldsymbol{\theta}_c, \sigma_c^2)$ for each $c \in (1, \ldots, K)$ conditional on $(\mathbb{Z}, \mathbb{Y}, \mathbb{V})$ by the density function in Proposition 2 as

$$\widetilde{f}(\boldsymbol{\theta}_c, \sigma_c^2) \sim \text{NIG}(\boldsymbol{\tau}^*, \Sigma^*, a^*, b^*)$$

9: **end for**
10: **end for**

3.2. Post MCMC Estimation

Let $\boldsymbol{\theta}_k^{(m)}$ and $\sigma_k^{2(m)}$ with $1 \le k \le K^{(m)}$ be the mth estimation after the number of burn-in iterations of the MCMC algorithm. Correspondingly denote $\mathbb{Z}^{(m)} = (z_i^{(m)} : 1 \le i \le N)^\top$ as the estimated memberships of the network nodes. Particularly we note that the estimated number of groups $K^{(m)}$ can be different for $1 \le m \le M$. As a result, the direct posterior estimation of the parameters would be difficult since the number of estimated parameters are not the same across different iterations. To address this issue, we adopt Dahl's method (Dahl 2006) to select the best post burn-in iteration with the least squares criterion. The estimate output by the best post burn-in iteration is then taken as the final post MCMC estimator.

Specifically, we take advantage of the co-membership matrix to help us select the best post burn-in iteration. Define the co-membership matrix as $\mathbf{B} = (b_{ij}) \in \mathbb{R}^{N\times N}$, where $b_{ij} = I(z_i = z_j)$. Therefore, the (i, j)th element of $\mathbf{B}$ denotes whether the ith node is in the same group with the jth node. We can see that $\mathbf{B}$ is well defined for different $K^{(m)}$s and it does not suffer from the label switching issue. We then choose the best post burn-in iteration as

$$m_b = \text{argmin}_{1\le m\le M} \big\|\mathbf{B}^{(m)} - \overline{\mathbf{B}}\big\|_F^2,$$

where $\mathbf{B}^{(m)}$ is the estimated co-membership matrix in the mth post burn-in iteration and $\overline{\mathbf{B}} = M^{-1}\sum_{m=1}^M \mathbf{B}^{(m)}$. As a consequence, the best post burn-in iteration is selected by the closest $\mathbf{B}^{(m)}$ to the mean grouping co-membership matrix $\overline{\mathbf{B}}$. The number of groups is then determined as $K^{(m_b)}$. Accordingly, the post MCMC estimations of the parameters and memberships are given by $(\boldsymbol{\theta}_k^{(m_b)}, \sigma_k^{2(m_b)})$ with $1 \le k \le K^{(m_b)}$ and $\mathbb{Z}^{(m_b)}$, respectively.

3.3. Selection of Smoothing Parameter h

The selection of smoothing parameter h is redesigned as a model selection problem. Specifically, we use the logarithm of the pseudo marginal likelihood (LPML) (Ibrahim et al. 2001) based on conditional predictive ordinate (CPO) (Gelfand, Dey, and Chang 1992) to select h.

The LPML given a specified h is defined as

$$\text{LPML}(h) = \sum_{i=1}^N \log\{\text{CPO}_i(h)\}, \tag{3.2}$$

where $\text{CPO}_i(h)$ is the CPO for the node i. The CPO is defined as $\text{CPO}_i(h) = p(\mathbb{Y}_{(i)}|\mathbb{Y}_{(-i)})$ (Pettit 1990), where $\mathbb{Y}_{(-i)} = \{\mathbb{Y}_{(j)} : j \ne i\}$. As one can see, LPML is a pseudo log-likelihood function and we prefer a smoothing parameter h to maximize LPML. Borrowing the idea of Chen, Shao, and Ibrahim (2012), we obtain the Monte Carlo estimate of CPO within the Bayesian framework as

$$\widehat{\text{CPO}}_i(h) = \Big\{\frac{1}{M}\sum_{m=1}^M \frac{1}{L(\boldsymbol{\theta}_{z_i}^{(m)}, \sigma_{z_i}^{2(m)}; h)}\Big\}^{-1},$$

where M is the total number of Monte Carlo iterations, and

$$L(\boldsymbol{\theta}_{z_i}^{(m)}, \sigma_{z_i}^{2(m)}; h) = \prod_{t=2}^T p(Y_{it}|\boldsymbol{\theta}_{z_i}^{(m)}, \sigma_{z_i}^{2(m)}; h)$$

is the likelihood function for the node i with the specified smoothing parameter h. Correspondingly we have $\widehat{\text{LPML}}(h) = \sum_{i=1}^N \log\{\widehat{\text{CPO}}_i(h)\}$ and we choose the optimal h by $h_b = \arg\max_h \widehat{\text{LPML}}(h)$.

4. Simulation

4.1. Simulation Models

To demonstrate the finite sample performance of our proposed method, we conduct a number of numerical studies in this section. Specifically, we use three types of graphs and assign two parameter settings to each of them.

For each simulation setting, we generate the random noise ε_{it} from a standard normal distribution. In addition, node covariates $V_i \in \mathbb{R}^p$ are independently sampled from a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$. For each graph, two different scenarios of parameters $(\beta_{0k}, \beta_{1k}, \beta_{2k}, \gamma_k, \sigma_k^2)$ are considered, which are listed in Table 1. Given the initial value $\mathbb{Y}_0 = \mathbf{0}$, the time series $\mathbb{Y}_t$ is generated according to the grouped network vector autoregression model in (2.1). In each scenario, 100 replicated datasets are generated, and in each replicate we set the time length $T = 20$. In addition, for the prior distribution gaCRP(α, h), we set $\alpha = 1$ and select the smoothing parameters $h \in \{0, 0.2, 0.4, \ldots, 5.0\}$ by LPML (3.2). A total of 1500 MCMC iterations are run for each replicate, with the first 500 iterations treated as the burn-in stage. Besides, the hyperparameters set for the base distribution in (2.6) are $(a_0, b_0, \boldsymbol{\tau}_0, \Sigma_0) = (0.01, 0.01, \mathbf{0}, 100\mathbf{I})$, which are noninformative priors.

Example 1 (Stochastic Block Model). We first consider the stochastic block model (SBM) (Wang and Wong 1987; Nowicki and Snijders 2001; Zhao, Levina, and Zhu 2012). The SBM assumes that nodes in the same group (block) are more likely to be connected, when compared with nodes from different groups. The model is widely used to discover community structures for network data. We follow Nowicki and Snijders (2001) to randomly assign each node a group label $k = 1, \ldots, K$ with equal probability $1/K$ and we set $K = 3$. Next, let $P(a_{ij} = 1) = 20N^{-1}$ if node i and node j are in the same group, and $P(a_{ij} = 1) = 2N^{-1}$ otherwise. The network size is set as $N = 100$. As a result, the generating mechanism of this example ensures that the nodes within the group should share denser connections than between groups.

Example 2 (Chinese Cities Graph). In this example we construct the graph of Chinese cities by using the geographical information. We collect 151 cities of mainland China and treat each city as a network node in the graph. The edge between two cities is defined as whether they share the common boarder (Cao, Liang, and Niu 2017; Zhang et al. 2018). Specifically, $a_{ij} = 1$ illustrates that two cities are connected, otherwise $a_{ij} = 0$. Lastly, we let $K = 5$ be the total number of groups and assign the group labels using k-means clustering on the distance matrix.

