Statistica Sinica Preprint No: SS-2024-0435	
Title	Mixed Membership Network with the Autoregressive Structure
Manuscript ID	SS-2024-0435
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0435
Complete List of Authors	Tianyi Sun, Bo Zhang, Baisuo Jin and Yuehua Wu
Corresponding Authors	Bo Zhang
E-mails	zhangbo890301@outlook.com
Notice: Accepted author version.	

Statistica Sinica

MIXED MEMBERSHIP NETWORK WITH THE AUTOREGRESSIVE STRUCTURE

Tianyi Sun[1], Bo Zhang[1*], Baisuo Jin[2,1], and Yuehua Wu[3]

[1] *University of Science and Technology of China,* [2] *Xinjiang Normal University and* [3] *York University*

Abstract: Community network research has received extensive attention, and the study of dynamic random block networks is one of the emerging directions of network analysis. This paper first proposes a mixed membership autoregressive network model with a first-order autoregressive structure, then discusses the estimation of the membership matrix and community detection, and gives an AR-1 mixed spectral clustering algorithm for this model to estimate the membership matrix for community detection. In addition, an empirical eigenvalue threshold estimator is introduced to estimate the number of communities. Simulation results show that our method shows stronger generalization ability than previous methods for both random block community detection problems and mixed member community detection problems. Under different settings, the explicit error rate of our method does not exceed that of previous methods, and in many cases our method performs better. Application to real data proves the effectiveness of our method.

Key words and phrases: Community detection; Empirical eigenvalue-threshold estimator; First-order autoregressive networks; Mixed membership stochastic block model; AR-1 mixed spectral clustering algorithm.

1. Introduction

The study of network data covers various disciplines (Easley et al., 2010; Kolaczyk and Csárdi, 2014), including social networks in social sciences (Lee et al., 2019; Tian et al., 2024), trade networks in statistics (Goldenberg et al., 2010)), the Internet and information networks in computer science (Newman, 2004), the Internet of Things in industry (Majid et al., 2022; Ben-Daya et al., 2019), and genetic networks and cell networks in biology (Barabási and Oltvai, 2004). In these empirical network datasets, nodes usually represent research objects, such as individuals or web pages, and edges represent connections between them. In the past few years, the analysis and modeling of network data have made rapid progress. The study of networks containing community structures is an important direction of network research. It explores the internal structure of the network and considers the homogeneity between nodes and objects by assuming that there are communities or groups in the network. Such assumptions are very common in real network data, and corresponding research has been applied to some practical problems (Newman, 2003; Girvan and Newman, 2002).

In the study of community structured networks, the most commonly used model is the stochastic block model (SBM) proposed in Holland and Leinhardt (1981) and Holland et al. (1983), which is based on a random graph model and assumes that nodes in the network belong to different communities and members in the same community have equal connection probability. SBM has been continuously studied since its proposal (Bickel et al., 2013; Lei and Rinaldo, 2015; Rohe et al., 2011), and those interested can refer to Abbe (2017)

for some of the latest progress. In addition, researchers have established some extensions based on SBM. To solve the problem of node heterogeneity under the framework of community structures, Karrer and Newman (2011) established the degree-corrected stochastic block model (DCSBM), and some subsequent studies gave theoretical results of DCSBM including community estimation error and consistency (Zhao et al., 2012; Chen et al., 2018; Jin, 2015; Ma et al., 2021). Cai and Li (2015) introduced the generalized stochastic block model (GSBM) that allows adversarial outliers. Airoldi et al. (2008) proposed the mixed membership stochastic block model (MMSBM), which allows each node in the network to have mixed membership from different communities. Mao et al. (2021) and Anandkumar et al. (2014) developed effective estimation algorithms for membership in MMSBM. The study of community structure networks also includes other types of models, such as graphon (Chatterjee, 2015; Xu, 2018), latent position clustering models (Handcock et al., 2007; Hoff, 2007), and graph root distribution parameterization (Lei, 2021).

Dynamic networks are another focus in the field of network research, and their modeling requires simulating the changes of networks over time. Research in this field has been developing, and several models have been proposed, including Markov chain network models (Ludkin et al., 2018), exponential random graph models (Krivitsky and Handcock, 2014), and the latent process-based models (Durante and Dunson, 2016; Matias and Miele, 2017)). Evolutionary network analysis based on network snapshots (Donnat and Holmes, 2018) has also been applied to this problem, which can reflect the dynamic changes of the network,

but it is difficult to give the principles behind it.

When considering dynamic networks that incorporate community structures, most past studies (Bhattacharjee et al., 2020; Pensky, 2019) have assumed that networks observed over time are independent of each other and have used static connection probabilities to model network characteristics. Jiang et al. (2023) creatively incorporated autoregressive structures into a dynamic SBM. By replacing connection probabilities with transition probabilities, the new model explicitly represents dynamic dependencies in the network data, and it also simplifies additional inferences such as model diagnostic checks. However, their study only addresses networks where nodes belong to a single community, and their estimation method assumes that the number of communities in the network is known, which is often not satisfied in real-world data.

Community detection has always been the focus of SBM. Past approaches to this problem include proportional cutting (Hagen and Kahng, 1992), regularized cutting (Shi and Malik, 2000), convex optimization (Cai and Li, 2015), Newman-Girvan modularity (Amini et al., 2013), etc. Among them, the spectral clustering algorithm (Von Luxburg, 2007) is the most widely used method. The estimation of the number of communities is also a valuable research direction, and most of the past methods use Akaike Information Criterion (AIC) (Akaike, 1973) or Bayesian Information Criterion (BIC) (Schwarz, 1978) for estimation.

In this paper, inspired by Jiang et al. (2023), we extend the autoregressive structure to MMSBM, i.e., we allow the nodes in the dynamic autoregressive network to have different

degrees of membership in multiple communities. Such a setting expands the community structure of the network while incorporating time series information, making it closer to many network forms in real life. We propose a new algorithm based on spectral clustering, in which the static connection probabilities are replaced by transition probabilities, and a projection step is added to estimate the mixed membership of dynamic networks. We also introduce a new empirical eigenvalue threshold estimator and compare its effect with past methods. Thus, we can study the community detection and community number estimation based on our model. Empirical results show that compared with the method in Jiang et al. (2023), the communities detected by our method are more stable with changes in sample size.

This paper is organized as follows: Section 2 introduces the framework of our proposed mixed membership autoregressive network model. Section 3 introduces the AR-1 mixed spectral clustering algorithm for membership estimation and the empirical eigenvalue threshold estimator. Section 4 gives theoretical results of our proposed method. Section 5 shows that our method outperforms previous methods on simulated data, while Section 6 presents the performance of the new method on real datasets. Section 7 concludes this paper.

The following notations will be used in subsequent sections: For any matrix $\boldsymbol{M}$, we use $\boldsymbol{M}(i,:)$ and $\boldsymbol{M}(:,j)$ to denote the ith row and jth column of the matrix $\boldsymbol{M}$, respectively. $\boldsymbol{M}(S,:)$ and $\boldsymbol{M}(:,S)$ define the submatrices consisting of rows and columns in the set S of the matrix $\boldsymbol{M}$, respectively. $i:j$ denotes the set of indices from i to j. We use $\|\boldsymbol{M}\|_F$ to denote

the Frobenius norm of the matrix $\boldsymbol{M}$, and use $\|v\|_1$ and $\|v\|_2$ to denote the Manhattan norm and Euclidean norm of the vector v, respectively. We denote $[\boldsymbol{X}|\boldsymbol{Y}]$ as the concatenation of the columns of the matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$.

2. Mixed membership autoregressive network model

2.1 Model

Let $\{\boldsymbol{X}_t = (X^t_{i,j})_{p\times p}, t = 0, 1, 2...n\}$ be a network process with p fixed nodes, where for each t, $\boldsymbol{X}_t$ is the $p \times p$ adjacency matrix of an undirected, unweighted, and self-loop-free network. We assume that all networks are Erdös-Renyi, i.e., $X^t_{ij}, 1 \le i < j \le p$ are independent and take the value 1 or 0, i.e., $X^t_{ij} = X^t_{ji} = 1$ when there is an edge between nodes i and j at time t, and $X^t_{ij} = X^t_{ji} = 0$ otherwise.

In light of Jiang et al. (2023), we assume that there is a first-order autoregressive structure in the network process such that

$$X^t_{ij} = X^{t-1}_{ij} I(\varepsilon^t_{i,j} = 0) + I(\varepsilon^t_{i,j} = 1), \quad t \ge 1, \tag{2.1}$$

where

$$P(\varepsilon^t_{i,j} = 1) = \alpha_{i,j}, \quad P(\varepsilon^t_{i,j} = -1) = \beta_{i,j}, \quad P(\varepsilon^t_{i,j} = 0) = 1 - \alpha_{i,j} - \beta_{i,j}. \tag{2.2}$$

Here, $\alpha_{i,j}$ and $\beta_{i,j}$ are fixed constants in [0,1], satisfying $\alpha_{i,j}+\beta_{i,j} \le 1$. Under this assumption, $\alpha_{i,j}$ can be understood as the probability of connection between nodes i and j at any time t when there is no edge between nodes i and j at time $t-1$. Similarly, $1-\beta_{i,j}$ is the probability

of connection between nodes i and j at any time t when there is an edge at time $t-1$. This fact ensures that X_{ij}^t is binary and $\{\boldsymbol{X}_t,\ t=0,1,2...n\}$ is a homogeneous Markov chain.

