Overlapping Community Detection in Mixed Membership Vector Autoregression

Siao Xu*†

September 8, 2026

Abstract

In this work, we introduce a novel cluster affiliation vector autoregressive model, termed the mixed membership stochastic block vector autoregression (MMSB-VAR(p)). This model assumes that the VAR coefficients at all lags are random matrices following a joint distribution defined by a probabilistic generative framework. The panel data generated by this model can exhibit either single or multiple membership structures with the estimated VAR coefficients recovering the corresponding non-overlapping or overlapping block structures. To uncover the latent group structure, we propose a two-step algorithm leveraging the key insight that spillover effects within groups are significantly stronger than those between groups. We establish the consistency of the proposed algorithm and the simulation study demonstrates its effectiveness. Finally, we apply our methodology to analyze public opinions about the Ukraine-Russia war on Reddit.

Keywords: Vector Autoregression, Community detection, Random graph, Spectral clustering, Probabilistic graph model, Mixed membership model

JEL Codes: key1, key2, key3

*I am deeply indebted to my advisor, Carsten Trenkler. I am very grateful to Mengshan Xu, So Jin Lee, Xia Wang for their very insightful suggestions.

†Graduate School of Economic and Social Sciences, University of Mannheim. E-mail: sixu@mail.uni-mannheim.de

1 Introduction

Over the past decade, network analysis has become a widely used tool in econometrics and statistics. Network-based approaches analyzing multiple time series have gained significant attention recently, e.g., financial system connectedness network using forecast error variance decomposition (FEVDs) as network weights have been extensively studied (Diebold and Yılmaz, 2014; Demirer et al., 2018), other notable contributions include the development of a LASSO-type algorithm for sparse VAR coefficients and its partial correlation matrices which are both interpreted as networks (Barigozzi and Brownlees, 2019), as well as a connectedness measure for financial entities based on Granger causality networks (Billio et al., 2012). Additionally, risk networks driven by tail events have been explored in the context of systemic risk (Härdle, Wang and Yu, 2016). On the other hand, detecting group structures within panel data has become a prominent research area in econometrics, see e.g., Bonhomme and Manresa (2015); Ando and Bai (2016); Chen, Härdle and Klochkov (2022); Su, Shi and Phillips (2016) and Zhang, Wang and Zhu (2019). In line with this literature, our paper contributes to the growing body of research on group structure detection in panel data through network analysis. Specifically, we introduce a novel vector autoregressive (VAR) model that generates panels with multiple-membership group structures, along with a set of algorithms designed to detect these structures.

In general, a large panel can be represented as a graph, where the vertices represent the individual time series and the edges capture the measures of dependence between them. One specific way to represent such a panel is through a Granger causality network based on a vector autoregression (VAR) (Barigozzi and Brownlees, 2019). The dependence structure within the panel is foundational for identifying latent group structures. Specifically, time series in the panel can be classified into different groups, where intra-group dependence is strong, while inter-group dependence is relatively weak. The variation in spillover effects between series serves as the basis for learning these group structures. In Guðmundsson and Brownlees (2021), the concept of a stochastic block vector autoregression (SB-VAR(p))

is introduced, which generates the panels with a single-membership group structure. This means that each variable is assigned to only one group. However, in many real-world applications, multiple memberships are common. For instance, econometricians may belong to both economics and statistics communities, proteins often perform multiple functions, fin-tech companies operate in both financial and high-tech sectors, children participate in multiple hobby groups at school, and individuals can hold multiple nationalities. These observations motivate our work. We extend this framework by introducing a mixed membership stochastic block vector autoregression (MMSB-VAR(p)), which allows for both single- and multiple-membership group structures. It should be noted that the stochastic block vector autoregression introduced by Guðmundsson and Brownlees (2021) can be viewed as a special case of the mixed membership stochastic block autoregression proposed in this paper. When all series in the panel belong exclusively to one group, the MMSB-VAR(p) reduces to the SB-VAR(p).

In our model, we assume that the VAR coefficients are random matrices, following a joint distribution. More specifically, we model the vector autoregression as a probabilistic generative process, which can be analyzed using probabilistic graphs. To the best of our knowledge, this is the first study to analyze VAR models within a probabilistic graphical framework.

We begin by introducing the mixed membership model and its application in network analysis, which leads to the development of the mixed membership stochastic block model (MMSBM) (Airoldi et al., 2008). We then extend this framework by introducing the mixed membership stochastic block vector autoregression. The MMSB-VAR(p) builds on the original MMSBM by incorporating temporal dynamics through all lags. In this model, a spillover effect between time series i and j exists if and only if an edge is present between nodes i and j in the latent network. A key feature of the model is that the autoregressive coefficients inherit the block structure defined by the MMSBM.

We propose a two-step algorithm to detect panel data with either single- or multiple-

membership group structures. The first step involves spectral clustering (Ng, Jordan and Weiss, 2001; Von Luxburg, 2007), which classifies each time series to one of several groups. In the second step, we refine the partitions in the first step by identifying series that belong to multiple groups. Our simulation study demonstrates the effectiveness of the algorithm in detecting both single- and multiple-membership structures.

We analyze the consistency of our algorithm and demonstrate that it reliably detects group structures in panel data when both the number of time series N and the number of observations T are sufficiently large. However, the ratio N/T must remain bounded. Additionally, N_O/N, where N_O is the number of time series having multiple memberships, must also be limited. Our approach is inspired by the method of bounding the misclassification rate from Rohe, Chatterjee and Yu (2011). We establish consistency by bounding the misclassification rates in both the first and second step of the algorithm.

We conduct an empirical application to illustrate the effectiveness of our method. Specifically, we apply the proposed algorithm to a word count panel of public comments about recent Ukraine-Russia war on Reddit.

Our paper is related to several strands of the literature. Firstly, it contributes to the body of work on group detection in panel data, see e.g., Brownlees, Guðmundsson and Lugosi (2022), Su, Shi and Phillips (2016), Zhang, Wang and Zhu (2019), and Chen, Härdle and Klochkov (2022). In particular, this work is closely related to Guðmundsson and Brownlees (2021), where the network representation for panel data is the Granger causality network based on vector autoregression. Secondly, it connects to the literature on network autoregressive models including, among others, Zhu et al. (2017), Chen, Fan and Zhu (2023), and Zhu et al. (2019). Thirdly, we build upon the mixed membership models, see e.g., Blei, Ng and Jordan (2003), Airoldi et al. (2008), Erosheva (2003), and Jin, Ke and Luo (2024). Finally, the paper is also related to the methodology for detecting overlapping structure in graphs, see e.g., Ahn, Bagrow and Lehmann (2010), Latouche, Birmelé and Ambroise (2009), Yang and Leskovec (2013), and Palla et al. (2005).

The rest of the paper is organized as follows: Section 2 introduces the mixed membership model. In Section 3, we present the main model and methodology. Section 4 describes the proposed algorithm. Section 5 provides the theoretical analysis. Sections 6 and 7 are devoted to simulation and empirical studies, respectively. Finally, Section 8 concludes the paper. Proofs and additional material are provided in the Appendix.

1.1 Notation and the definition for Graph

Let N denote the number of time series and T represent the number of observations. We define N_O as the number of time series with multiple memberships and $N - N_O$ as the number of time series with single membership. Lowercase letters, such as y, denote scalars, while capital letters, like Y, represent matrices. For any matrix Y, we use $y_{i\cdot}$, $y_{\cdot j}$ and $y_{i,j}$ to denote its ith row, jth column and ij-th entry, respectively. For an arbitrary square matrix $Y \in \mathbb{R}^{n\times n}$, $\lambda_i(Y)$ refers to the ith eigenvalue of Y with $|\lambda_1(Y)| \geq |\lambda_2(Y)| \geq \cdots \geq |\lambda_n(Y)|$. The spectral radius of a square matrix Y is defined as $\rho(Y) = |\lambda_1(Y)|$. For a symmetric matrix Y, $\lambda_{max}(Y)$ and $\lambda_{min}(Y)$ denote the largest and smallest eigenvalues respectively. The following matrix norms are used: ℓ_2-norm: $\|Y\|$ and Frobenius norm: $\|Y\|_F = Tr\big(Y'Y\big)$.