Example 3 (Common Shareholder Network). We lastly construct the stock graph from Chinese A stock market, which contains $N = 180$ actively traded stocks in the Shanghai and the Shenzhen Stock Exchange. The stocks used in this simulation example are the subset of our empirical study for computational convenience. Specifically, $a_{ij} = 1$ indicates that two stocks share at least two of the top 10 shareholders, otherwise $a_{ij} = 0$. The graph structure is shown in the bottom right panel of Figure 1. We set the number of groups $K = 6$ and assign the group memberships by implementing the spectral clustering algorithm to the adjacency matrix.

The three network structures with the true group labels are visualized in Figure 1.

4.2. Performance Measurements and Simulation Results

Let $(\hat{\beta}_{0k}^{(r)}, \hat{\beta}_{1k}^{(r)}, \hat{\beta}_{2k}^{(r)}, \hat{\gamma}_k^{(r)}, \hat{\sigma}_k^{2(r)})$ be the estimated parameters of the kth group in the rth replicate $(1 \le r \le R)$. For each node i, we obtain its group label as $\hat{z}_i^{(r)} (i = 1, \ldots, N)$ by Dalh's method. Subsequently, the estimated parameters for each node i is given as $\left(\hat{\beta}_{0\hat{z}_i^{(r)}}^{(r)}, \hat{\beta}_{1\hat{z}_i^{(r)}}^{(r)}, \hat{\beta}_{2\hat{z}_i^{(r)}}^{(r)}, \hat{\gamma}_{\hat{z}_i^{(r)}}^{(r)}, \hat{\sigma}_{\hat{z}_i}^{2(r)}\right)$. We consider the following measurements to evaluate the finite sample performance. First,

Table 1. Simulation parameters setting for three examples.

	σ_k^2	β_{0k}	β_{1k}	β_{2k}	γ_k	σ_k^2	β_{0k}	β_{1k}	β_{2k}	γ_k
	Example 1									
	Scenario 1					Scenario 2				
Group 1	2.0	5.0	0.2	0.1	$(0.5, 0.7, 1.0)^\top$	2.0	0.0	0.1	0.3	$(0.5, 0.7, 1.0)^\top$
Group 2	1.0	−5.0	−0.4	0.2	$(0.1, 0.9, 0.4)^\top$	4.0	0.2	−0.3	0.2	$(0.1, 0.9, 0.4)^\top$
Group 3	3.0	0.0	0.2	0.4	$(0.2, -1.0, 2.0)^\top$	3.0	0.5	0.2	0.7	$(0.2, -0.2, 1.4)^\top$
	Example 2									
	Scenario 1					Scenario 2				
Group 1	2.0	5.0	0.2	0.1	$(0.5, 0.7, 1.0)^\top$	2.0	0.0	0.1	0.3	$(0.5, 0.7, 1.0)^\top$
Group 2	1.0	−5.0	−0.4	0.2	$(0.1, 0.9, 0.4)^\top$	1.0	0.2	−0.3	0.2	$(0.1, 0.9, 0.4)^\top$
Group 3	3.0	0.0	0.2	0.4	$(0.2, -1.0, 2.0)^\top$	3.0	0.5	0.2	0.7	$(0.2, -0.2, 1.4)^\top$
Group 4	4.0	−0.1	0.1	0.2	$(1.0, -1.0, 1.5)^\top$	4.0	−0.1	0.1	0.2	$(1.0, -1.0, 1.5)^\top$
Group 5	2.0	3.0	0.5	0.2	$(0.8, 0.5, -2.0)^\top$	2.0	0.8	0.5	0.2	$(0.8, 0.5, -1.0)^\top$
	Example 3									
	Scenario 1					Scenario 2				
Group 1	2.0	5.0	0.2	0.1	$(0.5, 0.7, 1.0)^\top$	2.0	0.0	0.1	0.3	$(0.5, 0.7, 1.0)^\top$
Group 2	1.0	−5.0	−0.4	0.2	$(0.1, 0.9, 0.4)^\top$	1.0	3.0	−0.3	0.2	$(0.1, 0.9, 0.4)^\top$
Group 3	3.0	0.0	0.2	0.4	$(0.2, -1.0, 2.0)^\top$	3.0	−3.0	0.2	0.7	$(0.2, -0.2, 1.4)^\top$
Group 4	4.0	3.0	0.1	0.2	$(1.0, -1.0, 1.5)^\top$	1.5	4.5	0.1	0.2	$(1.0, -1.0, 1.5)^\top$
Group 5	2.0	−3.0	0.5	0.2	$(0.8, 0.5, -2.0)^\top$	2.5	−2.0	0.5	0.2	$(0.8, 0.5, -1.0)^\top$
Group 6	3.0	2.0	−0.6	−0.2	$(-0.8, 0.5, 2.0)^\top$	1.0	2.0	−0.6	−0.2	$(-0.8, 0.5, 2.0)^\top$

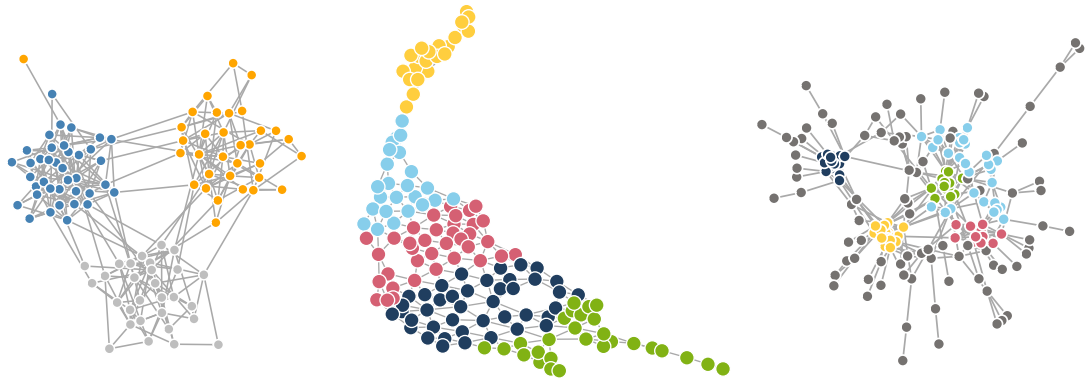

Figure 1. Three graphs with true group labels, each color representing a group. Left: graph generated from SBM with $K = 3$. Middle: geographical graph of $N = 151$ Chinese cities with $K = 5$. Right: common shareholder graph of $N = 180$ stocks with $K = 6$.

we employ the root mean square error (RMSE) to evaluate the estimation accuracy. For β_{sk} $(0 \le s \le 2, 1 \le k \le K)$, the RMSE is defined as

$$\mathrm{RMSE}_{\beta_s} = \left\{ (RN)^{-1} \sum_{r=1}^{R} \sum_{i=1}^{N} \left(\hat{\beta}^{(r)}_{s\widehat{z}_i^{(r)}} - \beta_{sz_i} \right)^2 \right\}^{1/2},$$

and the RMSE for γ is calculated as

$$\mathrm{RMSE}_{\gamma} = \left\{ (RN)^{-1} \sum_{r=1}^{R} \sum_{i=1}^{N} \left\| \hat{\gamma}^{(r)}_{\widehat{z}_i^{(r)}} - \gamma_{z_i} \right\|^2 \right\}^{1/2}.$$