In addition, assume that there are K communities $V_1, V_2...V_K$ in the network, and each node belongs to at least one of them, where K is known. For mixed-membership networks, a node can belong to one or more communities. We assume that node i belongs to a community (also called a cluster) V_k, whose membership degree $\theta_i(k)$ satisfies $0 \leq \theta_i(k) \leq 1$ for $1 \leq k \leq K, 1 \leq i \leq p$, and $\sum_{k=1}^{K} \theta_i(k) = 1$.

The membership matrix is denoted as a $p \times K$ matrix $\boldsymbol{\Theta}$, where $\boldsymbol{\theta}_i = (\theta_i(1), \theta_i(2), \ldots, \theta_i(K))^\top$ is its ith row. Therefore, the matrix $\boldsymbol{\Theta}$ is considered as a mixed membership matrix containing node and community relationships.

For simplicity, we follow the literature (Airoldi et al., 2008; Mao et al., 2021; Anandkumar et al., 2014) to define pure and mixed nodes. For each pure node i, it belongs to a unique degree-1 community, that is, there exists $1 \leq k \leq K$ such that $\theta_i(k) = 1$, and all other elements of $\boldsymbol{\theta}_i$ are zero. Otherwise, the node is a mixed node. A convenient assumption is that nodes belonging to the same community share the same connection probability. Therefore, the connection probabilities between nodes can be simplified to the connection probabilities between communities. Let the connection probabilities between communities be defined as

$$\boldsymbol{B}_1 = (\zeta_{k,\ell})_{K\times K}, \quad \boldsymbol{B}_2 = (\eta_{k,\ell})_{K\times K},$$

then, the connection probabilities between nodes, i.e. $\alpha_{i,j}$ and $\beta_{i,j}$ satisfy the following

conditions

$$\alpha_{i,j} = \sum_{k=1}^{K}\sum_{l=1}^{K}\theta_i(k)\theta_j(l)\zeta_{k,l}, \quad \beta_{i,j} = \sum_{k=1}^{K}\sum_{l=1}^{K}\theta_i(k)\theta_j(l)\eta_{k,l}. \tag{2.3}$$

It is easy to find that, as a special case, if nodes i and j are both pure nodes, the connection probabilities are

$$\alpha_{i,j} = \zeta_{k,\ell}, \quad \beta_{i,j} = \eta_{k,\ell}, \quad i \in V_k,\ j \in V_l.$$

Thus, if we define the connection probability matrices between nodes as

$$\boldsymbol{P}_1 = (\alpha_{i,j})_{p\times p} \quad \text{and} \quad \boldsymbol{P}_2 = (\beta_{i,j})_{p\times p},$$

then we obtain that

$$\boldsymbol{P}_1 = \boldsymbol{\Theta}\boldsymbol{B}_1\boldsymbol{\Theta}^\top \quad \text{and} \quad \boldsymbol{P}_2 = \boldsymbol{\Theta}\boldsymbol{B}_2\boldsymbol{\Theta}^\top. \tag{2.4}$$

The core of the community detection problem is to use the adjacency matrix $\{\boldsymbol{X}_t\}$ to estimate the unknown mixed membership matrix $\boldsymbol{\Theta}$, which is equivalent to discovering the latent community structure of the network. However, before exploring the structure of the matrix, we need to use the adjacency matrix to estimate the connection probabilities $\alpha_{i,j}$ and β_{ij}. We directly use the results of Jiang et al. (2023), where $\alpha_{i,j}$ and $\beta_{i,j}$ can be estimated by

$$\hat{\alpha}_{i,j} = \frac{\sum_{t=1}^{n} X_{i,j}^{t}(1 - X_{i,j}^{t-1})}{\sum_{t=1}^{n}(1 - X_{i,j}^{t-1})}, \quad \hat{\beta}_{i,j} = \frac{\sum_{t=1}^{n}(1 - X_{i,j}^{t})X_{i,j}^{t-1}}{\sum_{t=1}^{n} X_{i,j}^{t-1}}. \tag{2.5}$$

We make the following assumptions.

Assumption 1. There exists a constant $l > 0$ such that $\alpha_{i,j}, \beta_{i,j} \geq l$ and $\alpha_{i,j} + \beta_{i,j} \leq 1$ for all $1 \leq i, j \leq p$.

Assumption 2. $n, p \to \infty$, and $(\log n)(\log\log n)\sqrt{\log p/n} \to 0$.

Jiang et al. (2023) show that under Assumption 1 and Assumption 2, the error of the estimator of (2.5) is $O_p\left(\sqrt{\log p/n}\right)$ when $n, p \to \infty$, specific conclusion can be found in the supplementary material. Such a conclusion is the basis for our subsequent research.

2.2 Latent community structures

We first give some theoretical results for identifying the latent community structure. In our mixed membership autoregressive network model, from equations (2.3) and (2.4), we can find that both $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ contain information about the membership matrix Θ. Therefore, we need to use both of them in the exploration of the latent community structure of the network.

In Jiang et al. (2023), the normalized Laplacian matrix is expected to recover the latent community structure, which is given below. Let $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ be two $p \times p$ diagonal matrices with $d_{i,1}, d_{i,2}$ being their (i,i)th elements, where $d_{i,1} = \sum_{j=1}^{p} \alpha_{i,j}$ and $d_{i,2} = \sum_{j=1}^{p}(1-\beta_{i,j})$. Then, the normalized Laplacian matrices are defined as

$$\boldsymbol{L}_1 = \boldsymbol{D}_1^{-1/2}\boldsymbol{P}_1\boldsymbol{D}_1^{-1/2}, \quad \boldsymbol{L}_2 = \boldsymbol{D}_2^{-1/2}(\mathbf{1}_{n\times n} - \boldsymbol{P}_2)\boldsymbol{D}_2^{-1/2}, \quad \boldsymbol{L} = \boldsymbol{L}_1 + \boldsymbol{L}_2, \tag{2.6}$$

where $\mathbf{1}_{n\times n}$ represents an $n \times n$ matrix with all elements being 1.

However, when we apply the above approach to our mixed membership autoregressive network model, we find that it may cause some problems. The first problem is that $\boldsymbol{L}_1$ and

$\boldsymbol{L}_2$ may be unreliable to recover the latent community structure. In $\boldsymbol{L}_1$ (similar problems also occur for $\boldsymbol{L}_2$), the community structure information contained in $\boldsymbol{P}_1$ may not be fully preserved because there may be more than K different values in the diagonal matrix $\boldsymbol{D}_1$. In short, it is not guaranteed that we can find a $K \times K$ matrix $\boldsymbol{\Psi}$ that satisfies $\boldsymbol{L}_1 = \boldsymbol{\Theta}\boldsymbol{\Psi}\boldsymbol{\Theta}^\top$. Therefore, it is no longer appropriate to use $\boldsymbol{L}_1$ to obtain the community structure. We provide a specific example in the supplementary material for further illustration.

On the other hand, in the Jiang et al. (2023) algorithm, the authors estimated the sum $\boldsymbol{L}$ of the normalized Laplacian matrices $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ and attempt to discover community structure through the structural information contained in $\boldsymbol{L}$. However, because $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ contain information about $\alpha_{i,j}$ and $1-\beta_{i,j}$, respectively, we find that this approach can lead to large errors in simulation results when $\alpha_{i,j}$ and $\beta_{i,j}$ are close for any i and j. Similarly, let $\boldsymbol{L}_3 = \boldsymbol{D}_2^{-1/2}\boldsymbol{P}_2\boldsymbol{D}_2^{-1/2}$. When $\alpha_{i,j}$ and $\beta_{i,j}$ are close, community detection using the $\boldsymbol{L}_1 + \boldsymbol{L}_3$ estimate works well, but when $\alpha_{i,j}$ is close to $1-\beta_{i,j}$, the algorithm performs poorly. In fact, simply adding matrices containing community structure information can, in some cases, lead to a rank reduction, thereby masking the structural information in the sum matrix. This approach also makes the final result more susceptible to errors in the $\boldsymbol{L}_1 + \boldsymbol{L}_2$ estimate.

The rank reduction problem not only appears in methods using normalized Laplacian matrices but also in methods using the probability matrices $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$. It should be noted that using $\boldsymbol{P}_1 + \boldsymbol{P}_2$ can also mask structural information in some cases. On the other hand, because the network's underlying community structure affects both the connection

probabilities α and β, algorithms using only one of the probability matrices ($\boldsymbol{P}_1$ or $\boldsymbol{P}_2$) will obviously lose some information, thus affecting the reliability of the estimation results. It is necessary to construct an estimation matrix that can better integrate the information from both probability matrices.