A graph is defined as a triple: $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$ where $\mathcal{V} = 1, \cdots, N$ stands for the vertices of the graph, $\mathcal{E} \in \mathcal{V} \times \mathcal{V}$ the set of edges, and, accordingly, $\mathcal{W} \in \mathbb{R}$ the weights of all edges if necessary. A is the adjacency matrix which is the matrix representation of the graph. In this work, I assume that the adjacency matrix A is a $N \times N$ matrix where the elements $[A]_{ij} = w_{ij} \neq 0$ if there exists an edge connecting node j and node i and $[A]_{ij} = w_{ij} = 0$ otherwise.

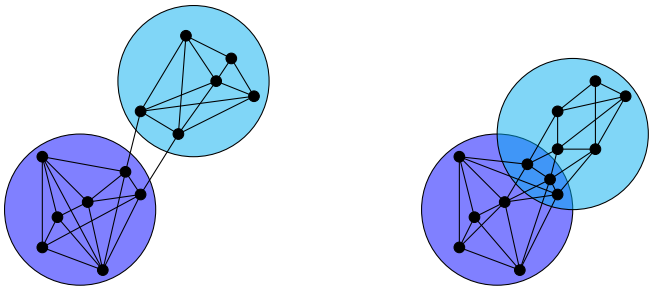
2 The Mixed Membership Model

The mixed membership model is an extension of traditional mixture models. Both are generative probabilistic models, but they differ in how they assign group membership. While a mixture model assumes that each individual belongs to a single group, the mixed membership model allows individuals to belong to multiple groups simultaneously, with each group association specified by a degree of membership. One of the most influential mixed membership models is Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan, 2003), a probabilistic generative model used in text analysis. LDA models a corpus by specifying how topics are distributed across documents and how words are generated within each topic.

2.1 Mixed Membership Stochastic Block Model

Another example is MMSBM, which is a probabilistic generative model for random graphs. Unlike the traditional stochastic block model (SBM) (Holland, Laskey and Leinhardt, 1983; Karrer and Newman, 2011), which assumes single membership group structure, the MMSBM allows for both single and multiple membership group structures. This flexibility arises because the nodes can possess multiple group memberships leading to overlapping community structures, whereas in the SBM, each node is assigned to a single group resulting in non-overlapping communities. Figure 1 illustrates the distinction between overlapping and non-overlapping group structures in a network.

Before giving the formal definition of MMSBM, We introduce additional notations. The matrix Φ_l for $l = 1, \ldots, p$ are the VAR coefficients up to the order p. The $(k \times k)$ matrix B is the community-specific probability matrix, where the on-diagonal elements specify the probability of links between variables within the same community and the off-diagonal elements indicate the probability of links between variables from different communities. The $(N \times k)$ matrix Z is the membership matrix with Z_i being the membership vector for variable



(a) Non-overlaps (b) Overlaps

Figure 1: In (a), all nodes are classified into two distinct, non-overlapping groups. There are more links within each group than between the groups, indicating stronger interactions among variables within the same group. In contrast, (b) shows a scenario where some variables belong to more than one group, resulting in overlapping structure. These nodes with multiple memberships exhibit strong interactions with variables in multiple groups, reflecting the mixed group affiliations.

$i = 1, \cdots, N$. The pair $(\underline{u}, \overline{u})$ defines the lower and upper bound for the $Uniform$ distribution. The vector $\vec{\alpha}$ serves as the hyper-parameter for $Dirichlet$ distribution. The vector $\vec{\pi}_i$ for $i = 1, 2, \ldots, N$ are the community stickiness vectors, specifying the degree to which each variable associates with each community. The $(N \times k)$ matrix $\boldsymbol{\pi}$ is the communities stickiness matrix, formed by stacking all the $\vec{\pi}_i$ vectors.

In the MMSBM, the membership matrix Z is no longer a fixed parameter. Instead, it serves as an indicator for a latent variable Z^*, which is sampled from a $Multinomial$ distribution with parameter $\boldsymbol{\pi}$. Here, $\boldsymbol{\pi}$ is composed of the vectors $\vec{\pi}_i$ for $i = 1, \ldots, N$, where each $\vec{\pi}_i$ represents a stickiness vector that specifies the degree to which a variable is associated with different groups. For example, if students in a school can participate in multiple hobby groups, $\vec{\pi}_i$ can be thought of as representing the proportion of time each student spends in the various hobby groups. Each element in $\vec{\pi}_i$ lies within the interval $[0, 1]$ and the elements sum to 1. This property is guaranteed by assuming that $\vec{\pi}_i$ for $i = 1, \ldots, N$ are drawn from a $Dirichlet$ distribution with hyper-parameter $\vec{\alpha}$ which is assumed to be given. The $Dirichlet$ distribution is commonly used for modeling probability vectors because of this normalization property. Additionally, the $Multinomial$ distribution is conjugate to the Dirichlet distribution, meaning that if the prior distribution is $Dirichlet$ and the likelihood

is $Multinomial$, then the posterior distribution remains $Dirichlet$. This conjugacy relationship simplifies the inference process. Since Z_i is sampled from a $Multinomial$ distribution parameterized by $\vec{\pi}_i$, the model allows for vertices to have multiple memberships across different groups. This characteristic results in an overlapping structure in the adjacency matrix.

The standard mixed membership stochastic block model can be formally defined as follows:

Definition 1. *(Mixed Menbership Stochastic Block Model). Given the hyper-parameter $\vec{\alpha}$ and the community-specific probability matrix B, the random graph is sampled in the following way:*

$$\vec{\pi}_i \sim Dirichlet(\vec{\alpha}) \; for \; i = 1, \ldots, N$$

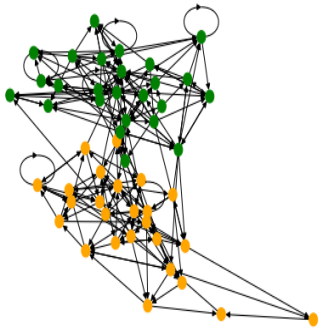
$$\begin{cases} Z_i^* | \vec{\pi}_i \sim Multinomial(\vec{\pi}_i) \\ Z_j^* | \vec{\pi}_j \sim Multinomial(\vec{\pi}_j) \end{cases} \quad for \; i, j = 1, \ldots, N$$

$$\begin{cases} Z_i = Indictor(Z_i^*) \\ Z_j = Indictor(Z_j^*) \end{cases} \quad for \; i, j = 1, \ldots, N$$

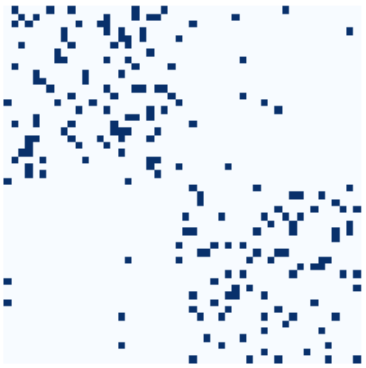
$$A_{ij} | Z_i, Z_j, B \sim Bernoulli(Z_i B Z_j') \; for \; i, j = 1, \ldots, N$$

Figure 2 illustrates an example of the MMSBM with single and multiple memberships. The top left panel shows a stochastic block graph with single membership. In this example, the number of nodes is set to 50 with two communities. The on-diagonal and off-diagonal elements of the community-specific probability matrix are set to 0.15 and 0.01, respectively. The latent membership vectors are configured such that the first 25 nodes belong entirely to group 1, while the remaining 25 nodes belong to group 2. To achieve this, the stickiness vectors $\vec{\pi}_i$ are set to $[1, 0]^T$ for the first 25 nodes and $[0, 1]^T$ for the last 25 nodes by fixing $p(\vec{\pi}|\vec{\alpha})$ to one. This effectively eliminates the influence of the hyperparameter $\vec{\alpha}$ in