To measure the similarity between the estimated group memberships and the true group memberships, we use the Adjusted Rand Index (ARI) (Rand 1971; Hubert and Arabie 1985). Define $\{z_i : 1 \le i \le N\}$ as the set of the true group memberships. Then ARI is defined as

$$\mathrm{ARI}^{(r)} = \frac{\mathrm{RI}^{(r)} - E\left(\mathrm{RI}^{(r)}\right)}{\max\left(\mathrm{RI}^{(r)}\right) - E\left(\mathrm{RI}^{(r)}\right)}, \quad \mathrm{RI}^{(r)} = \frac{a^{(r)} + b^{(r)}}{C_N^2},$$

where $a^{(r)} = \sum_{i,j} I(z_i = z_j, \widehat{z}_i^{(r)} = \widehat{z}_j^{(r)})$, $b^{(r)} = \sum_{i,j} I(z_i \ne z_j, \widehat{z}_i^{(r)} \ne \widehat{z}_j^{(r)})$, and $C_N^2 = N(N-1)/2$ is the total number of possible node pairs formed by the N nodes. Hence, ARI measures the alignment level of two grouping results. A higher ARI value implies the estimated group memberships are more consistent with the true memberships. The ARI can be calculated using the R-package *mclust*. We compare the parameter estimation of proposed model with that of EM algorithm and two-step estimation for the GNAR model (Zhu and Pan 2020). One should note that, since the number of groups could not be inferred in the other two baseline methods, we apply the true value of K as the input of EM and two-step methods.

Table 2 summarizes the average RMSE of estimated parameters under the optimal h selected by LPML. For all three simulation examples, we see that the overall RMSEs of estimated parameters by the GAGNAR model are much lower than those of the EM and two-step methods. This is consistent with the weakness we mentioned that the GNAR model does not use the network structure information. In addition, we see that graphs in example 2 and 3 are more complicated than that generated from SBM, due to the larger number of groups and the group patterns which are not easily captured (see Figure 1). In both examples, the estimation accuracy of our model is shown to be much higher than that of the EM and two-step estimators. This further confirms the usefulness of the proposed GAGNAR model in the intricate reality.

Table 2. RMSEs of parameter estimation.

	β_0		β_1		β_2		γ		σ^2	
					Example 1					
LPML	0.626	0.471	0.179	0.377	0.057	0.108	0.264	0.532	0.330	0.624
EM	1.706	0.233	0.164	0.167	0.132	0.123	0.514	0.279	1.450	1.466
2-step	3.774	0.226	0.318	0.242	0.496	0.258	0.717	0.380	1.128	1.474
					Example 2					
LPML	0.722	0.265	0.088	0.176	0.085	0.135	0.404	0.429	0.450	0.596
EM	2.452	0.317	0.329	0.186	0.157	0.149	0.963	0.482	1.245	1.241
2-step	3.191	0.370	0.329	0.282	0.508	0.371	0.956	0.726	1.188	1.309
					Example 3					
LPML	0.920	0.644	0.158	0.180	0.093	0.089	0.460	0.318	0.443	0.246
EM	1.942	1.416	0.300	0.288	0.110	0.120	0.748	0.573	1.183	0.576
2-step	2.880	1.890	0.356	0.315	0.568	0.480	0.878	0.723	0.948	0.576

NOTE: For each parameter, the left column comes from scenario 1, and the right column is from scenario 2.

Figure 2 shows the estimated number of groups in the left panels, together with ARIs in the right panels. When $h = 0$, the gaCRP method reduces to the traditional CRP method, which always tends to over-cluster and shows smaller ARI than results by LPML selection. From histograms in Figure 2 we see that, when h increases, the estimated number of groups first decreases and then increases. The reason is that as $h \to \infty$, the graphical weight w_{ij} for disconnected nodes becomes particularly small, which means only connected nodes can be classified into the same group, therefore, leading to the over-cluster problem, as we discussed in Remark 2. The group concordance under optimal h selected by LPML generally performs better than those of EM method, with the proportion of selecting true number of groups always greater than 80%. In addition, we also compare the group number estimation with the group panel data models (Liu et al. 2020), whose performance could be sensitive to the tuning parameter specification. We present the detailed analysis in Appendix C.2.

5. Empirical Case Study

In this section, we study the local government economic competition by using city-level fiscal revenue (FR) data. We first