Based on the above discussion, we let

$$\boldsymbol{P}^* = \boldsymbol{P}_1^* \boldsymbol{P}_1^* + \boldsymbol{P}_2^* \boldsymbol{P}_2^*, \tag{2.7}$$

where $\boldsymbol{P}_1^* = \frac{\boldsymbol{P}_1}{\|\boldsymbol{P}_1\|_F}$ and $\boldsymbol{P}_2^* = \frac{\boldsymbol{P}_2}{\|\boldsymbol{P}_2\|_F}$. The matrix $\boldsymbol{P}^*$ contains the necessary information regarding the connection probabilities α and β. Furthermore, its construction avoids masking structural information, since both the $\boldsymbol{P}_1^* \boldsymbol{P}_1^*$ and $\boldsymbol{P}_2^* \boldsymbol{P}_2^*$ matrices are proven to be positive semi-definite.

Remark 1. Let $\lambda_i(A)$ be the ith largest eigenvalue of the square matrix A. Since $\boldsymbol{P}_1^* \boldsymbol{P}_1^*$ and $\boldsymbol{P}_2^* \boldsymbol{P}_2^*$ are both positive semi-definite matrices,

$$\lambda_K(\boldsymbol{P}^*) \geq \max\{\lambda_K(\boldsymbol{P}_1^* \boldsymbol{P}_1^*), \lambda_K(\boldsymbol{P}_2^* \boldsymbol{P}_2^*)\}.$$

Thus, the community structure information contained in the probability matrices $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ is completely preserved. When either $\boldsymbol{P}_1$ or $\boldsymbol{P}_2$ has K distinct communities, $\boldsymbol{P}^*$ also contains K distinct communities.

Moreover, when $\max\{\lambda_K(\boldsymbol{P}_1^* \boldsymbol{P}_1^*), \lambda_K(\boldsymbol{P}_2^* \boldsymbol{P}_2^*)\} = 0$, $\lambda_K(\boldsymbol{P}^*)$ is still possibly larger than 0. In other words, $\boldsymbol{P}^*$ may contain K distinct communities even though neither $\boldsymbol{P}_1$ nor $\boldsymbol{P}_2$ contain K distinct communities. It's the advantage of using the matrix $\boldsymbol{P}^*$.

Remark 2. The reason for using normalized matrices $\boldsymbol{P}_1^*$ and $\boldsymbol{P}_2^*$ instead of the original probability matrices is that the magnitude difference between the values in $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ may be large. Directly using $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ may cause the structural information in the matrices with smaller values to be ignored, that is, the information contained in one of the two matrices may be masked by the estimation error.

Denote $\boldsymbol{B}^* = \frac{\boldsymbol{B}_1\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\boldsymbol{B}_1}{\|\boldsymbol{P}_1\|_F^2} + \frac{\boldsymbol{B}_2\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\boldsymbol{B}_2}{\|\boldsymbol{P}_2\|_F^2}$. Then, by (2.4) and (2.7), we have

$$\boldsymbol{P}^* = \boldsymbol{\Theta}\left(\frac{\boldsymbol{B}_1\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\boldsymbol{B}_1}{\|\boldsymbol{P}_1\|_F^2} + \frac{\boldsymbol{B}_2\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\boldsymbol{B}_2}{\|\boldsymbol{P}_2\|_F^2}\right)\boldsymbol{\Theta}^\top = \boldsymbol{\Theta}\boldsymbol{B}^*\boldsymbol{\Theta}^\top. \tag{2.8}$$

To show that the community structure in the membership matrix $\boldsymbol{\Theta}$ can be recovered by the eigenvectors of $\boldsymbol{P}^*$, we make the following assumption:

Assumption 3. $\boldsymbol{B}^* \in R^{K\times K}$ is full rank. There is at least one pure node for each of the K communities.

We have the following theorem.

Theorem 1. *Assume that Assumption 3 holds. Then rank*$(\boldsymbol{P}^*) = K$. *Let* $\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top$ *be the eigen-decomposition of* $\boldsymbol{P}^*$. *There are at least* K *pure nodes in the network, one for each community. Let* $\mathcal{I}$ *be the row indices corresponding to the* K *pure nodes belonging to* K *different communities. For* $1 \le k \le K$, *the* k*th element in* $\mathcal{I}$ *is the row index of the pure node in community* V_k. *Then* $\boldsymbol{\Gamma}(\mathcal{I},:)$ *is full rank, and* $\boldsymbol{\Gamma} = \boldsymbol{\Theta}\boldsymbol{\Gamma}(\mathcal{I},:)$.

Theorem 1 shows that it is feasible to recover the real membership matrix $\boldsymbol{\Theta}$ using $\boldsymbol{P}^*$, which establishes the theoretical basis for our subsequent proposed algorithm.

2.3 Discussion of identifiability

Theorem 1 shows that the membership matrix $\boldsymbol{\Theta}$ can be recovered if Assumption 3 holds true. However, Assumption 3 is a relatively strict condition, and although we cannot find a way to recover the community structure of $\boldsymbol{\Theta}$ without Assumption 3, a problem worth studying is under what conditions $\boldsymbol{\Theta}$ can be uniquely determined by the real probability matrix $\boldsymbol{P}^*$. For convenience, we define the identifiability as follows. The model is considered identifiable if and only if $\boldsymbol{\Theta}$ can be uniquely determined by $\boldsymbol{P}^*$ up to a permutation of the community labels, or in other words by exchanging the columns of $\boldsymbol{\Theta}$. In stochastic block models, the identifiability is self-evident. However, when it comes to mixed membership models, the identifiability of the model needs to be discussed carefully, because the members corresponding to the nodes are no longer community labels but membership vectors. Some previous discussions of such issues can be found in Mao et al. (2017), Kaufmann et al. (2018), and Zhang et al. (2020). Mao et al. (2021) was the first to propose necessary and sufficient conditions for identifiability in mixed membership models.

We present identifiability results in the following two propositions, which are essentially generalizations of the results in Mao et al. (2021) to our model. Define a "fully mixed" node as a node i that satisfies $\theta_i(k) > 0$ for any community V_k, $1 \leq k \leq K$.

Proposition 1. *Assume that there are K communities, each with at least one pure node. Then, we have the following results.*

(a) If $\operatorname{rank}(\boldsymbol{B}^*) = K$, *then* $\boldsymbol{\Theta}$ *can be uniquely determined by* $\boldsymbol{P}^*$ *up to a permutation of the*

community labels.

(b) If $\text{rank}(\boldsymbol{B}^*) = K - 1$*, and no row in* $\boldsymbol{B}^*$ *is an affine combination of any other rows in* $\boldsymbol{B}^*$*, then* $\boldsymbol{\Theta}$ *can be uniquely determined by* $\boldsymbol{P}^*$ *up to a permutation of the community labels.*

(c) In any other case, if there exists a fully mixed node, then $\boldsymbol{\Theta}$ *cannot be uniquely determined by* $\boldsymbol{P}^*$*.*

The proof of Proposition 1 is similar to the proof of Theorem 2.1 in Mao et al. (2021), and hence omitted.

Proposition 2. *Assume that* $b^*_{k,l} \in (0, 1)$ *for all* $1 \leq k, l \leq K$*.* $\boldsymbol{\Theta}$ *can be uniquely determined by* $\boldsymbol{P}^*$ *up to a permutation of the community labels only if each of the* K *communities has at least one pure node.*

The proof of Proposition 2 follows the result of Theorem 2.2 in Mao et al. (2021) and is therefore omitted.,

It can be seen that the existence of pure nodes is of great significance in the study of mixed membership autoregressive network models.

3. Methodology

3.1 AR-1 mixed spectral clustering algorithm

In order to estimate the membership matrix $\boldsymbol{\Theta}$, we first estimate the node connect probability matrices $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ by

$$\widehat{\boldsymbol{P}}_1 = (\hat{\alpha}_{i,j})_{p\times p}, \quad \widehat{\boldsymbol{P}}_2 = (\hat{\beta}_{i,j})_{p\times p},$$

where $\hat{\alpha}_{i,j}$ and $\hat{\beta}_{i,j}$ are defined in (2.5). Since there are no self-loops in our network, the diagonal elements of $\widehat{\boldsymbol{P}}_1$ and $\widehat{\boldsymbol{P}}_2$ are all zero, while all the diagonal elements in $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are the internal connection probabilities of the communities to which the corresponding nodes belong, which will lead to estimation errors. As shown in Jiang et al. (2023), for large p, the effect of the diagonal elements in both $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ can be neglected.