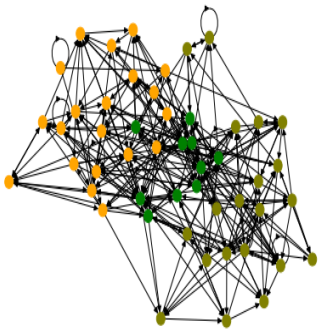
the generative model by treating $\vec{\pi}_i$ for $i = 1, \cdots, N$ as fixed parameters. The bottom left panel presents a stochastic block graph with multiple memberships. In this case, the stickiness vectors $\vec{\pi}_i$ for the first 20 and last 20 nodes are set to $[1, 0]^T$ and $[0, 1]^T$ respectively while while the vectors for the middle 10 nodes are set to $[0.5, 0.5]^T$, representing equal exposure to both communities. The top right and bottom right panels display heatmaps of the adjacency matrices corresponding to the graphs in the top left and bottom left panels, respectively. In these heatmaps, the elements are either "1" or "0" in which "1" represents the existence of the edge and "0" otherwise. The MMSBM generates a clear distinction between overlapping and non-overlapping structures based on the underlying latent stickiness vectors, reflecting spillover effects across communities.



(a) Realization of non-overlapping case in MMSBM.



(b) Heatmap of adjacency matrix for (a).



(c) Realization of overlapping case in MMSBM.

(d) Heatmap of adjacency matrix for (c).

Figure 2: the Mixed Membership Stochastic Block Model.

In the non-overlapping case, the nodes cluster into two distinct groups with no intersections between them. In contrast, the overlapping case features two sets of nodes that cluster into groups with regions of overlap, where some nodes belong to both groups. This difference highlights the flexibility of the MMSBM, which extends the traditional SBM by allowing for the simultaneous modeling of both overlapping and non-overlapping network structures.

3 Methodology

3.1 The Mixed Membership stochastic block VAR Model

Building on MMSBM, we introduce a novel network-based vector autoregressive model, termed the mixed membership stochastic block vector autoregression. The sampled panel data adopt the multiple-membership group structures through the MMSB-VAR(p). Specifically, the VAR coefficients in this model are treated as random matrices, sampled from a joint distribution defined by a probabilistic graphical model.

Before giving the formal definition of MMSB-VAR(p), we introduce additional notation. Let D_l for $l = 1, 2, \ldots, p$ represent N-dimensional normalized diagonal matrices defined as the sum of in-degree and out-degree for each node in an adjacency matrix. The in-degree of a vertex refers to the number of edges directed toward the vertex, while the out-degree is the number of edges directed away from it. The scalars ϕ_l for $l = 1, 2, \ldots, p$ are auxiliary parameters that assist in ensuring the stability of the vector autoregression.

We define the MMSB-VAR(p) as follows:

Definition 2. *Given the mixed membership stochastic block model as defined in Definition 1, the lower and upper bound of weights* $(\underline{u}, \overline{u})$ *and the scalars* ϕ_l *for* $l = 1, \ldots, p$*, the MMSB-*

VAR(p) is defined in the following way:

$$A_l \sim MMSBM(B, \vec{\alpha}) \; for \; l = 1, \dots, p$$

$$\Phi_{l,ij} | A_{l,ij}, (\underline{u}, \overline{u}) = \begin{cases} Uniform(\underline{u}, \overline{u}) & if \; A_{l,ij} = 1 \\ 0 & if \; A_{l,ij} = 0 \end{cases} \quad for \; i, j = 1, \dots, N$$

$$\mathbf{\Phi}_l = \phi_l D_l^{-1/2} \Phi_l D_l^{-1/2} \; for \; l = 1, \dots, p$$

$$Y_t = \beta_0 + \mathbf{\Phi}_1 Y_{t-1} + \mathbf{\Phi}_2 Y_{t-2} + \cdots + \mathbf{\Phi}_p Y_{t-p} + \epsilon_t$$

We remark that ϕ_l for $l = 1, \dots, N$ are given scalars. D_l represents the degree of Φ_l for $l = 1, \dots, N$. Accordingly, the matrix $\mathbf{\Phi}_l$ is the normalized version of Φ_l, preserving the same distribution as Φ_l for $l = 1, \dots, N$.

As noted by Sarkar and Bickel (2015) and Joseph and Yu (2016), the normalization step reduces the spread of the data points which improves the quality of group detection. This enhancement can significantly boost the algorithm's performance and establish consistency when the sample size goes to infinity. Furthermore, normalization is essential for ensuring the stability of the MMSB-VAR(p) along with the auxiliary scalars ϕ_l for $l = 1, \dots, N$ as discussed in Guðmundsson and Brownlees (2021). Lastly, normalization plays a key role in establishing the convergence properties of the VAR coefficients Φ_l when the sample size goes to infinity.

Based on the law of Bayesian network (Koller and Friedman, 2009), the MMSB-VAR(p) can be represented by a joint distribution:

$$\begin{aligned} &\mathbb{P}(\Phi, \{\vec{\pi}\}_{1:N}, Z, A | \underline{u}, \overline{u}, B, \vec{\alpha}) \\ &= \sum_{l=1}^{p} \mathbb{P}(\Phi_l | A, \underline{u}, \overline{u}) \{ \prod_{i,j}^{N} \mathbb{P}(A_{i,j} | B, \vec{z}_i, \vec{z}_j) \mathbb{P}(\vec{z}_i | \vec{\pi}_i) \mathbb{P}(\vec{z}_j | \vec{\pi}_j) \{ \prod_{n}^{N} \mathbb{P}(\vec{\pi}_n | \vec{\alpha}) \} \} \end{aligned} \tag{1}$$

We remark that the condition to ensure the stability of the MMSB-VAR(p) is the same as that for SB-VAR(p) (Guðmundsson and Brownlees, 2021). When the auxiliary scalars ϕ_l

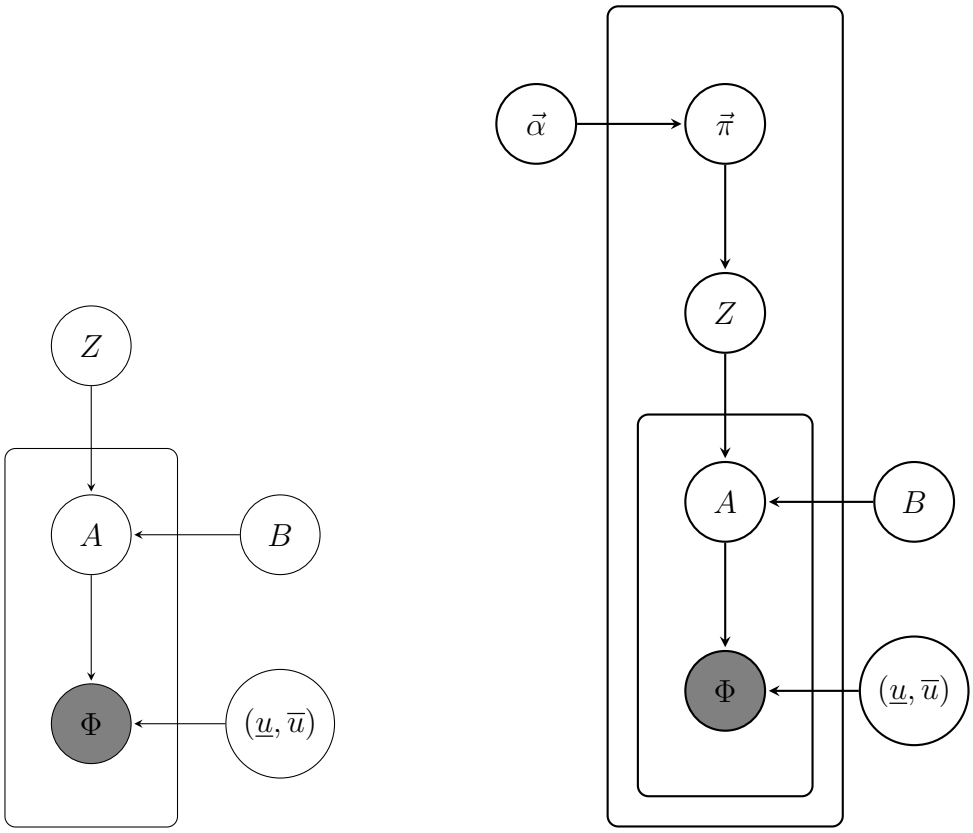
for $l = 1, \ldots, p$ meet the condition such that $\sum_{l=1}^{p} \phi_l < 1$, the MMSB-VAR(p) is stable.