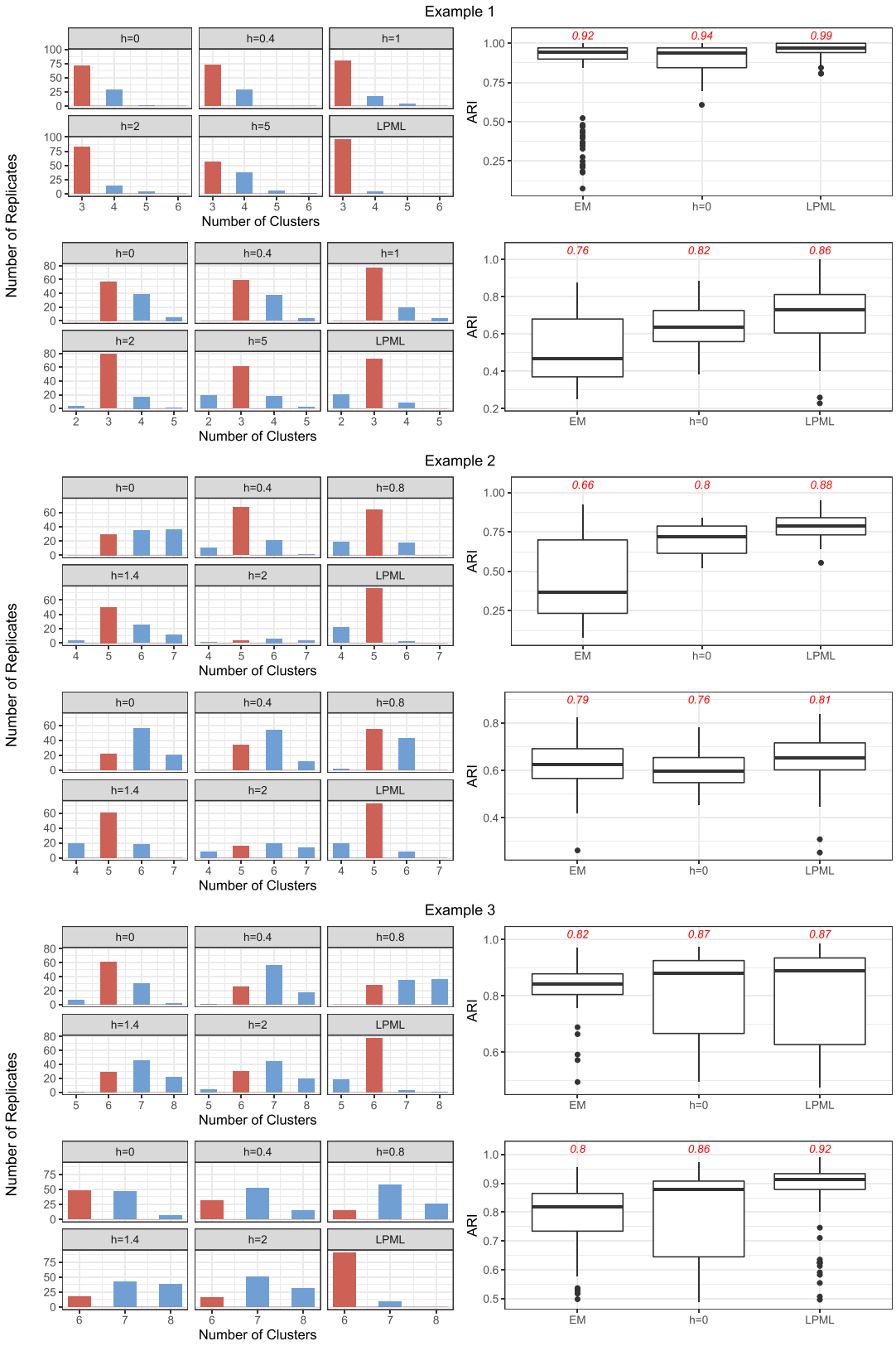

Figure 2. Grouping results of three examples, with results of scenario 1 in the first line and scenario 2 in the second line. Left panel: histogram of estimates of K under different h. Right panel: boxplots of ARI under EM method, $h = 0$ and optimal h. The red texts show the average accuracy of group memberships. The optimal hs selected by LPML are listed at the top of each example.

conduct a descriptive analysis of the real dataset. We calculate the average ratio of fiscal revenue to GDP across all cities and the yearly average ratio from 2005 to 2016. As shown in Figure 3, the average FR/GDP ratio is 0.067, and the ratio increases steadily with a little fluctuation. As the histogram shows, a slightly right skewed distribution pattern can be observed. The data

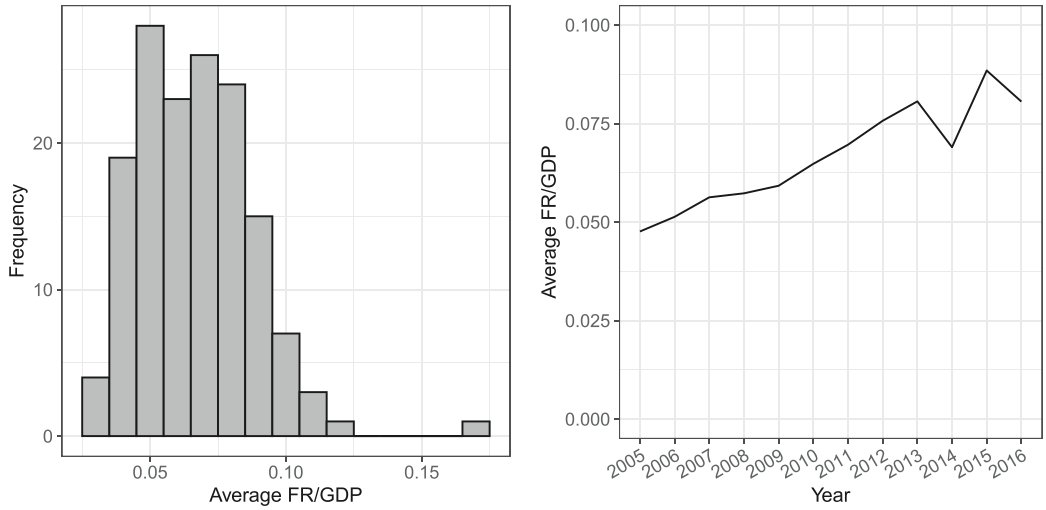

Figure 3. Histogram and line plot of $N = 151$ cities' yearly average fiscal FR/GDP.

cleaning procedure is included in the Appendix D. In addition, We also conduct another empirical application on stock return rate analysis. The details can be found in Appendix E.2.

5.1. Background and Data Description

In the study for the economic growth of China, regional government's economic strategies are shown to be closely related to the whole economic growth (Xu 2011; Yu, Zhou, and Zhu 2016). Consequently, it is important to study the regional government competitions within different time periods. In spatial economics, evidence has shown that the fiscal competition exists widely among the local government with its spatial adjacent neighbors (Cassette, Di Porto, and Foremny 2012; Janeba and Osterloh 2013; Parchet 2014; Agrawal 2015). We investigate how the city fiscal revenue relates to its spatial adjacent cities and its historical information. The model with space-time lags has been widely used on economic panel data research, including analysis about equation systems with time dynamics and spatial spillover effects in regional science (De Graaff, Van Oort, and Florax 2012; Gebremariam, Gebremedhin, and Schaeffer 2011), fiscal policy with government competition (Hauptmeier, Mittermaier, and Rincke 2012; Allers and Elhorst 2011), and many other related fields (Korniotis 2010; Brown and Laschever 2012). In addition, the latent group structure has been considered in recent macro-economic researches. For instance, Bonhomme and Manresa (2015) proposed the time-varying grouped patterns of heterogeneity in linear panel data models, and model the relationship between the degree of democracy and GDP. They found the group patterns exist across countries. A simple and fast approach was proposed by Liu et al. (2020) to identify and estimate the unknown group structure in panel data models, which is used to analyze the aggregate production function. They discovered both individual-level and group-level heterogeneity for women's labor force participation.

We collect $N = 151$ cities' public financial statements from *Fiscal Statistics of Cities and Counties in China* during the period 2005–2016. The yearbook is published by China Financial and Economic Publishing House, a state-owned press under the supervision of the Ministry of Finance of the People's Republic of China. It collects detailed information on city-level fiscal statistics, such as fiscal revenues and expenditures, fiscal accounting balances, transfer payments, and the fiscally supported population (Yu, Zhou, and Zhu 2016). In this empirical study, the response Y_{it} is the fiscal revenue divided by the local GDP. Following Zhang and Zou (1998), Devereux, Lockwood, and Redoano (2007), and Lv, Liu, and Li (2020), we consider four covariates, which are POP (population at the end of year), GDP1st (the proportion of primary industry to GDP), SAV (year-end savings of urban and rural residents), and FOR (actual foreign investment).

5.2. Model Estimation

We next implement the GAGNAR model (2.6) to the China fiscal revenue data. First, we conduct parameter estimation by using the data from 2005 to 2016. The smoothing parameter h is chosen from $\{0, 0.2, 0.4, \ldots, 2.0, 3.0, 4.0, 5.0\}$ and we set $\alpha = 1$. For each h, we run 1500 MCMC iterations and drop the first 500 as burn-in. The best h selected by LPML is 2, and the estimated number of groups is $K = 4$. The city group pattern under $h = 2$ is displayed in the top left panel of Figure 5. Cities in the first group are mainly from Guangdong, Jiangxi and Zhejiang, which are located in the south China, as well as Anhui province in the central south of China. The second group contains mainly the cities in Shandong province (east China) and Henan province (central east China), with the average FR/GDP at the lowest level as the top right panel of Figure 5 shows. Most cities in Liaoning, Jilin, Heilongjiang province and others constitute the third group, which are mostly in the Northeast China. Some cities in Jiangsu are in the fourth group, with the highest average FR/GDP values. According to the top right panel of Figure 5, cities in four groups exhibit different economic features, as the average FR/GDPs among four groups show different levels.

Table 3 reports the estimated parameters for each group under the optimal LPML criterion. As one could see, the network effect and momentum effect are different among four groups. Specifically, the cities in the second and the third group

Table 3. Parameters estimated by China Fiscal Revenue dataset with optimal $h = 2$.