As mentioned in Subsection 3.1, since $\hat{\alpha}_{i,j}$ and $\hat{\beta}_{i,j}$ are the transition probability in different situations, it is obvious that $\widehat{\boldsymbol{P}}_1$ and $\widehat{\boldsymbol{P}}_2$ both contain information of the membership matrix. Therefore, we use the normalized forms $\widehat{\boldsymbol{P}}_1^*$ and $\widehat{\boldsymbol{P}}_2^*$ to construct

$$\widehat{\boldsymbol{P}}^* = \widehat{\boldsymbol{P}}_1^*\widehat{\boldsymbol{P}}_1^* + \widehat{\boldsymbol{P}}_2^*\widehat{\boldsymbol{P}}_2^* = \frac{\widehat{\boldsymbol{P}}_1\widehat{\boldsymbol{P}}_1}{\|\widehat{\boldsymbol{P}}_1\|_F^2} + \frac{\widehat{\boldsymbol{P}}_2\widehat{\boldsymbol{P}}_2}{\|\widehat{\boldsymbol{P}}_2\|_F^2}.$$

We now introduce our AR-1 mixed spectral clustering (AMSC) algorithm. First, we set up the PCA procedure like the classical spectral method. Let the K eigenvectors $\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2 \ldots \hat{\boldsymbol{p}}_K$ corresponding to the first K largest eigenvalues of $\widehat{\boldsymbol{P}}^*$ be combined into a $p \times K$ matrix

$$\widehat{\boldsymbol{\Gamma}} = (\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2, \ldots, \hat{\boldsymbol{p}}_K),$$

Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_K\}$ be any set of K-dimensional vectors. We can then apply K-medians clustering to obtain K estimated centers of $\widehat{\boldsymbol{\Gamma}}$ by

$$\{\hat{\boldsymbol{v}}_1, \hat{\boldsymbol{v}}_2, \ldots, \hat{\boldsymbol{v}}_K\} = \arg \min_{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_K} \sum_{i=1}^{p} \min_{\boldsymbol{u}\in\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_K\}} \|\widehat{\boldsymbol{\Gamma}}(i,:) - \boldsymbol{u}\|_2. \tag{3.1}$$

Denote $(\hat{\boldsymbol{v}}_1, \hat{\boldsymbol{v}}_2, \ldots, \hat{\boldsymbol{v}}_K)^\top$ as a $K \times K$ matrix $\widehat{\boldsymbol{V}}$. In the classical SDM model, the membership of each node can be directly estimated by finding the row of $\widehat{\boldsymbol{V}}$ that is closest to the row in $\widehat{\boldsymbol{\Gamma}}$. However, in the mixed membership setting, since the actual membership matrix $\boldsymbol{\Theta}$ is not a Boolean matrix, we need to reconstruct the membership to obtain a proper estimate. By projecting $\widehat{\boldsymbol{\Gamma}}$ onto the span of $(\hat{\boldsymbol{v}}_1, \hat{\boldsymbol{v}}_2, \ldots, \hat{\boldsymbol{v}}_K)^\top$, we can compute the estimated membership matrix by

$$\widehat{\boldsymbol{Y}} = \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{V}}^\top(\widehat{\boldsymbol{V}}\widehat{\boldsymbol{V}}^\top)^{-1}. \tag{3.2}$$

Here, we can directly estimate the rows of the real membership matrix $\boldsymbol{\Theta}$ using the normalization step on $\widehat{\boldsymbol{Y}}$. However, since every element in $\boldsymbol{\Theta}$ is non-negative, we find that in practice, adding an extra correction step can significantly reduce the estimation error without affecting the theoretical results, i.e.,

$$\widetilde{y}_{ij} = \max\{\widehat{y}_{ij}, 0\},\ i = 1, \ldots, p,\ j = 1, \ldots, K,$$

where $\widehat{y}_{ij}$ is the (i,j)th element of $\widehat{\boldsymbol{Y}}$. In addition, when all elements in the ith row of $\widehat{\boldsymbol{Y}}$, i.e., $\widehat{\boldsymbol{y}}_i$, are non-positive, we define each element of $\widetilde{\boldsymbol{y}}_i$ to be $\frac{1}{K}$. Let $\widetilde{\boldsymbol{Y}} = (\widetilde{\boldsymbol{y}}_1, \ldots, \widetilde{\boldsymbol{y}}_p)^\top$. Then, the rows of the actual membership matrix $\boldsymbol{\Theta}$ can be estimated by:

$$\hat{\boldsymbol{\theta}}_i = \widetilde{\boldsymbol{y}}_i / \|\widetilde{\boldsymbol{y}}_i\|_1, \quad 1 \le i \le p.$$

The estimated membership matrix $\widehat{\boldsymbol{\Theta}}$ is defined as

$$\widehat{\boldsymbol{\Theta}} = (\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \ldots, \hat{\boldsymbol{\theta}}_p)^\top.$$

The details of the AMSC algorithm can be found in Algorithm 1 below.

Remark 3. It is worth noting that by executing this algorithm, we obtain an estimate of the membership vector for each node in the network. Therefore, when a discrete classification result is required, i.e., an unambiguous community label for each node (whether it is a pure or mixed node), we can classify the node by simply assigning it to the community with the largest membership, which means classifying it to the cluster corresponding to the maximum value in the node's estimated membership vector.

3.2 Estimation of the number of communities K

The AMSC algorithm requires that the number of communities K is known. However, in actual data analysis, this condition is usually not met. In this subsection, we try to discuss the estimation method of the number of communities K when K is unknown. Objectively, finding the number of communities is equivalent to counting the number of non-zero eigenvalues of the matrix containing community structure information. Since the real probability matrix is unobservable, we usually estimate the number of communities by counting the number of eigenvalues far from zero in the estimated probability matrix. We still use the matrix $\widehat{\boldsymbol{P}}^*$ mentioned in Section 3.1. The estimator we use is the empirical eigenvalue threshold

Algorithm 1 AR-1 Mixed Spectral Clustering Algorithm (AMSC) algorithm

Input: The network process $\{\boldsymbol{X}_t,\ t = 0, 1, 2, \ldots, n\}$, the number of nodes p, the number of communities K.

Output: The estimate membership matrix $\widehat{\boldsymbol{\Theta}}$.

1: **for** i in $1 : p$ **do**

2: **for** j in $1 : p$ **do**

3: $\hat{\alpha}_{i,j} = \frac{\sum_{t=1}^{n} X_{i,j}^{t}(1-X_{i,j}^{t-1})}{\sum_{t=1}^{n}(1-X_{i,j}^{t-1})}, \quad \hat{\beta}_{i,j}\frac{\sum_{t=1}^{n}(1-X_{i,j}^{t})X_{i,j}^{t-1}}{\sum_{t=1}^{n} X_{i,j}^{t-1}}$

4: **end for**

5: **end for**

6: $\widehat{\boldsymbol{P}}_1 = (\hat{\alpha}_{i,j})_{p\times p}, \quad \widehat{\boldsymbol{P}}_2 = (\hat{\beta}_{i,j})_{p\times p}$

7: $\widehat{\boldsymbol{P}}^* = \frac{\widehat{\boldsymbol{P}}_1\widehat{\boldsymbol{P}}_1}{\|\widehat{\boldsymbol{P}}_1\|_F^2} + \frac{\widehat{\boldsymbol{P}}_2\widehat{\boldsymbol{P}}_2}{\|\widehat{\boldsymbol{P}}_2\|_F^2}$

8: $\widehat{\boldsymbol{\Gamma}} = [\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2, \ldots, \hat{\boldsymbol{p}}_K]$ ▷ $\hat{\boldsymbol{p}}_i$ is the eigenvector corresponding to the ith largest eigenvalue of $\widehat{\boldsymbol{P}}^*$.

9: $\{\hat{\boldsymbol{v}}_1, \hat{\boldsymbol{v}}_2, \ldots, \hat{\boldsymbol{v}}_K\} = \arg\min_{\hat{\boldsymbol{u}}_1,\ldots,\hat{\boldsymbol{u}}_K} \sum_{i=1}^{p} \min_{\hat{\boldsymbol{u}}\in\{\hat{\boldsymbol{u}}_1,\ldots,\hat{\boldsymbol{u}}_K\}} \|\widehat{\boldsymbol{\Gamma}}(i,:) - \hat{\boldsymbol{u}}\|_2$

10: $\widehat{\boldsymbol{V}} = (\hat{\boldsymbol{v}}_1, \hat{\boldsymbol{v}}_2, \ldots, \hat{\boldsymbol{v}}_K)^\top$

11: $\widehat{\boldsymbol{Y}} = \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{V}}^\top(\widehat{\boldsymbol{V}}\widehat{\boldsymbol{V}}^\top)^{-1}$

12: $\widetilde{y}_{ij} = \max\{\widehat{y}_{ij}, 0\}, i = 1, \ldots, p, j = 1, \ldots, K$. In addition, for each i, if all elements in $\widehat{\boldsymbol{y}}_i$ are non-positive, let each element of $\widetilde{\boldsymbol{y}}_i$ to be $\frac{1}{K}$.