For the sake of brevity, we assume that the MMSB-VAR(p) is stable throughout the paper.

3.2 Comparison between MMSB-VAR(p) and SB-VAR(p)

In this sub-section, we reanalyze SB-VAR(p) within the context of a probabilistic graphic model. When framed this way, SB-VAR(p) can be interpreted as a mixture model. We also compare SB-VAR(p) with MMSB-VAR(p) using probabilistic graph. In Figure 3, the distinction between SB-VAR(p) and MMSB-VAR(p) is evident. In the SB-VAR(p), Z_i is a membership vector that specifies the affiliation of variable i with groups, while B is a community-specific probability matrix as previously defined. Given these parameters, the adjacency matrix A_l is sampled from a $Bernoulli$ distribution. Using this graph structure, the VAR coefficients are then sampled from a $Uniform$ distribution with the parameters $(\underline{u}, \overline{u})$, representing the lower and upper bound, respectively. In contrast, the MMSB-VAR(p) shares the same two bottom tiers of the probabilistic generative model as the SB-VAR(p). However, the key difference lies in the treatment of the membership matrix Z. In the SB-VAR(p), Z is taken as a given matrix for all variables, whereas in the MMSB-VAR(p), Z is a random matrix sampled from a $Multinomial$ distribution with stickiness matrix $\boldsymbol{\pi}$. The stickiness matrix $\boldsymbol{\pi}$ is sampled from a $Dirichlet$ distribution with a given hyper-parameter vector $\vec{\alpha}$.

The MMSB-VAR(p) has several limitations. Firstly, although it captures both global and local patterns, it is not capable of generating hubs. Hubs, or central nodes, are those connected to a large number of other nodes in the graph, often resulting in a skewed degree distribution. This feature is commonly observed in real-world networks, where degree distributions often follow a power law. Secondly, the model assumes homogeneity in the expected graph, meaning that the probability of connections is assumed to be the same for all variables within the same group. However, real-world networks frequently exhibit heterogeneity within groups. A potential remedy for this issue is to incorporate a degree-corrected approach as proposed by Karrer and Newman (2011). Thirdly, the MMSB-VAR(p) model



(a) SB-VAR(p) (b) MMSB-VAR(p)

Figure 3: The graphical representation of SB-VAR(p) and MMSB-VAR(p) illustrates the structure and dependencies within these models. In the diagram, circles outside the 'plates' represent parameters, while circles inside the 'plates' denote variables of interest. Observations are indicated by grey circles. Both the adjacency matrix A and the VAR coefficients Φ share the same block structure, which is why they appear within the inner 'plate'.

struggles to represent hierarchical or densely overlapping group structures, which are common in social, business, and natural networks. A hierarchical structure occurs when nodes can be divided into groups, which can then be further subdivided into sub-groups. On the other hand, a densely overlapping structure implies that the more memberships two variables share, the more likely they are to be connected, leading to a denser block structure in overlapping regions. Finally, the model faces computational challenges as the number of states increases. The computational effort grows exponentially with the number of variables and communities, making it computationally expensive to apply the model to large-scale networks.

4 Algorithm

In this section, we introduce a two-step algorithm to detect a group structure within a panel. The first step employs a spectral clustering-based algorithm (Ng, Jordan and Weiss, 2001) and (Von Luxburg, 2007), which reliably classifies series with a single group membership into their true groups, while assigning series with multiple memberships to one of the potential groups to which they belong . In the second step, we analyze all pairs of partitions obtained from the first step identifying series with multiple memberships by applying a defined criterion to each pairwise comparison.

4.1 Step 1

It is worth noting that the first part of our algorithm is based on the "VAR Blockbuster" approach introduced in Guðmundsson and Brownlees (2021). For the current analysis, we assume that the number of communities is known; methods for determining the optimal number of communities will be discussed in the empirical section.

We assume that $\{Y_t\}_{t=1}^{T}$ represents the sample of observations from the MMSB-VAR(p).

Algorithm 1 Step 1: VAR Blockbuster Algorithm

Input: a set of panel data , Y_t, for t = 1, 2, 3, ..., T, with autoregressive order p and k communities.

1: Estimate the VAR coefficients, $\hat{\Phi}_l \; for \; l = 1, \ldots, p$ by using OLS.

2: Symmetrize the estimated coefficients: $\hat{\Phi}^S = \sum_{l}^{p} [\hat{\Phi}_l + \hat{\Phi}_l^T]$.

3: Form a matrix $\hat{U} \in \mathbb{R}^{N \times k}$ by computing the first k eigenvectors corresponding to the k largest eigenvalues of $\hat{\Phi}^S$.

4: Update $\hat{U} \in \mathbb{R}^{N \times k}$ from $\hat{U}$ by normalizing the rows of $\hat{U}$.

5: Cluster the nodes into different groups by applying k-means to the rows of $\hat{U}$.

Output: return the k-means partitions.

We introduce additional notation. Let $\hat{U}$ denote the $(N \times k)$ estimated eigenvector matrix

where the k columns correspond to the eigenvectors associated with the k largest eigenvalues of a given matrix. We define a $(k \times k)$ estimated diagonal eigenvalue matrix as $\hat{\Lambda}$ where the k diagonal elements correspond to the k largest eigenvalues of a given matrix.

First, assuming the invertibility of $\sum_{t=1}^{T} X_t X_t'$, we estimate the VAR coefficients of order p using ordinary least squares (OLS) for the panel Y_t. Next, we sum all estimated VAR coefficients and their transposes up to the order p. The summation is refered to as symmertised VAR coefficient. We denote this symmetrized VAR coefficient as $\hat{\Phi}^S$. It ensures that $\hat{\Phi}^S$ is symmetric and thus diagonalizable. Given that $\hat{\Phi}^S$ is diagonalizable, it can be decomposed into $\hat{U}\hat{\Lambda}\hat{U}'$ where $\hat{U}$ and $\hat{\Lambda}$ are the eigenvector matrix and eigenvalue matrix of $\hat{\Phi}^S$ respectively. Furthermore, the expected value $\mathbb{E}[\Phi^S]$, with respect to Equation 1, can be decomposed into $U\Lambda U'$. This decomposition reveals a unique relationship between U and $\boldsymbol{\pi}$: specifically, U represents a quasi-rotation of $\boldsymbol{\pi}$. Consequently, classifying U is approximately equivalent to classifying $\boldsymbol{\pi}$. The theoretical details of this relationship will be discussed in the following section. In the next step, we calculate the eigenvector matrix $\hat{U}$. We then normalize $\hat{U}$ on a row-wise basis, a step that has been shown to enhance the performance of spectral clustering algorithms (Joseph and Yu, 2016) and (Sarkar and Bickel, 2015). Finally, we cluster all rows of $\hat{U}$ using k-means clustering. This approach classifies all series into one group with any series exhibiting multiple memberships being assigned to one of the groups that they belong to. Details on the consistency of this classification method will be discussed in the theoretical section.