N_k		Group 1 83	Group 2 38	Group 3 25	Group 4 5
$\widehat{\beta}_{0k}$		0.003 (0.001, 0.007)	0.037 (0.027, 0.049)	0.038 (0.023, 0.067)	0.047 (0.025, 0.163)
$\widehat{\beta}_{1k}$	Network effect	−0.011 (−0.026, 0.056)	0.101 (−0.004, 0.202)	0.212 (0.008, 0.421)	−0.179 (−0.757, 0.416)
$\widehat{\beta}_{2k}$	Momentum effect	0.996 (0.925, 1.001)	0.229 (0.101, 0.473)	0.345 (0.210, 0.547)	0.178 (−0.064, 0.502)
$\widehat{\gamma}_k$	POP	−0.001 (−0.004, 0.004)	−0.025 (−0.037, −0.012)	0.022 (−0.035, 0.019)	0.113 (−0.156, 0.054)
	GDP1st	−0.001 (−0.005, 0.012)	0.175 (0.048, 0.258)	−0.128 (−0.234, 0.211)	0.186 (−0.270, 0.420)
	SAV	0.008 (−0.020, 0.034)	−0.066 (−0.122, −0.013)	0.001 (−0.174, 0.070)	−0.074 (−0.798, 0.274)
	FOR	0.008 (−0.001, 0.013)	0.070 (−0.093, 0.149)	0.078 (−0.089, 0.243)	−0.383 (−0.337, 0.243)
$\widehat{\sigma}_k^2$		0.0001	0.0003	0.0006	0.0078

NOTE: The 95% HPDs of parameters estimated upon four specific cities from different groups are reported in the parentheses.

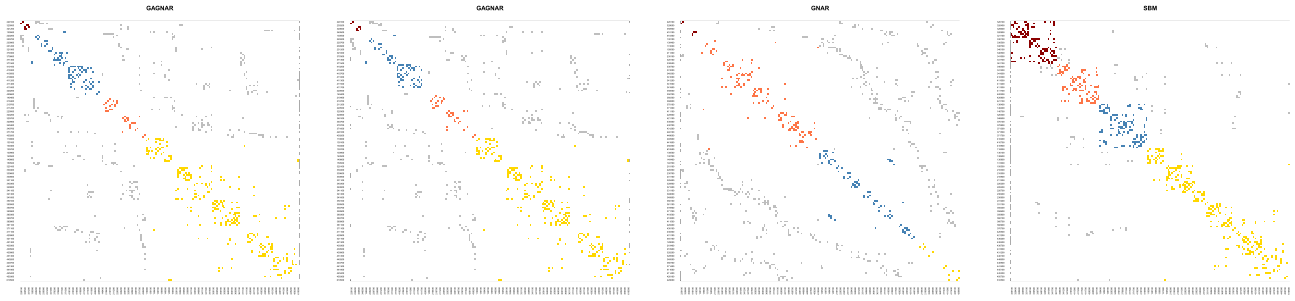

Figure 4. Adjacency matrices among $N = 151$ cities, where the colored grids mean that ith city is connected with jth city. In each subgraph, cities are ordered by the membership from highest to lowest within-group density, with darkred, blue, pink, and yellow representing for four groups. The gray grids correspond to between-group connections.

have positive network effects, which means there might exist imitation economic strategies for those cities with their spatial neighbors. On the contrary, cities in the first and the last group are negatively related to their neighbors. This implies that they tend to take opposite macroeconomic decisions to their adjacent cities. All of the four groups have positive momentum effects, which shows that the responses are positively related to their historical performances. Furthermore, the covariates tend to have different effects on the response. For example, the population has negative effect on the fiscal revenue for cities in the first and the second group while its influence is positive for other cities. The foreign investment, which represents the openness of a city, negatively affects the fiscal revenue for cities in the fourth group, while it is positive for other cities.

To summarize the posterior distribution of each parameter, the Highest Posterior Density (HPD) (Chen and Shao 1999; Kruschke 2014) interval that spans the majority of the posterior distribution is specified. We calculate the 95% HPD intervals of parameters of four typical cities (Beijing, Linyi, Fushun and Taizhou) from four groups using the R package *HDInterval*. The results are visualized in the middle left panel of Figure 5. Besides, we report the 95% HPD intervals for the corresponding four cities in the parentheses following each estimation in Table 3. To better understand the differences of parameter estimates among the groups, for each parameter (e.g., network effect) we collect the estimates of each city for the 1000 iterations after the burn-in stage. Then we pool the estimates of cities within the same group together, which leads to the boxplots in the bottom panel of Figure 5. Particularly here we show the boxplots of network effects and momentum effects for illustration. First, we can observe that the momentum effects are positive for most cities. In addition, we find that the network effects are highest for the Group 3 and lowest for Group 4. This implies a positive spillover effect of cities in Group 3 and a negative spillover effect for Group 4.

Subsequently, we compare the estimation results with other grouping methods, which are CRP (Pitman 1995), GNAR (Zhu and Pan 2020), and the SBM (Wang and Wong 1987) methods. Each method can output the estimated group memberships. For convenience, we visualize the adjacency matrix with each within-group connections marked in different colors. This leads to Figure 4, where the cities are reordered according to their memberships. We can observe that the estimated groups of SBM have the densest within-group connections, while the groups estimated by GNAR model do not exhibit such a clear pattern. This is mainly because that the SBM only uses the adjacency matrix for membership estimation while the GNAR model totally ignores the graph information but only using $\{Y_{it}\}$ for nodes clustering. The GAGNAR model and CRP model exhibit similar results and strike a balance between the SBM and GNAR model. To further confirm the idea, we calculate the within- and between-group densities for each model. The detailed result is presented in Appendix E.1.2. It shows that the GAGNAR model has slightly higher within-group density and lower between-group density than the CRP model. Furthermore, We present the boxplot of average responses of the cities clustered by SBM in Figure S6 of Appendix E.1.3. One can observe that the responses of group 2–4 are close to each other, which indicates the lack of ability of the SBM to distinguish the responses from different groups. We also include parameters estimated by the CRP and GNAR in Appendix E.1.4.

Lastly, we compare the prediction accuracy of our model with CRP, GNAR, NAR, and traditional time series models, namely AR and ARMA. Denote T_{train} and T_{test} as the time