13: **for** i in $1 : p$ **do**

14: $\hat{\boldsymbol{\theta}}_i = \widetilde{\boldsymbol{y}}_i/\|\widetilde{\boldsymbol{y}}_i\|_1, \quad 1 \le i \le p$

15: **end for**

16: $\widehat{\boldsymbol{\Theta}} = (\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \ldots, \hat{\boldsymbol{\theta}}_p)^\top$

estimator (EETE) inspired by Zhang et al. (2021). Let $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p$ be the eigenvalues of $\widehat{\boldsymbol{P}}^*$, and the empirical eigenvalue threshold estimator is defined as

$$\hat{K} = \max_{1 \leqslant k \leqslant p} \left\{ k : \hat{\lambda}_k > \xi \min \left(1, \left(\frac{\sqrt{p \log p} + \sqrt{n}}{\sqrt{np}}\right)^r\right) \right\}, \tag{3.3}$$

where $\xi > 0$ and $0 < r < 1$ are two given constants.

By the definition of $\boldsymbol{P}^*$, it is easy to see that $\boldsymbol{P}^*$ is positive semi-definite. Let $\lambda_1 \geq \cdots \geq \lambda_K > 0$ be the eigenvalues of $\boldsymbol{P}^*$, and let $\hat{K}$ be the estimator of K defined by (3.3). The following proposition gives the consistency of the EETE.

Proposition 3. *Assume that $\lambda_K > c_0$ for a positive constant c_0. Under Assumptions 1-3, with properly chosen constants ξ and r in* (3.3)*, we have $\hat{K} \to K$ almost surely as $n, p \to \infty$.*

In Proposition 3, to ensure that the structure of all communities in the actual connection probability matrix $\boldsymbol{P}^*$ is unique enough, we assume that $\lambda_K > c_0$, where c_0 is a positive constant. This assumption is feasible in practice. In fact, if λ_K is too small, then there is at least one community in the corresponding network that is not obvious, that is, it is very close to some other communities. In this case, even if K is underestimated, the impact of ignoring some less obvious communities due to underestimation of K is acceptable.

Remark 4. Regarding the value of ξ and r in practical applications, we have found that $\xi = 0.02$ and $r = 0.5$ perform well in both simulation and real data applications, so we adopted this value in subsequent analysis. Further discussion on the constant values is

provided in supplementary material.

4. Theoretical results

In this section, we establish the consistency of the estimated membership based on the AMSC algorithm in Subsection 3.1.

First, we show the error bound resulting from using $\widehat{\boldsymbol{P}}^*$ in estimation. Recall that $\lambda_1 \geq \cdots \geq \lambda_K > 0$ are the eigenvalues of $\boldsymbol{P}^*$. We have the following theorem.

Theorem 2. *Assume that when* $n, p \to \infty$, $(\sqrt{p \log p} + \sqrt{n})/(\sqrt{np}\lambda_K) \to 0$. *Under Assumptions 1-3, we have* $\widehat{\boldsymbol{P}}^*$ *is a consistent estimator of* $\boldsymbol{P}^*$ *in the sense that*

$$\|\widehat{\boldsymbol{P}}^* - \boldsymbol{P}^*\|_F = O_p\left(\sqrt{\frac{\log p}{n}} + p^{-1/2}\right). \tag{4.1}$$

In addition, for any constant $c > 0$, *there exists a constant* $C_\gamma > 0$ *and a* $K \times K$ *orthogonal matrix* $\boldsymbol{O}$, *such that the inequality*

$$\|\widehat{\boldsymbol{\Gamma}}\boldsymbol{O} - \boldsymbol{\Gamma}\|_F \leq \left(\frac{C_\gamma(\sqrt{p \log p} + \sqrt{n})}{\sqrt{np}\lambda_K}\right) \tag{4.2}$$

holds true with probability larger than $1 - 4\exp\{-c \log p\}$.

Define $\mathcal{S} = \{\boldsymbol{U} \in R^{K \times K}\}$. Let $\boldsymbol{V} = \boldsymbol{\Gamma}(\mathcal{I}, :)$, where $\mathcal{I}$ is the indices of the rows corresponding to K pure nodes, one for each community. The next goal is to find a bound of $\|\widehat{\boldsymbol{V}}\boldsymbol{O} - \boldsymbol{V}\|_F$, where $\boldsymbol{O}$ is an orthogonal matrix.

We define the loss function $\mathcal{L}(\boldsymbol{Q}, \boldsymbol{U})$ as

$$\mathcal{L}(\boldsymbol{Q}, \boldsymbol{U}) = \frac{1}{p} \sum_{i=1}^{p} \min_{1 \le k \le K} \|\boldsymbol{Q}(i,:) - \boldsymbol{U}(k,:)\|_F.$$

Note that

$$\boldsymbol{\Gamma}(i,:) = \boldsymbol{\theta}_i \boldsymbol{\Gamma}(\mathcal{I},:).$$

It can be found that if the centers of the K-medians clustering fall on the rows of $\boldsymbol{\Gamma}$ corresponding to pure nodes (i.e. $\boldsymbol{\Gamma}(\mathcal{I},:)$). Then, the membership matrix $\boldsymbol{\Theta}$ can be recovered from $\boldsymbol{\Gamma}$ using projection.

Let $\boldsymbol{V}_{\mathcal{L}} = \arg\min_{\boldsymbol{U} \in \mathcal{S}} \mathcal{L}(\boldsymbol{\Gamma}, \boldsymbol{U})$ be the minimizer of the loss function of $\boldsymbol{\Gamma}$. We make the following assumption on $\boldsymbol{\Gamma}$:

Assumption 4. $\boldsymbol{V}_{\mathcal{L}} = \arg\min_{\boldsymbol{U} \in \mathcal{S}} \mathcal{L}(\boldsymbol{\Gamma}, \boldsymbol{U})$ is the unique minimizer, and $\boldsymbol{V} = \boldsymbol{V}_{\mathcal{L}}$ up to a row permutation. Further, there is a global constant κ such that $\kappa K^{-1} \min_{\boldsymbol{\Sigma}} \|\boldsymbol{U} - \boldsymbol{V}\boldsymbol{\Sigma}\|_F \le \mathcal{L}(\boldsymbol{\Gamma}; \boldsymbol{U}) - \mathcal{L}(\boldsymbol{\Gamma}; \boldsymbol{V})$ for any $\boldsymbol{U} \in \mathcal{S}$, where $\boldsymbol{\Sigma}$ is a $K \times K$ permutation matrix.

Assumption 4 states that the unique minimum of the K-median loss function on $\boldsymbol{\Gamma}$ falls exactly at the matrix $\boldsymbol{V}$, and there is curvature around the minimum. A similar assumption and some analyses can be found in Section 4.2 of Zhang et al. (2020).

Then, We have the following theorem.

Theorem 3. *If Assumptions 1-4 hold and* $\left(\sqrt{p \log p} + \sqrt{n}\right) / (\sqrt{np}\lambda_K) \to 0$ *as* $n, p \to \infty$, *then for any constant* $c > 0$, *there exists a constant* C_v *such that*

$$\|\widehat{\boldsymbol{V}}\boldsymbol{O} - \boldsymbol{V}\|_F \le \frac{C_v K(\sqrt{p \log p} + \sqrt{n})}{p\sqrt{n}\lambda_K} \tag{4.3}$$

with probability at least $1 - 4\exp\{-c\log p\}$.

To investigate the bound of $\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_F$. First, let $m = \frac{min_i\{\|\boldsymbol{\Gamma}(i,:)\|_2^2\}}{max_i\{\|\boldsymbol{\Gamma}(i,:)\|_2^2\}}$. Since for each i, $\boldsymbol{\Gamma}(i,:) = \boldsymbol{\theta}_i\boldsymbol{\Gamma}(\mathcal{I},:)$, together with the truth that $\|\boldsymbol{\theta}_i\| = 1$, we consider m to be controllable.

Next, we make the following assumption:

Assumption 5. There exists a global constant $m_V > 0$ such that $\lambda_{\min}(\boldsymbol{V}\boldsymbol{V}^\top) \geq \frac{m_V}{p}$.

Remark 5. Note that $\|\boldsymbol{\Gamma}\|_F = \sqrt{K}$, since $\boldsymbol{V} = \boldsymbol{\Gamma}(\mathcal{I},:)$ is a submatrix of $\boldsymbol{\Gamma}$, we have

$$\|\boldsymbol{V}\|_F = \sqrt{\frac{\|\boldsymbol{V}\|_F^2}{\|\boldsymbol{\Gamma}\|_F^2}\|\boldsymbol{\Gamma}\|_F^2} = \sqrt{\frac{\|\boldsymbol{\Gamma}(\mathcal{I},:)\|_F^2}{(\sum_{i=1}^p \|\boldsymbol{\Gamma}(i,:)\|_2^2)}K}.$$

Therefore, we have

$$\frac{\sqrt{m}K}{\sqrt{p}} = \sqrt{\frac{mK}{p}K} \leq \|\boldsymbol{V}\|_F \leq \sqrt{\frac{K}{mp}K} = \frac{K}{\sqrt{mp}}.$$

Since $\boldsymbol{V}$ is a $K \times K$ matrix, based on the above facts, we argue that Assumption 5 is not unattainable.