4.2 Step 2

Given the k partitions obtained in the first step, we identify series with multiple memberships on a pairwise basis. We define the weighted node-group degree as follows:

Definition 3. *(Weighted Node-Group Degree). Given an undirected adjacency matrix with group partitions, weighted node-group degree is the sum of the edge weights for edges in a group incident to a specific node.*

The weighted node-group degree serves as a criterion for determining whether a node belongs to a particular group.

Algorithm 2 Step 2: Pick-up Algorithm

Input: k **partitions from step one.**

1: Pair up k partitions.

2: Update $\hat{\Phi}^S$ by de-meaning $\hat{\Phi}^S$ on row basis.

3: For each pair, calculate weighted node-group degree for each node in one group against another group, sequentially move each node with positive weighted node-group degrees to a new overlapping set from original partitions.

4: Update k partitions by moving nodes in every overlapping set back to the partitions in a corresponding pair.

5: Repeat step 3 and 4 with updated k partitions $k-1$ times .

Output: k non-overlapping partitions and C_k^2 overlapping partitions.

In the second step, we begin by pairing up the k partitions resulting in C_k^2 pairs. Next, we de-mean $\hat{\Phi}^S$ on a row-wise basis. The MMSB-VAR(p) operates under the assumption that within-group spillovers are significantly larger than those between groups. Consequently, this step serves as a normalization making within-group spillovers positive and between-group spillovers negative. Following this, we identify nodes with multiple memberships in each pair. For each node in one partition of a given pair, we calculate the weighted node-group degree, as previously defined, with respect to another partition in the pair. Nodes with a positive weighted node-group degree are sequentially moved from their original partitions to a new set termed the "overlapping set." This operation is repeated for each pair. In each round, a node with multiple memberships is assigned to one more group. In total, a node can belong to up to k groups. To ensure that all multiple-membership nodes are classified into each group to which they belong, we perform the pair selection process $k-1$ times. At the end of each pair selection, we update the k partitions by moving nodes in the overlapping sets back to their corresponding partitions and we clear the C_k^2 overlapping sets. The operations correspond to step 3 and 4. Finally, we obtain k non-overlapping partitions

where each node belongs to only one group and C_k^2 overlapping partitions where each node belongs to both groups in a pair. The consistency of the algorithm will be discussed in the theoretical section.

5 Theoretical Analysis

In this section, we begin with a population-level analysis and, under specific assumptions, establish the existence of a quasi-rotation matrix in population for the membership stickiness matrix $\boldsymbol{\pi}$, which govern the group memberships of series in the panel. This result is essential for spectral clustering because $\boldsymbol{\pi}$ in the MMSB-VAR(p) model is latent and cannot be directly observed. Instead, we rely on panel data to compute the eigenvector matrix of the Granger causality network implied by vector autoregression. If a quasi-rotation of $\boldsymbol{\pi}$ exists, we demonstrate that clustering the eigenvector matrix U of $\mathbb{E}(\Phi)$ is equivalent to clustering the latent membership stickiness matrix $\boldsymbol{\pi}$.

The first step establishes the effectiveness of spectral clustering when applied to the MMSB-VAR(p) model in population. Subsequently, errors arise from the deviation of the estimated symmetrized VAR coefficient $\hat{\Phi}^S$ from its population counterpart, $\mathbb{E}(\Phi)$ with respect to the MMSB-VAR(p) defined above. This lays the foundation for proving the consistency of the algorithm, provided that such a kind of errors are bounded.

Secondly, we separately establish the consistency of steps 1 and 2 in our algorithm. Our method extends the approach used in Rohe, Chatterjee and Yu (2011). In both steps, we define a misclassified ratio. If the misclassified ratios for both step 1 and 2 are bounded, the consistency of the two-step algorithm is established.

5.1 Population Analysis

This subsection aims to establish that spectral clustering works well when applied to $\mathbb{E}(\Phi)$ with respect to the MMSB-VAR(p) model.

Assumption 1. *Given a $n \times n$ probability-specific matrix B for the family of stochastic block graphs, $(B + B')$ is positive definite.*

The probability-specific matrix B encodes the likelihood of edge formation between nodes, where diagonal elements represent the probabilities of edges forming between nodes within the same group, and off-diagonal elements represent the probabilities for nodes from different groups. Assumption 1 ensures that the diagonal elements of B are larger than the off-diagonal elements implying that intra-group connections are consistently more likely than inter-group connections. This condition is essential to establishing a group structure in the MMSB-VAR(p) model. As noted in Guðmundsson and Brownlees (2021), such random graphs are often homophilic meaning they exhibit a natural tendency for similar nodes to connect. Thus, Assumption 1 imposes a homophilic structure that underpins the group detection framework of the MMSB-VAR(p).

Lemma 1 establishes that the population normalized eigenvectors can be approximately factorized into a product of stickiness vectors and a square matrix. This result implies that clustering the population normalized eigenvectors is approximately equivalent to clustering the underlying stickiness vectors of all nodes. It extends Lemma 3.3 in Qin and Rohe (2013) and Lemma 2.1 in Lei and Rinaldo (2015).

Lemma 1. *Given a MMSB-VAR(p) defined above. Let Φ^S be symmertised autoregressive coefficient from the MMSB-VAR(p) and let $U\Lambda U'$ be the eigen-decomposition of $\mathbb{E}(\Phi^S)$ with $U'U = I_k$ and a diagonal eigenvalue matrix $\Lambda \in \mathbb{R}^{k\times k}$.*

With $\mathbb{E}(\Phi^S) = \frac{1}{2}(\underline{u} + \overline{u})M(\boldsymbol{\pi}B\boldsymbol{\pi}')$, where $\frac{1}{2}(\underline{u}, \overline{u})$ is the mean of Uniform distribution, M is the number of trials for Multinomial distribution and $M(\boldsymbol{\pi}B\boldsymbol{\pi}')$ is the mean of Multinomial distribution , then we roughly have $\mathbb{E}(\Phi^S) \approx \boldsymbol{\pi}B\boldsymbol{\pi}'$ since $\frac{1}{2}(\underline{u} + \overline{u})M$ is a scalar which does not affect the relative positions of $\boldsymbol{\pi}$,

1. *There exists a quasi-rotation matrix $R \in \mathbb{R}^{K\times K}$ that $U \approx \boldsymbol{\pi}R$, where the column space of $\boldsymbol{\pi}$ is approximately equal to that of U.*

2. $\vec{\pi}_i = \vec{\pi}_j$ *if and only if* $\vec{\pi}_i R = \vec{\pi}_j R$.

Lemma 1 establishes that, following a quasi-rotation transformation, the eigenvector representations of nodes within the same group are no longer perfectly aligned in the same positions. However, these representations remain closer to each other within the same group than to those of nodes in different groups. This preserved proximity within groups ensures that the algorithm continues to effectively distinguish and identify the group structure.

5.2 Consistency of Step 1

In this subsection, we extend the method from Rohe, Chatterjee and Yu (2011) to prove the consistency of the first step by bounding the misclassification ratio. Since we have proved that the cluster spectral works well with MMSB-VAR(p) in population, the deviation comes from the difference between observed graph and the expectation of the graph. In this case, the observed graph is the estimated symmertised VAR coefficient, $\hat{\Phi}^S$ estimated from the panel generated from the MMSB-VAR(p) and the expectation of the graph is $\mathbb{E}(\Phi)$ with respect to the MMSB-VAR(p). Hence, the misclustering comes from the deviation of $\hat{\Phi}^S$ from $\mathbb{E}(\Phi)$. If the the deviation of $\hat{\Phi}^S$ from $\mathbb{E}(\Phi)$ is bounded, then misclustering ratio is also bounded. If this deviation is bounded, the misclassification ratio is also bounded, ensuring the consistency of the algorithm.