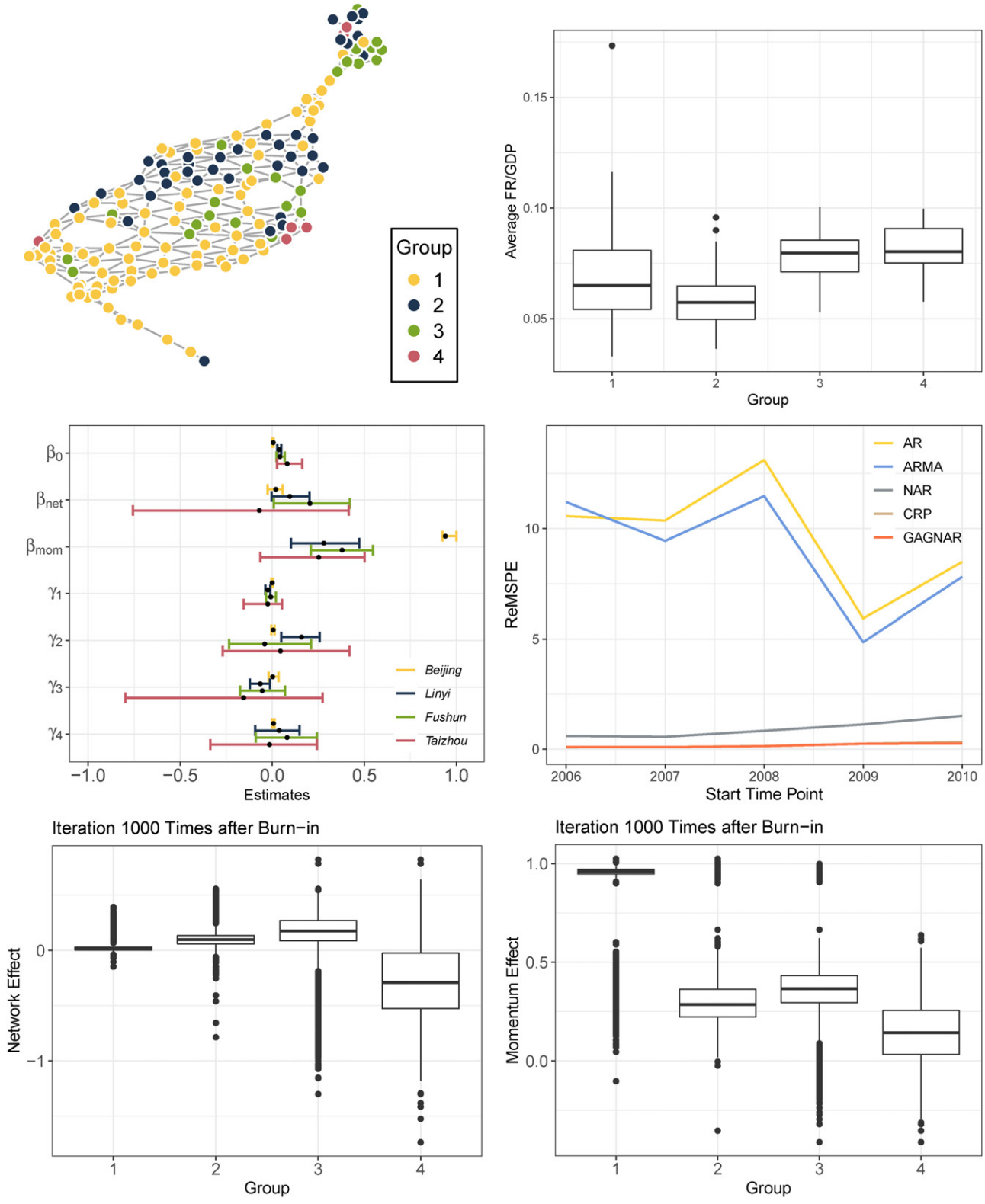

Figure 5. The top left panel shows the group pattern of cities estimated by GAGNAR under $h = 2$. The nodes in group 1–4 are marked in yellow, dark blue, green and pink. The top right panel shows the average FR/GDP of cities corresponding to four estimated groups. Middle left panel shows the 95% HPD intervals of four cities from different groups. The ReMSPE by rolling window scheme from different starting time points are displayed in the middle right panel. The bottom panel shows boxplots of the network effects and momentum effects, respectively, for four groups by pooling all estimates of cities in each group.

length of training set and testing set, respectively. To evaluate the prediction accuracy, we calculate the mean square prediction error $\text{MSPE} = (NT_{\text{test}})^{-1} \sum_{t=t_{\text{test},0}}^{t_{\text{test},0}+T_{\text{test}}} \sum_i (\widehat{Y}_{it} - Y_{it})^2$ where $\widehat{Y}_{it}$ is the predicted response for node i at time point t and $t_{\text{test},0}$ is the starting time point of the testing set. Define $\text{MSPE}_0 = (NT_{\text{test}})^{-1} \sum_{t=t_{\text{test},0}}^{t_{\text{test},0}+T_{\text{test}}} \sum_i (Y_{it} - \widehat{\mu}_{i,\text{train}})^2$ as the baseline MSPE,

Table 4. Predicting ReMSE of GAGNAR, CRP, GNAR, NAR, AR, ARMA under the rolling window setting.

Method	Start time point of testing set				
	2006	2007	2008	2009	2010
GAGNAR	0.0990	0.0973	0.1375	0.2468	0.2662
CRP	0.0952	0.0968	0.1435	0.2511	0.3238
GNAR	0.7249	0.6936	2.3679	5.0960	226.818
NAR	0.6045	0.5652	0.8372	1.1248	1.5153
AR	10.5640	10.3679	13.1174	5.9287	8.4905
ARMA	11.2059	9.4445	11.4805	4.8626	7.8136

where $\widehat{\mu}_{i,\text{train}} = T_{\text{train}}^{-1} \sum_{t=t_{\text{test},0}-T_{\text{train}}}^{t_{\text{test},0}-1} Y_{it}$ as the mean response of the ith node in the training set. The relative prediction error is defined as

$$\text{ReMSPE} = \text{MSPE}/\text{MSPE}_0, \tag{5.1}$$

which is used to evaluate the model performance. To illustrate the superior prediction performance of GAGNAR, a rolling window approach is adopted. First, we set the training data with $T_{\text{train}} = 7$ years. Then, the following $T_{\text{test}} = 7$ years are used for prediction evaluation. Subsequently, we make the prediction for every 1 year in a rolling window approach, which yields the middle right panel in Figure 5. As one could observe, the GAGNAR model (orange line) is able to achieve obviously higher prediction accuracy than the NAR, AR, and ARMA models. Although the performances of the GAGNAR model and CRP model are comparable, the GAGNAR model is generally better and stable in prediction. For instance, for prediction starting from the year 2008, the ReMSPE of GAGNAR model is 0.1375, which is lower than the second best model (i.e., the CRP model) with ReMSPE= 0.1435, and much better than others. Since the ReRMSE of GNAR is very large of the year 2010, we eliminate GNAR model in Figure 5. The detailed prediction results could be found in Table 4. This indicates that the node group memberships could be informative for the prediction work.

6. Discussion

In this article, we propose a novel Bayesian nonparametric grouping approach for learning the heterogeneity of dynamic response for network data. The proposed method could conduct group-specific parameter estimation, estimate the number of groups, and infer group configurations simultaneously. Building upon the gaCRP framework, we develop a collapsed Gibbs sampler for an efficient Bayesian inference. Specifically, we adopt the Dahl's method for post MCMC inference and introduce LPML for smoothing parameter selection. Numerical results have confirmed that the proposed method is able to simultaneously infer the number of groups and the group-wise parameters with high accuracy. Comparing to the traditional techniques such as EM algorithm and two-step approach in Zhu and Pan (2020), the proposed method is able to improve the grouping performance, especially when grouping configurations contain certain graphical information. Lastly, we illustrate the usefulness of the proposed method by studying two real data examples, including the China fiscal revenue analysis and the stock return prediction. The results indicate that incorporating graphical information on grouping process for dynamic observations will improve the interpretability and prediction power of the methodology.

A few topics are worth further investigation. A natural extension is designing an efficient algorithm for large scale networks with lower computational complexity. Two promising solutions for tackling this challenge are discovering low rank structure and imposing sparsity. Second, the proposed method is restricted to continuous responses and thus can be extended to binary or counted data. Finally, incorporating the GNAR model with the stochastic block model (SBM) will be another interesting topic for future study.