Then, we have

Theorem 4. *If Assumptions 1-5 hold and* $\left(\sqrt{p\log p} + \sqrt{n}\right)/(\sqrt{np}\lambda_K) \to 0$ *as* $n, p \to \infty$, *then for any constant* $c > 0$, *there exists a constant* C_z *such that*

$$\frac{\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_F}{\sqrt{p}} \leq \frac{C_z K^4(\sqrt{p\log p} + \sqrt{n})}{m\sqrt{np}\lambda_K {m_V}^2}, \tag{4.4}$$

with probability at least $1 - 4\exp\{-c\log p\}$, *where* $m = \frac{min_i\{\|\boldsymbol{\Gamma}(i,:)\|_2^2\}}{max_i\{\|\boldsymbol{\Gamma}(i,:)\|_2^2\}}$.

5. Numerical results

In this section, we present the simulation results of the proposed empirical eigenvalue-threshold estimator (EETE) for the community number estimation problem. In addition, for the community detection problem, we show the performance of the AMSC algorithm on stochastic block AR-1 networks and mixed-membership AR-1 networks.

5.1 Simulation on the estimation of the community number K.

First, we conduct simulations aimed at estimating the number of communities K. To examine the performance of EETE, we compared it to the Bayesian information criterion (BIC). The Bayesian information criterion was originally defined in Schwarz (1978) and has been improved in the subsequent work of many researchers. In our simulations, we use the definition in Bai et al. (2018), where the criterion BIC is defined as

$$\mathrm{BIC}_j = (p-1-j)\log\bar{\lambda}_j - \sum_{i=j+1}^{p-1}\log\hat{\lambda}_i - \frac{(p-j-2)(p-j+1)}{2p}\log p,$$

where

$$\bar{\lambda}_j := \frac{1}{p-1-j}\sum_{t=j+1}^{p-1}\hat{\lambda}_t, \quad j = 1, 2, \ldots, p-1,$$

and the community number is estimated by $\hat{K} = \arg\min_{j\le p-2}\mathrm{BIC}_j$.

Next, we examine the performance of these two methods in simulated networks. We set the number of real communities K to 2, 3, or 4, and the number of nodes p to 200. For $i = 1, \ldots, K$, we set $\zeta_{i,i} = \eta_{i,i} = 0.4$, and for $1 \le i < j \le K$, we let $\zeta_{i,j}$ and $\eta_{i,j}$ be drawn

independently from $U[0.05, 0.25]$.

For networks with two communities, we set 66 pure nodes for each community, and the remaining 68 nodes as mixed nodes. For each mixed node i, we first uniformly generate a random value x_i from $[0, 0.5]$ and then randomly assign the membership vector to $(x_i, 1-x_i)$ or $(1-x_i, x_i)$ with probability 1/2.

For networks with three communities, we set 50 pure nodes for each community and the remaining 50 nodes as mixed nodes. For each mixed node i, we first uniformly generate a random value x_i from $[0, 0.5]$ and then randomly assign the membership vector to $(x_i, x_i, 1-2x_i)$, $(x_i, 1-2x_i, x_i)$, or $(1-2x_i, x_i, x_i)$ with probability 1/3.

For networks with four communities, we set 40 pure nodes for each community and the remaining 40 nodes as mixed nodes. For each mixed node i, we first uniformly generate x_i from $[0, 0.25]$ and then randomly assign the membership vector to

$$\{(x_i, x_i, x_i, 1-3x_i),\ (x_i, x_i, 1-3x_i, x_i),\ (x_i, 1-3x_i, x_i, x_i),\ (1-3x_i, x_i, x_i, x_i)\},$$

with probability 1/4.

As described in Section 3.2, we set $\xi = 0.02$ and $r = 0.5$ for EETE. For each combination of K and sample size n, we perform 500 simulations and count the number of correct estimates. The simulation results are reported in Table 1, from which it can be seen that EETE outperforms BIC for various values of K and n. A possible reason for the comparatively unstable performance of the BIC is its dependence on the ratio $\hat{\lambda}_{K+1}/\bar{\lambda}_{K+1}$, which is heavily influenced by all the small, noisy eigenvalues. This may lead to less reliable estimates for the

Table 1: The simulation results of the BIC and EETE for different sample size n and the number of communities K.

		EETE		BIC	
K	n	No. of Correct	No. of Wrong	No. of Correct	No. or Wrong
2	10	500	0	500	0
3	10	493	7	334	166
	15	495	5	455	45
	20	500	0	497	3
4	10	402	98	47	453
	15	428	72	131	369
	20	484	16	339	161
	25	490	10	446	54

true community number K. We provide a detailed analysis of this dependence in Section A.1.5.

5.2 Simulation for AR-1 SBM

For the simulation on the AR-1 mixed spectral clustering algorithm, we first show its performance on AR-1 stochastic block models without mixed membership. We generate data according to an AR-1 stochastic block model with 3 clusters and 150 nodes. The three communities are set to contain $p_1, p_2 - p_1$, and $150 - p_2$ nodes respectively, where p_1 and p_2 are random integers uniformly drawn from $\{1, \ldots, 149\}$ and satisfy $p_1 < p_2$. For the community transition probabilities ζ and η, we consider two settings. In Setting 1, we set $\zeta_{i,i} = \eta_{i,i} = 0.4$ for $i = 1, \ldots, K$, and we let $\zeta_{i,j}$ and $\eta_{i,j}$ be independently drawn from $U[0.05, 0.25]$ when

$1 \leq i < j \leq K$. Under Setting 2, for $i = 1, \cdots, K$, $\zeta_{i,i} = \eta_{i,i} = 0.4$, for $1 \leq i < j \leq K$, $\zeta_{i,j}$ is drawn from $U[0.05, 0.25]$ and $\eta_{i,j}$ is drawn from $U[0.55, 0.75]$. We repeat the simulation 500 times for each setting and different n. Since the actual membership matrix $\mathbf{\Theta}$ is a Boolean matrix in this case, we convert the estimate $\widehat{\mathbf{\Theta}}$ (obtained by performing the AR-1 mixture spectral clustering algorithm) into a Boolean matrix by setting the largest value in each row to 1 and the others to 0. The standard adjacency matrix-based spectral clustering method for static networks in Rohe et al. (2011) (using an average adjacency matrix instead of a single adjacency matrix in our simulations) and the AR-1 spectral clustering algorithm in Jiang et al. (2023) are also used for comparison. We report the normalized mutual information (NMI) in [0,1] as a measure of the performance of each algorithm. A large NMI value indicates that the estimated membership is close to the true membership and vice versa. The results of Setting 1 are summarized in Table 2 and the results of Setting 2 are shown in Table 3.

It can be seen that under Setting 1, due to the small difference between the connection probability $\pi_{i,j} = \frac{\alpha_{i,j}}{\alpha_{i,j}+\beta_{i,j}}$ within the community and between the communities, it is difficult for the static network algorithm using information of $\pi_{i,j}$ to estimate the community membership. In addition, as we mentioned in Subsection 2.2, although the AR-1 algorithm in Jiang et al. (2023) outperforms the static network algorithm, it still performs poorly when $\alpha_{i,j}$ and $\beta_{i,j}$ are close for any i and j, because part of the structural information is masked in the sum of the Laplacian matrices. At the same time, our algorithm consistently outperforms

Table 2: Normalized mutual information of the actual and estimated membership matrices for 500 replicates of different sample sizes n for Setting 1 on the AR-1 stochastic block model without mixed membership.

			Normalized mutual information		
K	p	n	The AMSC algorithm	AR-1 algorithm	Static network algorithm
3	150	5	0.7206	0.5101	0.3999
		20	0.8127	0.4849	0.5302
		50	0.8703	0.6329	0.6420
		100	0.9078	0.7266	0.6922

other algorithms.

In Setting 2, due to the significant difference in the connection probabilities within and between communities, the static network method performs better than Setting 1, and even shows the best performance among all methods when $n = 5$. This is because other algorithms need to partition the samples to calculate the transition probabilities $\alpha_{i,j}$ and $\beta_{i,j}$, respectively, so they require more data than static methods to obtain satisfactory estimates. The AR-1 algorithm in Jiang et al. (2023) outperforms the static network method when $n \geq 20$. However, its performance is still slightly worse than the AMSC algorithm we propose. Finally, except for $n = 5$, the AMSC algorithm has the best performance in all other cases.

In summary, for autoregressive networks without mixed membership, the AMSC algorithm exhibits acceptable performance under different transition probability settings.

Table 3: Normalized mutual information of the actual and estimated membership matrices for 500 replicates of different sample sizes n for Setting 2 on the AR-1 stochastic block model without mixed membership.