Before proceeding to the details, we introduce additional notation and present some assumptions regarding the MMSB-VAR(p). We define $\|X\|$ as the spectral norm of any $(N \times N)$ matrix X. We define $\Omega\Big(\ \cdot\ \Big)$ as the asymptotic lower bound of a function meaning that a function $f(n)$ is said to be $\Omega\Big(g(n)\Big)$ if there exist positive constants C and n_0 such that $C|g(n)| \leq |f(n)|$ for all $n \geq n_0$. Similarly, we define $O\Big(\ \cdot\ \Big)$ as the asymptotic upper bound of a function meaning that a function $f(n)$ is said to be $O\Big(g(n)\Big)$ if there exist positive constants C and n_0 such that $|f(n)| \leq C|g(n)|$ for all $n \geq n_0$.

Assumption 2. *Let* $B_N = \min_{ij}\big[B\big]_{ij}$ *,* $\underline{\alpha} = \min_i\big[\alpha_i\big]$ *and* $\alpha_0 = \sum_{i=1}^{N}\alpha_i$ *, where* $\alpha_i \in \vec{\alpha}$ $for\ i = 1, \ldots, N$*. Assume that* $B_N = \Omega\Big(\frac{log(N)}{N}\Big)$.

Assumption 2 states that the probability of edges forming between nodes should be at least on the order of $log(N)/N$ (Sarkar and Bickel, 2015), which implies that the expected number of edges in the graph should also be at least $log(N)/N$. This secures the sparsest regime under which communities in a graph can still be recovered exactly (Abbe, Bandeira and Hall, 2015). Assumption 2 is satisfied as long as the expected number of edges grows proportionally to N^2 ensuring that the graph remains sufficiently dense for accurate community detection.

Next, we assume that the underlying mixed membership random graphs $\mathcal{G}_1, \ldots, \mathcal{G}_p$ are fixed, as is the symmetrized VAR coefficient matrix Φ^S. Based on these fixed graphs, we impose several assumptions regarding the sampling properties of the data, conditional on the realizations of the random graphs. Specifically, appropriate distributional and dependence assumptions are required for isotropic random vectors such as $\Sigma_Y^{-1/2}Y_t$ where $\Sigma_Y = \mathbb{E}\big[Y_tY_t'|\mathcal{G}_1, \ldots, \mathcal{G}_p\big]$. Let $\mathcal{B}_{-\infty}^{r}$ and $\mathcal{B}_{r+m}^{\infty}$ be the σ-algebras generated by $\{\Sigma_Y^{-1/2}Y_t : -\infty \leq t \leq r\}$ and $\{\Sigma_Y^{-1/2}Y_t : r+m \leq t \leq \infty\}$ respectively. Then, the α-mixing coefficient is defined as $\alpha(m) = \sup_r \sup_{A\in\mathcal{B}_{-\infty}^{r}, B\in\mathcal{B}_{r+m}^{\infty}} |\mathbb{P}\big(A\cap B|\mathcal{G}_1, ..., \mathcal{G}_p\big) - \mathbb{P}\big(A|\mathcal{G}_1, ..., \mathcal{G}_p\big)\mathbb{P}\big(B|\mathcal{G}_1, ..., \mathcal{G}_p\big)|$.

Assumption 3. *let* $\{Y_t\}$ *be MMSB-VAR(p), several assumptions follow:*

1. $\{\epsilon_t\}$ *is covariance-stationary, serially uncorrelated and* $\mathbb{E}[Y_{t-1}\epsilon_t'|\mathcal{G}_1, ..., \mathcal{G}_p] = 0$ *for all t.*

2. $\big|\big|\mathbb{E}[\epsilon_t\epsilon_t']\big|\big| < \infty$ *and* $\big|\big|\mathbb{E}[\epsilon_t\epsilon_t']^{-1}\big|\big| < \infty$.

3. $\{\Sigma_Y^{-1/2}Y_t\}$ *is strongly mixing with coefficient satisfying* $\alpha(m) \leq e^{-C_1m^{\gamma_1}}$ *where m is positive integer and* γ_1 *and* C_1 *are positive constants.*

4. *For any vector* $\mathbf{x}$ *satisfying* $||\mathbf{x}|| = 1$ *and for any s > 0,* $\mathbb{P}\big(\big|\mathbf{x}'\Sigma_Y^{-1/2}Y_t\big| > s\big|\mathcal{G}_1, ..., \mathcal{G}_p\big) \leq e^{1-(s/C_2)^{\gamma_2}}$*, for all t and* C_2 *and* γ_2 *are positive constants.*

5. $T = \Omega(N^{\frac{2}{\gamma}-1})$, *where* $1/\gamma = 1/\gamma_1 + 1/\gamma_2$ *and* $\gamma < 1$.

Assumption 3 consists of a set of standard conditions commonly used in the analysis of multiple time series (Fan, Liao and Mincheva, 2013). Assumption 3.1 imposes a standard independence condition on the innovations, conditional on the realizations of random graphs. Assumption 3.2 bounds the covariance of the innovations in the MMSB-VAR(p) from above and below by constants independent of N. Assumption 3.3 requires that the α-mixing coefficients, which measure temporal dependence, decay sufficiently fast. Assumption 3.4 imposes generalized exponential tails on the distribution of $\Sigma_Y^{-1/2} Y_t$ implying that all elements of this vector exhibit such tails (Fan, Liao and Mincheva, 2013). Finally, Assumption 3.5 specifies that the sample size T must grow faster than the cross-sectional dimension N. Together, Assumptions 3.3–3.5 enable the use of the concentration inequality from Merlevède, Peligrad and Rio (2011) which is instrumental in proving the consistency of Step 1 in the algorithm.

The following lemma demonstrates that the norm difference between the population and the estimated symmetrized autoregressive coefficient is bounded. This result forms the basis for establishing the algorithm's reliability under the stated assumptions.

Lemma 2. *Given a MMSB-VAR(p) in Definition 2 satisfying Assumption 2 and Assumption 3, the group-specific probability matrix B satisfying Assumption 1, let $\hat{\Phi}^S$ be the estimated symmertised autoregressive coefficient, Φ^S the sample symmertised autoregressive coefficient and $\overline{\Phi}^S$ the population symmertised autoregressive coefficient. Then, it is with high probability that*

$$\|\hat{\Phi}^S - \overline{\Phi}^S\| \leq O_p\left(\sqrt{\frac{\alpha_0 log(N)}{\underline{\alpha} N B_N}} + \frac{N^{3/2}}{\sqrt{T}} \lambda_{max}(\Phi^S)\right).$$

The result of Lemma 2 is instrumental in establishing Lemma 3, where we aim to bound the deviation of the estimated eigenvector representations from its population counterpart. These eigenvector representations are derived from the population and estimated symmetrized autoregressive coefficients, respectively. By leveraging the bounded differ-

ence established in Lemma 2, we ensure that the deviation between the estimated and true eigenvector representations remains controlled, providing a foundation for the algorithm's consistency.

Lemma 3. *Given a MMSB-VAR(p) defined in definition 2 which satisfies 1,2 and 3 of assumption (3). Let* $\hat{\mathbf{X}}$ *be the estimated row-normalized eigenvector matrix and* $\mathbf{X}$ *its population counterpart. Then, it is with high probability that*

$$||\hat{\mathbf{X}} - \mathbf{X}\mathcal{O}|| \leq O_p\left(\sqrt{\frac{\alpha_0 log(N)}{\underline{\alpha} B_N}} + \frac{N^2}{\sqrt{T}}\right),$$

where $\mathcal{O}$ *is a* $k \times k$ *orthonormal rotation matrix which relies on* $\hat{\mathbf{X}}$ *and* $\mathbf{X}$.

Lemma 3 establishes that the norm difference between the population and estimated eigenvector representations is bounded. This result is then used to show that, after applying k-means to the population and estimated eigenvector representations, the norm difference between the resulting population and estimated k-means centroids is also bounded. This result, in turn, forms the foundation for the following theorem.

The following theorem establish the consistency of the first step of the algorithm.