Supplementary Materials

All technique proofs, additional numerical results as well as another empirical application are given in the supplementary material.

Funding

Yimeng Ren and Xuening Zhu are supported by the National Natural Science Foundation of China (nos. 71991472, 72222009, 11901105, U1811461). Xiaoling Lu is supported by the National Natural Science Foundation of China (nos. 72171229). Guanyu Hu is supported by the NSF Grant BCS-2152822 and DMS-2210371.

ORCID

Xuening Zhu http://orcid.org/0000-0001-5824-5279
Guanyu Hu http://orcid.org/0000-0001-5824-5279

References

Agrawal, D. R. (2015), "The Tax Gradient: Spatial Aspects of Fiscal Competition," *American Economic Journal: Economic Policy*, 7, 1–29. [58]

Allers, M. A., and Elhorst, J. P. (2011), "A Simultaneous Equations Model of Fiscal Policy Interactions," *Journal of Regional Science*, 51, 271–291. [58]

Ando, T., and Bai, J. (2016), "Panel Data Models with Grouped Factor Structure under Unknown Group Membership," *Journal of Applied Econometrics*, 31, 163–191. [49,50,51]

Bester, C. A., and Hansen, C. B. (2016), "Grouped Effects Estimators in Fixed Effects Models," *Journal of Econometrics*, 190, 197–208. [50]

Blei, D. M., and Frazier, P. I. (2011), "Distance Dependent Chinese Restaurant Processes," *Journal of Machine Learning Research*, 12, 2461–2488. [52]

Bonhomme, S., and Manresa, E. (2015), "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83, 1147–1184. [50,58]

Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G. (2009), "Network Analysis in the Social Sciences," *Science*, 323, 892–895. [49]

Borsboom, D., and Cramer, A. O. (2013), "Network Analysis: An Integrative Approach to the Structure of Psychopathology," *Annual Review of Clinical Psychology*, 9, 91–121. [49]

Brown, K. M., and Laschever, R. A. (2012), "When they're Sixty-Four: Peer Effects and the Timing of Retirement," *American Economic Journal: Applied Economics*, 4, 90–115. [58]

Cao, Q., Liang, Y., and Niu, X. (2017), "China's Air Quality and Respiratory Disease Mortality based on the Spatial Panel Model," *International Journal of Environmental Research and Public Health*, 14, 1081. [55]

Cassette, A., Di Porto, E., and Foremny, D. (2012), "Strategic Fiscal Interaction across Borders: Evidence from French and German Local Governments along the Rhine Valley," *Journal of Urban Economics*, 72, 17–30. [58]

Chen, M.-H., and Shao, Q.-M. (1999), "Monte Carlo Estimation of Bayesian Credible and HPD Intervals," *Journal of Computational and Graphical Statistics*, 8, 69–92. [59]

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2012), *Monte Carlo methods in Bayesian Computation*, New York: Springer. [54]

Chen, Y., Wang, X., Bu, J., Tang, B., and Xiang, X. (2016), "Network Structure Exploration in Networks with Node Attributes," *Physica A: Statistical Mechanics and its Applications*, 449, 240–253. [52]

Chuang, H.-Y., Lee, E., Liu, Y.-T., Lee, D., and Ideker, T. (2007), "Network-based Classification of Breast Cancer Metastasis," *Molecular Systems Biology*, 3, 140. [49]

Cramer, A. O., Waldorp, L. J., Van Der Maas, H. L., and Borsboom, D. (2010), "Comorbidity: A Network Perspective," *Behavioral and Brain Sciences*, 33, 137–150. [49]

Dahl, D. B. (2006), "Model-based Clustering for Expression Data via a Dirichlet Process Mixture Model," *Bayesian Inference for Gene Expression and Proteomics*, 4, 201–218. [54]

De Graaff, T., Van Oort, F. G., and Florax, R. J. (2012), "Regional Population–Employment Dynamics Across Different Sectors of the Economy," *Journal of Regional Science*, 52, 60–84. [58]

Devereux, M. P., Lockwood, B., and Redoano, M. (2007), "Horizontal and Vertical Indirect Tax Competition: Theory and Some Evidence from the USA," *Journal of Public Economics*, 91, 451–479. [58]

Gao, J., Xia, K., and Zhu, H. (2020), "Heterogeneous Panel Data Models with Cross-Sectional Dependence," *Journal of Econometrics*, 219, 329–353. [50]

Gebremariam, G. H., Gebremedhin, T. G., and Schaeffer, P. V. (2011), "Employment, Income, and Migration in Appalachia: A Spatial Simultaneous Equations Approach," *Journal of Regional Science*, 51, 102–120. [58]

Gelfand, A. E., Dey, D. K., and Chang, H. (1992), "Model Determination using Predictive Distributions with Implementation via Sampling-based Methods," Technical report, STANFORD UNIV CA DEPT OF STATISTICS. [54]

Geng, J., Bhattacharya, A., and Pati, D. (2019), "Probabilistic Community Detection with Unknown Number of Communities," *Journal of the American Statistical Association*, 114, 893–905. [50]

Geng, L., and Hu, G. (2022), "Bayesian Spatial Homogeneity Pursuit for Survival Data with an Application to the Seer Respiratory Cancer Data," *Biometrics*, 78, 536–547. [50,52]

Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732. [51,52]

Guda, P., Chittur, S. V., and Guda, C. (2009), "Comparative Analysis of Protein–Protein Interactions in Cancer-Associated Genes," *Genomics, Proteomics & Bioinformatics*, 7, 25–36. [49]

Guðmundsson, G. S., and Brownlees, C. (2021), "Detecting Groups in Large Vector Autoregressions," *Journal of Econometrics*, 225, 2–26. [49,50,51]

Hanneman, R. A., and Riddle, M. (2005), *Introduction to Social Network Methods*, Riverside, CA: University of California, Riverside. [49]

Härdle, W. K., Wang, W., and Yu, L. (2016), "Tenet: Tail-Event Driven Network Risk," *Journal of Econometrics*, 192, 499–513. [49]

Hauptmeier, S., Mittermaier, F., and Rincke, J. (2012), "Fiscal Competition over Taxes and Public Inputs," *Regional Science and Urban Economics*, 42, 407–419. [58]

Horvath, S. (2011), *Weighted Network Analysis: Applications in Genomics and Systems Biology*, New York: Springer. [49]

Hu, J., Qin, H., Yan, T., and Zhao, Y. (2020), "Corrected Bayesian Information Criterion for Stochastic Block Models," *Journal of the American Statistical Association*, 115, 1771–1783. [50]

Hubert, L., and Arabie, P. (1985), "Comparing Partitions," *Journal of Classification*, 2, 193–218. [56]

Ibrahim, J. G., Chen, M.-H., Sinha, D., Ibrahim, J., and Chen, M. (2001), *Bayesian Survival Analysis* (Vol. 2), New York: Springer. [54]

Ishwaran, H., and Zarepour, M. (2002a), "Dirichlet Prior Sieves in Finite Normal Mixtures," *Statistica Sinica*, 12, 941–963. [52,53]

——— (2002b), "Exact and Approximate Sum Representations for the Dirichlet Process," *Canadian Journal of Statistics*, 30, 269–283. [51]

Janeba, E., and Osterloh, S. (2013), "Tax and the City'a Theory of Local Tax Competition," *Journal of Public Economics*, 106, 89–100. [58]