			Normalized mutual information		
K	p	n	The AMSC algorithm	AR-1 algorithm	Static network algorithm
3	150	5	0.6826	0.7217	0.7777
		20	0.8685	0.8473	0.8342
		50	0.9072	0.8694	0.8521
		100	0.9286	0.8818	0.8597

5.3 Simulation for AR-1 MMSB

We now consider the performance of our algorithm on the AR-1 mixed-membership stochastic block model. We generate data with 3 clusters and 200 nodes. We let each cluster contain 50 pure nodes, and for the remaining 50 nodes, we simulate two cases: let them have a "uniform membership" of 1/3,1/3,1/3 for each mixed node, or let their membership be random, which means for each mixed node i, the membership is randomly set to $(x_i, x_i, 1-2x_i)$, $(x_i, 1-2x_i, x_i)$ or $(1-2x_i, x_i, x_i)$, each with probability 1/3. As in the previous simulations, for node i, x_i is a random value uniformly drawn from the interval [0,0.5]. In addition to the AMSC algorithm, the AR-1 spectral clustering algorithm in Jiang et al. (2023) and the Overlapping Continuous Community Assignment Model (OCCAM) algorithm using the mean adjacency matrix $\boldsymbol{X}$ in Zhang et al. (2020) are also used for comparison. Since the membership estimated by OCCAM algorithm for each node i satisfies $\|\boldsymbol{\theta}_i\|_2 = 1$, we apply

Table 4: The Mixed-Hamming error rate of the estimated and real membership matrix with 500 replications for Setting 1 with different sample sizes and different membership on the AR-1 mixed membership model of 3 clusters and 200 nodes.

Uniform membership			Mixed-Hamming error rate		
K	p	n	The AMSC algorithm	AR-1 algorithm	OCCAM algorithm
3	200	5	0.1497	0.6458	0.6970
		20	0.0906	0.7087	0.5231
		50	0.0482	0.5544	0.4191
		100	0.0301	0.4594	0.3537
Random membership			Mixed-Hamming error rate		
K	p	n	The AMSC algorithm	AR-1 algorithm	OCCAM algorithm
3	200	5	0.1443	0.5277	0.6593
		20	0.0891	0.6367	0.5095
		50	0.0472	0.4745	0.4053
		100	0.0299	0.3622	0.3442

additional normalization to satisfy $\|\boldsymbol{\theta}_i\|_1 = 1$. We use the Mixed-Hamming error rate to measure the effectiveness of each algorithm, that is,

$$\min_{\boldsymbol{O}\in\{K\times K\text{permutation matrix}\}} \frac{1}{p}\|\widehat{\boldsymbol{\Theta}}\boldsymbol{O} - \boldsymbol{\Theta}\|_1.$$

The community transition probabilities $\zeta_{i,j}$ and $\eta_{i,j}$ are set in two settings, the same as in Experiment 1. In both settings, $\zeta_{i,i} = \eta_{i,i} = 0.4$, and for Setting 1, when $1 \le i < j \le K$, we let $\zeta_{i,j}$ and $\eta_{i,j}$ be drawn independently from $U[0.05, 0.25]$, while for Setting 2, when $1 \le i < j \le K$, $\zeta_{i,j}$ is drawn from $U[0.05, 0.25]$, and $\eta_{i,j}$ is drawn from $U[0.55, 0.75]$. The results are shown in Tables 4-5.

Table 5: The Mixed-Hamming error rate of the estimated and real membership matrix with 500 replications for Setting 2 with different sample sizes and different membership on the AR-1 mixed membership model of 3 clusters and 200 nodes.

Uniform membership			Mixed-Hamming error rate		
K	p	n	The AMSC algorithm	AR-1 algorithm	OCCAM algorithm
3	200	5	0.2094	0.3337	0.1244
		20	0.0685	0.3334	0.0661
		50	0.0425	0.3333	0.0430
		100	0.0295	0.3333	0.0314
Random membership			Mixed-Hamming error rate		
K	p	n	The AMSC algorithm	AR-1 algorithm	OCCAM algorithm
3	200	5	0.1884	0.2242	0.1177
		20	0.0659	0.2110	0.0632
		50	0.0415	0.2097	0.0409
		100	0.0292	0.2095	0.0299

From Table 4, we can see that under Setting 1, the performance of the AMSC algorithm is significantly better than that of other algorithms. From Table 5 centered on Setting 2, we can see that although the OCCAM algorithm performs best when the sample size is small (n=5), as the sample size increases, our algorithm gradually shows comparable performance to the OCCAM algorithm. In contrast, the estimation error of the original AR-1 algorithm in Jiang et al. (2023) only decreases slightly with the increase of sample size, which may be due to the fact that the estimated membership matrix by the original AR-1 algorithm is a Boolean matrix, resulting in a lower bound on the estimation error.

6. Real data analysis

In this section, we apply the proposed method to real-world data.

6.1 Global trade data

First, we apply our method to global trade data for 195 countries from 1991 to 2014. For simplicity, if there is trade between two countries in a year, we regard trade as an edge between the two countries, regardless of the direction, so we transform the data into a network process with $n = 24$ and $p = 195$. The data are taken from public trade data for 205 countries from 1870 to 2014 (Barbieri et al., 2009; Keshk, 2017)), and the reason for choosing data from this period is that with the collapse of the Soviet Union in 1991, the international situation has changed dramatically, and using data before this period may affect the stability of the estimation results. We only consider countries that have existed from 1991 to 2014.

Regarding the number of communities K, we first assume $K = 2$ in our analysis. This assumption is the same as Jiang et al. (2023), that is, the membership vector corresponding to each node is two-dimensional. The estimated membership results are shown in Figure 1. There are 53 pure nodes in the estimated Community 1, including Monaco, Kosovo, Somalia, Uzbekistan, Nepal, Nauru, and some developing countries in Asia and Africa. The estimated Community 2 contains 45 pure nodes, including the United States, the United Kingdom, France, China, Japan, India and other countries, most of which are developed

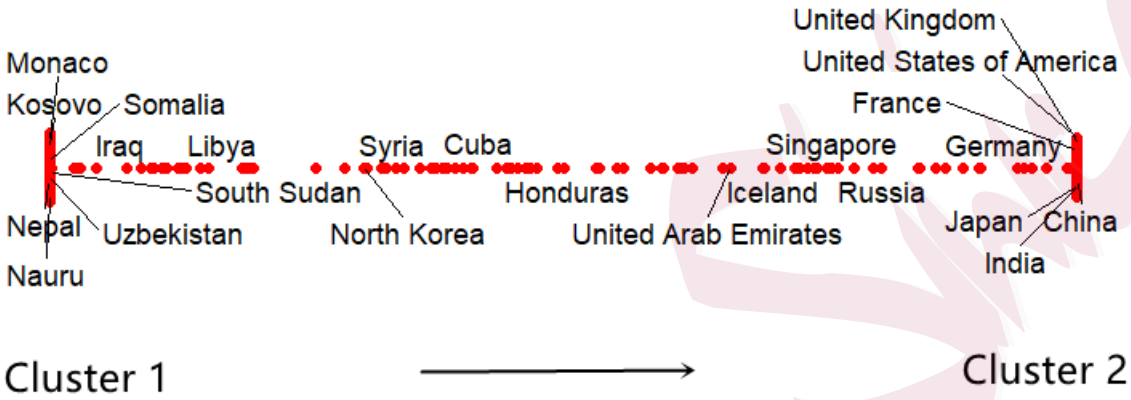

Figure 1: The estimated membership from the Global trade data for both communities.

countries.

Next, we estimate K using our proposed empirical eigenvalue threshold estimator, and the results show that for this dataset, the number of communities should be set to $K = 6$. We then apply the AMSC algorithm to the trade data with $K = 6$ and obtain $\widehat{\mathbf{\Theta}}$. To facilitate the presentation of our results, we group countries into several clusters as described in Remark 3. The classification results can be found in the supplementary materials.

In the classification results of countries and regions, Community 1 includes economically developed and trade-intensive countries and regions such as the United States, the United

Kingdom, and China, while Community 2 is mainly composed of less developed countries and regions in Latin America. Community 3 contains micro-states distributed throughout Europe and the major oceans, Community 4 contains most of the remaining countries and regions in Eastern Europe and Asia, Community 5 includes parts of Oceania, Asia and Africa surrounding the Indian Ocean, while most of the countries and regions in Africa and the Middle East make up Community 6. It is worth noting that the probability of some countries joining any community is less than 0.4, such as Iceland, North Korea, Hungary, and Sri Lanka. This means that their trade propensity is between different communities, but not significantly close to any communities.

We put some figures showing further details in the supplementary material. Meanwhile, to test the robustness of the results, We present the results of estimating data over different time spans in the supplementary material.

6.2 French high school contact data

In addition to the global trade data, we focus on a dataset corresponding to the contacts between 327 students in 9 classes of a high school in Marseille, France, in December 2013, taken from Mastrandrea et al. (2015). This dataset is also analyzed in Jiang et al. (2023). We applied the empirical eigenvalue-threshold estimator to the dataset and found that $K = 9$ is the most appropriate setting. Then, the AMSC algorithm identified students of the 9 original classes perfectly into 9 communities, which is better than the result given by Jiang

et al. (2023). To demonstrate the robustness of our estimation, we also try to randomly delete the contact data of some students and then use the AMSC algorithm to estimate the communities of the remaining students. The results show that our estimation has good robustness, and the estimation effect is better than that of the original AR-1 algorithm in Jiang et al. (2023) on this dataset. We present the results and details of our analysis in the supplementary materials.