Theorem 1. *Given a MMSB-VAR(p) in definition 2 satisfying Assumption 1, 2 and 3. Let* $\mathcal{M}$ *be the set of misclustered variables.*

Then, it is with high probability that

$$\frac{|\mathcal{M}|}{N} = O_p\left(\frac{\alpha_0 log(N)}{\underline{\alpha} N B_N} + \frac{N^3}{T}\right)$$

Theorem 1 establishes that the displacement ratio is bounded with high probability under specific conditions: N/T remains sufficiently small as N and T approach infinity, the latent random graphs are sufficiently dense, and the difference between the maximum and minimum elements of the hyper-parameter $\vec{\alpha}$ is relatively small. Additionally, the theorem further demonstrates that nodes with multiple memberships are correctly classified into

one of the groups to which they belong. Furthermore, the misclassification ratio is shown to be bounded above with high probability, guaranteeing robust group detection even in the presence of multiple memberships structure.

5.3 Consistency of Step 2

The approach to proving the consistency of step 2 parallels that of step 1. Specifically, we first prove that the algorithm is consistent in the population case, meaning it performs well when applied to $\mathbb{E}(\Phi^S)$. The errors arise from the deviation when the algorithm is applied to the observed graph, $\hat{\Phi}^S$. We then prove that this deviation is bounded, thereby ensuing the consistency of step 2.

Before delving into the details, we present the following assumptions:

Assumption 4. *let $\{Y_t\}$ be panel from MMSB-VAR(p) and $\mathbb{E}(\Phi^S) \approx 2\boldsymbol{\pi}B\boldsymbol{\pi}'$ the expected VAR coefficient with respect to the MMSB-VAR(p), several assumptions follow:*

1. *If $N_{total} = \sum_i \Big(\#\ of\ memberships\ for\ node\ i\Big)$, then $N_{total} = \Omega\Big(cN\Big)$ and $N_{total} = O\Big(dN\Big)$ where $1 \leq c \leq d$ are positive constants.*
2. *$B_{i,i} = B_{j,j}$ for $i \neq j$ and $i, j = 1, \cdots, N$.*
3. *Given a non-pure mixed membership π_i , $\pi_i(k) \approx \pi_i(d)$ for $k \neq d$ and $k, d = 1, \cdots, K$.*

Assumption 4.1 bounds the complexity of the group structure by requiring the number of multiple-membership series to grow as functions of the cross-sectional dimension N. This growth rate is governed by two positive constants, c and d, which define the lower and upper bounds of the structure's complexity. To ensure the group structure remains manageable, we argue that d should be moderate, limiting the number of multiple-membership nodes and preventing the structure from becoming overly complex. Assumption 4.2 requires that the probabilities of edge formation within each group of the panel are identical. This condition ensures that, regardless of which group a multiple-membership series is classified into,

its edge-forming probability with existing series in the group remains the same. Although this is a strict assumption, it is essential for proving the next lemma. Once consistency in the population case is established under this assumption, any deviation can be treated as source of error, without invalidating the overall proof strategy. Assumption 4.3 establishes that, for any multiple-membership series, the stickiness with each group is identical. This strict assumption ensures that the density of the overlapping regions is the equal to that of non-overlapping regions. Consequently, when a multiple-membership series is classified into one of its groups, the edge-forming probability between this series and all other series in the group matches the edge-forming probability between nodes already within the group. As with previous assumptions, if consistency in the population case is established under this condition, any deviation can be treated as source of error without invalidating the proof strategy. Finally, given a $\mathbb{E}(\Phi^S)$ satisfying assumption 4, we have that $\mathbb{E}(\Phi^S) \approx 2\boldsymbol{\pi}B\boldsymbol{\pi}' \approx 2$ where

$$\begin{bmatrix} Z_1(\pi_1)BZ_1(\pi_1) & \cdots & Z_1(\pi_1)BZ_{N-1}(\pi_{N-1}) & Z_1(\pi_1)BZ_N(\pi_N) \\ Z_2(\pi_2)BZ_1(\pi_1) & \cdots & Z_2(\pi_2)BZ_{N-1}(\pi_{N-1}) & Z_2(\pi_2)BZ_N(\pi_N) \\ \vdots & \ddots & \ddots & \vdots \\ Z_N(\pi_N)BZ_1(\pi_1) & \cdots & Z_N(\pi_N)BZ_{N-1}(\pi_{N-1}) & Z_N(\pi_N)BZ_N(\pi_N) \end{bmatrix}_{(N\times N)}$$

$Z_i(\pi_i)$ is a function of π_i. It's no longer fixed taken as given. Under assumption 4, for each block of $\boldsymbol{\pi}B\boldsymbol{\pi}'$, all edge-forming probabilities inside are the same. This assumption serves as a foundation for proving Lemma 4.

In the next lemma, we prove that step 2 of the algorithm is effective in the population case building on the partition results established in step 1.

Lemma 4. *Given $\mathbb{E}(\Phi^S)$ with respect to the MMSB-VAR(p) satisfying assumption 4, A_{group} is a reconstructed group-specific sub-adjacency of $\mathbb{E}(\Phi^S)$ building on the partition results in step 1.*

We define d_j as the degree of vertex j in a graph, let λ and $\vec{v}$ be a specific eigenvalue and eigenvector of A_{group}, let v_j be j-th element in $\vec{v}$ and v_{max} the maximum in $\vec{v}$, then

$$0 \leq \lambda \leq d_j.$$

Lemma 4 establishes that, for each reconstructed sub-adjacency matrix A_{group} of $\mathbb{E}(\Phi^S)$ based on the partitions from step 1, all node degrees within a group are non-negative. Specifically, in step 1 of the algorithm, we form k non-overlapping groups and C_k^2 overlapping groups. In step 2, to determine whether a series in the overlapping set belongs to a specific group, we check whether its node degree with that group is positive or negative. If a series belongs to the group, Lemma 4 ensures that its degree with the group is non-negative; otherwise, it is negative. Thus, Lemma 4 confirms that step 2 of the algorithm is effective in the population case.

Next, we focus on bounding the deviation of the eigenvalues of the estimated group sub-adjacency matrix from those of its population counterpart. In the population case, for each sub-adjacency matrix formed by the partition results in step 1, all eigenvalues are non-negative and are less than or equal to the corresponding node degrees. Thus, bounding the deviation of the eigenvalues is sufficient to bound the deviation of the node degrees between the estimated and population versions.

The next lemma bounds the deviation of the eigenvalues in the estimated group-specific sub-adjacency matrix from those in its population counterpart.

Lemma 5. *Given the partition results in step 1, the estimated VAR symmertised coefficient,* $\hat{\Phi}^S$ *and the population VAR coefficient,* $\mathbb{E}(\Phi^S)$*, define* $\hat{A}_{group}$ *and* $\overline{A}_{group}$ *as the reconstructed group-specific sub-adjacency of* $\hat{\Phi}^S$ *and* $\mathbb{E}(\Phi^S)$ *based on partition results in step 1, let* $\hat{\lambda}^*_{i,l}$ *and* $\overline{\lambda}^*_{i,l}$ *be eigenvalues of* $\hat{A}_{group}$ *and* $\overline{A}_{group}$ *for each series* n *and each time lag* l *, then*

$$\sum_{l=1}^{p}\sum_{i=1}^{N}|\hat{\lambda}^*_{i,l}-\overline{\lambda}^*_{i,l}| \leq O_p\left(\sqrt{\sqrt{\frac{\alpha_0 log(N)}{\underline{\alpha}NB_N}}+\frac{N^{3/2}}{\sqrt{T}}\lambda_{max}(\Phi^S)}\right).$$

The result of Lemma 5 is instrumental in establishing Theorem 2. Lemma 5 bounds the deviation of the eigenvalues between their estimated and population versions, which

accounts for the misplacement errors during the implementation of step 2 of the algorithm. In Theorem 2, we bound the misplacement ratio in step 2 of the algorithm.