Ke, Z. T., Fan, J., and Wu, Y. (2015), "Homogeneity Pursuit," *Journal of the American Statistical Association*, 110, 175–194. [49]

Korniotis, G. M. (2010), "Estimating Panel Models with Internal and External Habit Formation," *Journal of Business and Economic Statistics*, 28, 145–158. [58]

Kruschke, J. (2014), *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, Academic Press. [59]

Lee, L.-f., and Yu, J. (2009), "Spatial Nonstationarity and Spurious Regression: The Case with a Row-Normalized Spatial Weights Matrix," *Spatial Economic Analysis*, 4, 301–327. [49]

Lei, J., and Rinaldo, A. (2015), "Consistency of Spectral Clustering in Stochastic Block Models," *The Annals of Statistics*, 43, 215–237. [50]

Leung, A. C. M., Agarwal, A., Konana, P., and Kumar, A. (2017), "Network Analysis of Search Dynamics: The Case of Stock Habitats," *Management Science*, 63, 2667–2687. [49]

Li, K., Cui, G., and Lu, L. (2020), "Efficient Estimation of Heterogeneous Coefficients in Panel Data Models with Common Shocks," *Journal of Econometrics*, 216, 327–353. [50]

Liu, R., Shang, Z., Zhang, Y., and Zhou, Q. (2020), "Identification and Estimation in Panel Models with Overspecified Number of Groups," *Journal of Econometrics*, 215, 574–590. [50,51,56,58]

Liu, X., Patacchini, E., and Rainone, E. (2017), "Peer Effects in Bedtime Decisions among Adolescents: A Social Network Model with Sampled Data," *The Econometrics Journal*, 20, S103–S125. [49]

Lu, J., Li, M., and Dunson, D. (2018), "Reducing Over-Clustering via the Powered Chinese Restaurant Process," arXiv preprint arXiv:1802.05392. [52]

Lv, B., Liu, Y., and Li, Y. (2020), "Fiscal Incentives, Competition, and Investment in China," *China Economic Review*, 59, 101371. [58]

Ma, S., Su, L., and Zhang, Y. (2021), "Determining the Number of Communities in Degree-Corrected Stochastic Block Models," *Journal of Machine Learning Research*, 22, 1–63. [50]

Marin, J.-M., Mengersen, K., and Robert, C. P. (2005), "Bayesian Modelling and Inference on Mixtures of Distributions," *Handbook of Statistics*, 25, 459–507. [51]

Miller, J. W., and Harrison, M. T. (2018), "Mixture Models with a Prior on the Number of Components," *Journal of the American Statistical Association*, 113, 340–356. [52]

Newman, M. E., and Leicht, E. A. (2007), "Mixture Models and Exploratory Analysis in Networks," *Proceedings of the National Academy of Sciences*, 104, 9564–9569. [52]

Nowicki, K., and Snijders, T. A. B. (2001), "Estimation and Prediction for Stochastic Blockstructures," *Journal of the American Statistical Association*, 96, 1077–1087. [55]

O'Malley, A. J., and Marsden, P. V. (2008), "The Analysis of Social Networks," *Health Services and Outcomes Research Methodology*, 8, 222–269. [49]

Parchet, R. (2014), "Are Local Tax Rates Strategic Complements or Strategic Substitutes?" *American Economic Journal: Economic Policy*, 11, 189–224. [58]

Pettit, L. (1990), "The Conditional Predictive Ordinate for the Normal Distribution," *Journal of the Royal Statistical Society*, Series B, 52, 175–184. [54]

Pitman, J. (1995), "Exchangeable and Partially Exchangeable Random Partitions," *Probability Theory and Related Fields*, 102, 145–158. [50,59]

Rand, W. M. (1971), "Objective Criteria for the Evaluation of Clustering Methods," *Journal of the American Statistical Association*, 66, 846–850. [56]

Rohe, K., Chatterjee, S., and Yu, B. (2011), "Spectral Clustering and the High-Dimensional Stochastic Blockmodel," *The Annals of Statistics*, 39, 1878–1915. [50]

Sewell, D. K., and Chen, Y. (2017), "Latent Space Approaches to Community Detection in Dynamic Networks," *Bayesian Analysis*, 12, 351–377. [50]

Shi, W., and Lee, L.-f. (2017), "Spatial Dynamic Panel Data Models with Interactive Fixed Effects," *Journal of Econometrics*, 197, 323–347. [49]

Simmel, G. (1950), *The Sociology of Georg Simmel* (Vol. 92892), New York: Simon and Schuster. [49]

Sojourner, A. (2013), "Identification of Peer Effects with Missing Peer Data: Evidence from Project Star," *The Economic Journal*, 123, 574–605. [49]

Su, L., Shi, Z., and Phillips, P. C. (2016), "Identifying Latent Structures in Panel Data," *Econometrica*, 84, 2215–2264. [49,50]

van der Pas, S., and van der Vaart, A. (2018), "Bayesian Community Detection," *Bayesian Analysis*, 13, 767–796. [50]

Von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., and Bork, P. (2002), "Comparative Assessment of Large-Scale Data Sets of Protein–Protein Interactions," *Nature*, 417, 399–403. [49]

Wang, Y. J., and Wong, G. Y. (1987), "Stochastic Blockmodels for Directed Graphs," *Journal of the American Statistical Association*, 82, 8–19. [55,59]

Wasserman, S., and Faust, K. (1994), *Social Network Analysis: Methods and Applications*, Cambridge: Cambridge University Press. [49]

Wu, G., Feng, X., and Stein, L. (2010), "A Human Functional Protein Interaction Network and its Application to Cancer Data Analysis," *Genome Biology*, 11, 1–23. [49]

Xu, C. (2011), "The Fundamental Institutions of China's Reforms and Development," *Journal of Economic Literature*, 49, 1076–1151. [58]

Yu, J., Zhou, L.-A., and Zhu, G. (2016), "Strategic Interaction in Political Competition: Evidence from Spatial Effects across Chinese Cities," *Regional Science and Urban Economics*, 57, 23–37. [58]

Zhang, J., Chang, Y., Zhang, L., and Li, D. (2018), "Do Technological Innovations Promote Urban Green Development?—A Spatial Econometric Analysis of 105 Cities in China," *Journal of Cleaner Production*, 182, 395–403. [55]

Zhang, T., and Zou, H.-f. (1998), "Fiscal Decentralization, Public Spending, and Economic Growth in China," *Journal of Public Economics*, 67, 221–240. [58]

Zhao, Y., Levina, E., and Zhu, J. (2012), "Consistency of Community Detection in Networks Under Degree-Corrected Stochastic Block Models," *The Annals of Statistics*, 40, 2266–2292. [50,55]

Zhu, X., and Pan, R. (2020), "Grouped Network Vector Autoregression," *Statistica Sinica*, 30, 1437–1462. [49,50,51,56,59,61]

Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H. (2017), "Network Vector Autoregression," *The Annals of Statistics*, 45, 1096–1123. [49,51]

Zhu, X., Wang, W., Wang, H., and Härdle, W. K. (2019), "Network Quantile Autoregression," *Journal of Econometrics*, 212, 345–358. [49]

Zou, T., Lan, W., Wang, H., and Tsai, C.-L. (2017), "Covariance Regression Analysis," *Journal of the American Statistical Association*, 112, 266–281. [49]