7. Conclusion

In this paper, we present a mixed membership stochastic block model incorporating an autoregressive structure, which overcomes the limitations of previous approaches by integrating temporal information with a hybrid community structure. Subsequently, we propose a novel algorithms to estimate the number of communities and address the community detection problem for this model. We also give the corresponding theoretical findings. Finally, through comprehensive simulation experiments and real data analysis, we demonstrate the superior performance of our proposed methods compared to existing approaches in terms of their efficacy and accuracy.

Supplementary Materials

The supplementary material offers some additional remarks to the proposed model, the supplement to the real data analysis, and details of the proofs of all propositions and theorems.

References

Abbe, E. (2017). Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research 18*(1), 6446–6531.

Airoldi, E. M., D. Blei, S. Fienberg, and E. Xing (2008). Mixed membership stochastic blockmodels. *Advances in neural information processing systems 21*.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, pp. 267–281. Akadémiai Kiadó Location Budapest, Hungary.

Amini, A. A., A. Chen, P. J. Bickel, and E. Levina (2013). Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics*, 2097–2122.

Anandkumar, A., R. Ge, D. Hsu, and S. M. Kakade (2014). A tensor approach to learning mixed membership community models. *The Journal of Machine Learning Research 15*(1), 2239–2312.

Bai, Z., K. P. Choi, and Y. Fujikoshi (2018). Consistency of aic and bic in estimating the number of significant components in high-dimensional principal component analysis. *The Annals of Statistics 46*(3), 1050–1076.

Barabási, A.-L. and Z. N. Oltvai (2004). Network biology: understanding the cell's functional organization. *Nature reviews genetics 5*(2), 101–113.

Barbieri, K., O. M. Keshk, and B. M. Pollins (2009). Trading data: Evaluating our assumptions and coding rules. *Conflict Management and Peace Science 26*(5), 471–491.

Ben-Daya, M., E. Hassini, and Z. Bahroun (2019). Internet of things and supply chain management: a literature review. *International journal of production research 57*(15-16), 4719–4742.

REFERENCES

Bhattacharjee, M., M. Banerjee, and G. Michailidis (2020). Change point estimation in a dynamic stochastic block model. *The Journal of Machine Learning Research 21*(1), 4330–4388.

Bickel, P., D. Choi, X. Chang, and H. Zhang (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics 41*(4), 1922–1943.

Cai, T. T. and X. Li (2015). Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. *The Annals of Statistics*, 1027–1059.

Chatterjee, S. (2015). Matrix estimation by universal singular value thresholding. *The Annals of Statistics 43*(1), 177–214.

Chen, Y., X. Li, and J. Xu (2018). Convexified modularity maximization for degree-corrected stochastic block models. *The Annals of Statistics 46*(4), 1573.

Donnat, C. and S. Holmes (2018). Tracking network dynamics: A survey using graph distances. *The Annals of Applied Statistics 12*(2), 971–1012.

Durante, D. and D. B. Dunson (2016). Locally adaptive dynamic networks. *The Annals of Applied Statistics*, 2203–2232.

Easley, D., J. Kleinberg, et al. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*, Volume 1. Cambridge university press Cambridge.

Girvan, M. and M. E. Newman (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences 99*(12), 7821–7826.

Goldenberg, A., A. X. Zheng, S. E. Fienberg, E. M. Airoldi, et al. (2010). A survey of statistical network models.

Foundations and Trends® in Machine Learning 2(2), 129–233.

Hagen, L. and A. B. Kahng (1992). New spectral methods for ratio cut partitioning and clustering. *IEEE transactions on computer-aided design of integrated circuits and systems 11*(9), 1074–1085.

Handcock, M. S., A. E. Raftery, and J. M. Tantrum (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 170*(2), 301–354.

Hoff, P. (2007). Modeling homophily and stochastic equivalence in symmetric relational data. *Advances in neural information processing systems 20*.

Holland, P. W., K. B. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps. *Social networks 5*(2), 109–137.

Holland, P. W. and S. Leinhardt (1981). An exponential family of probability distributions for directed graphs. *Journal of the american Statistical association 76*(373), 33–50.

Jiang, B., J. Li, and Q. Yao (2023). Autoregressive networks. *Journal of Machine Learning Research 24*(227), 1–69.

Jin, J. (2015). Fast community detection by score. *The Annals of Statistics 43*(1), 57–89.

Karrer, B. and M. E. Newman (2011). Stochastic blockmodels and community structure in networks. *Physical review E 83*(1), 016107.

Kaufmann, E., T. Bonald, and M. Lelarge (2018). A spectral algorithm with additive clustering for the recovery of overlapping communities in networks. *Theoretical Computer Science 742*, 3–26.

Keshk, O. M. (2017). Correlates of war project trade data set codebook, version 4.0 1870-2014 katherine barbieri, university of south carolina. *Science 26*(5), 471–491.

Kolaczyk, E. D. and G. Csárdi (2014). *Statistical analysis of network data with R*, Volume 65. Springer.

Krivitsky, P. N. and M. S. Handcock (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society Series B: Statistical Methodology 76*(1), 29–46.

Lee, E., F. Karimi, C. Wagner, H.-H. Jo, M. Strohmaier, and M. Galesic (2019). Homophily and minority-group size explain perception biases in social networks. *Nature human behaviour 3*(10), 1078–1087.

Lei, J. (2021). Network representation using graph root distributions. *The Annals of Statistics 49*(2).

Lei, J. and A. Rinaldo (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 215–237.

Ludkin, M., I. Eckley, and P. Neal (2018). Dynamic stochastic block models: parameter estimation and detection of changes in community structure. *Statistics and Computing 28*, 1201–1213.

Ma, S., L. Su, and Y. Zhang (2021). Determining the number of communities in degree-corrected stochastic block models. *The Journal of Machine Learning Research 22*(1), 3217–3279.

Majid, M., S. Habib, A. R. Javed, M. Rizwan, G. Srivastava, T. R. Gadekallu, and J. C.-W. Lin (2022). Applications of wireless sensor networks and internet of things frameworks in the industry revolution 4.0: A systematic literature review. *Sensors 22*(6), 2087.

Mao, X., P. Sarkar, and D. Chakrabarti (2017). On mixed memberships and symmetric nonnegative matrix factorizations. In *International Conference on Machine Learning*, pp. 2324–2333. PMLR.

Mao, X., P. Sarkar, and D. Chakrabarti (2021). Estimating mixed memberships with sharp eigenvector deviations. *Journal of the American Statistical Association 116*(536), 1928–1940.

Mastrandrea, R., J. Fournet, and A. Barrat (2015). Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PloS one 10*(9), e0136497.

Matias, C. and V. Miele (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society Series B: Statistical Methodology 79*(4), 1119–1141.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review 45*(2), 167–256.

Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the national academy of sciences 101*(suppl_1), 5200–5205.

Pensky, M. (2019). Dynamic network models and graphon estimation. *The Annals of Statistics 47*(4).

Rohe, K., S. Chatterjee, and B. Yu (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 1878–1915.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461–464.

Shi, J. and J. Malik (2000). Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence 22*(8), 888–905.

Tian, H., B. Zhang, R. Jiang, and X. Han (2024+). A new preferential model with homophily for recommender systems. *Statistica Sinica. DOI:10.5705/ss.202022.0136*.

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing 17*, 395–416.

Xu, J. (2018). Rates of convergence of spectral methods for graphon estimation. In *International Conference on Machine Learning*, pp. 5433–5442. PMLR.

Zhang, W., B. Jin, and Z. Bai (2021). Learning block structures in u-statistic-based matrices. *Biometrika 108*(4),

933–946.

Zhang, Y., E. Levina, and J. Zhu (2020). Detecting overlapping communities in networks using spectral methods. *SIAM Journal on Mathematics of Data Science 2*(2), 265–283.

Zhao, Y., E. Levina, and J. Zhu (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics 40*(4), 2266–2292.

Acknowledgements

Bo Zhang (Corresponding author) is partially supported by National Key R&D Program of China (2022YFA1008000) and National Natural Science Foundation of China (12471268 & 12001517 & 72091212).

Baisuo Jin is partially supported by the National Natural Science Foundation (Grants 12231017,72293573)

Yuehua Wu is supported by the Natural Science and Engineering Research Council of Canada (No. RGPIN-2023-05655).

Tianyi Sun, Department of Statistics and Finance, University of Science and Technology of China, Hefei, China.

E-mail: (sty76@mail.ustc.edu.cn)

Bo Zhang, Department of Statistics and Finance, University of Science and Technology of China, Hefei, China.

E-mail: (wbchpmp@ustc.edu.cn)

Baisuo Jin, School of Mathemtical Sciences, Xinjiang Normal University; Department of Statistics and Finance, University of Science and Technology of China, Hefei, China. E-mail: (jbs@ustc.edu.cn)

Yuehua Wu, Department of Mathematics and Statistics, York University, Toronto, Canada

E-mail: (wuyh@yorku.ca)