Theorem 2. *Given a MMSB-VAR(p) in definition 2 satisfying Assumption 1, 2, 3 and 4. Let* $\mathcal{N}$ *be the set of misclustered series and* $N_{total} = \sum_{i=1}^{N} \Big(\# \; of \; memberships \; for \; node \; i \Big)$.

Then, it is with high probability that

$$\frac{|\mathcal{N}|}{N_{total} - N} = O_p\left(\frac{1}{c-1} \sqrt{ \sqrt{\frac{\alpha_0 log(N)}{\underline{\alpha} N^5 B_N}} + \frac{N^{1/2}}{\sqrt{T}} } \right).$$

Theorem 2 establishes the similar argument as in theorem 1 but adds an additional conclusion: as the number of multiple-membership series increases, the block structure becomes more complicated, leading to a higher misplacement ratio.

With the consistencies of both step 1 and step 2 established, their combination confirms the overall consistency of the two-step classification algorithm.

6 Simulation

In this section, we conduct two simulation exercises to evaluate the finite-sample properties of our two-step VAR classification algorithm. We examine two scenarios: one in which each time series is restricted to a single group membership and another in which multiple group memberships are permitted.

Repeated samples of size T, set respectively to 2000, 5000, 8000, 10,000, 15,000, and 20,000, are drawn from an MMSB-VAR(1) model. The number of groups is fixed at $k = 5$ for both studies. The community pair-specific probabilities are specified such that $b_{i,i} = b_{j,j}$ and $b_{i,j} = b_{j,i}$. We test three parameter settings for $(b_{i,i}, b_{i,j})$: (0.25,0.01), (0.25,0.05) and (0.5, 0.01). The total number of time series, N, is fixed at 100. Edge weights are drawn uniformly from the interval $(\underline{u}, \overline{u}) = [0.3, 1.0]$. To ensure stationarity, the autoregressive coefficient ϕ_1 is set to 0.99. The model's innovations are drawn independently and identically from

a multivariate standard normal distribution. Stickiness vectors $\vec{\pi}_i$ are specified such that $P(\vec{\pi}_i|\vec{\alpha}) = 1$ for $i = 1, 2, .., N$. This approach effectively fixes $\vec{\pi}_i$ and treats it as given. The configuration of $\vec{\pi}_i$ determines the group membership for each series and, consequently, the block structure of the graph, as discussed in Sections 6.1. Each experiment is repeated 500 times, and the results are summarized and discussed separately for the single-membership and multiple-membership cases in Sections 6.2 and 6.3. Results from additional simulation studies are provided in the Online Appendix.

6.1 Interpretation of Measurement

First of all, if the stickiness vector $\vec{\pi}_i$ is degenerate, meaning that there is only one element equal to 1 while all other elements are 0, then the membership vectors z_i, sampled from a Multinomial distribution with parameter $\vec{\pi}_i$ and guided by an indicator function, always mirrors $\vec{\pi}_i$. Simply put, for any degenerate $\vec{\pi}_i$, we have $\vec{z}_i = \vec{\pi}_i$ under this setup. For example, if $\vec{\pi}_i = [1, 0, 0]$, then $\vec{z}_i$ is always sampled as $[1, 0, 0]$. In this case, the MMSB-VAR(p) reduces to SB-VAR(p). On the other hand, if $\vec{\pi}_i$ is non-degenerate containing more than one non-zero element with all elements summing to 1, then $\vec{z}_i$ is always sampled as a degenerate vector in which the value 1 is likely to appear at any position corresponding to a non-zero element in $\vec{\pi}_i$. For instance, in a three-group case, if $\vec{\pi}_i = [0.9, 0.1, 0]$, it is highly likely to sample $\vec{z}_i = [1, 0, 0]$, less likely to sample $\vec{z}_i = [0, 1, 0]$ and impossible to sample $\vec{z}_i = [0, 0, 1]$. In the MMSB-VAR(p) model, non-degenerate $\vec{\pi}_i$ is assumed to be evenly distributed. For example, by taking $\vec{\pi}_i$ as given, we set $\vec{\pi}_i = [0.5, 0.5]$ for a two-group case or $\vec{\pi}_i = [0.33, 0.33, 0.33]$ for a three-group case. This setup ensures that, for a time series i, $\vec{z}_i$ evenly points to any group with which it has affiliation e.g. where the corresponding element in $\vec{\pi}_i$ is non-zero. In this work, we define a time series as exhibiting multiple memberships if it is likely to belong to multiple groups as reflected variability in its sampled $\vec{z}_i$.

Secondly, in the SB-VAR(p) model, all membership vectors $\vec{z}_i$ are assumed to be fixed

and predetermined when calculating each element-wise parameter, $\vec{z}_i B \vec{z}_j'$, with which $A_{i,j}$, an element of the adjacency matrix, is sampled. In contrast, in the MMSB-VAR(p) model, the membership vectors $\vec{z}_i$ are no longer fixed but are instead sampled element by element when calculating element-wise parameter $\vec{z}_i B \vec{z}_j'$ which is then used to sample each $A_{i,j}$. As a result, the variability in element-wise $\vec{z}_i$ reflects multiple memberships for series i corresponding to the stickiness vector $\vec{\pi}_i$.

Thirdly, as analyzed in Section 4, the algorithm outputs k sets, referred to as single-membership sets, where all series within each set have single membership. Additionally, the algorithm produces $2^{\mathrm{C}_k^2}$ sets, referred to as pair-membership sets, which correspond to pairs of groups where all series are affiliated with both groups. By pooling all pairs, we simply derive the multiple-membership structure for all series. Therefore, we evaluate the performance of algorithm separately for the k single-membership sets and the $2^{\mathrm{C}_k^2}$ pair-membership sets.

Finally, we define the measurement criteria and evaluate our two-step classification algorithm using panels sampled under the setup described above. We define four criteria: hit ratio (Single Membership), misplacement ratio (Single Membership), hit ratio (Multiple Membership), and misplacement ratio (Multiple Membership). The hit ratio (Single Membership) is the ratio of series with degenerate $\vec{\pi}_i$ correctly classified into their corresponding groups since degenerate $\vec{\pi}_i$ leads to unique $\vec{z}_i$. The numerator represents the number of series with single membership accurately assigned to the correct groups. The denominator is the total number of series with degenerate $\vec{\pi}_i$. In contrast, the misplacement ratio (Single Membership) is the ratio of series with degenerate $\vec{\pi}_i$ incorrectly classified into groups. The numerator is the number of series with degenerate $\vec{\pi}_i$ assigned to groups inconsistent with their $\vec{\pi}_i$, while the denominator is the total number of series with degenerate $\vec{\pi}_i$. Notably, the sum of the hit and misplacement ratios for single membership is not 1, as some single-membership series may be classified as multiple-membership series and counted in the hit and misplacement ratios for multiple memberships. In addition to the k sets representing

clusters of series with degenerate $\vec{\pi}_i$, the algorithm outputs $2^{\mathrm{C}_k^2}$ sets representing clusters of series with non-degenerate $\vec{\pi}_i$. For the hit ratio (Multiple Membership), the denominator is the total number of group affiliations across all series. For instance, if $\vec{\pi}_1 = [0.5, 0.5, 0]$ and $\vec{\pi}_1 = [0, 0.5, 0.5]$, series 1 is affiliated with groups 1 and 2, and series 2 is affiliated with groups 2 and 3, resulting in a denominator of 4. Accordingly, the numerator is the total number of affiliations correctly identified by the algorithm across all series. Similarly, the misplacement ratio (Multiple Membership) is the ratio of misclassified group affiliations. For example, if a series affiliated with groups 1 and 2 is incorrectly classified as affiliated with groups 2 and 3, such misclassifications contribute to the numerator of the misplacement ratio. The denominator for the misplacement ratio is the same as that for the hit ratio in multiple-membership cases. We note that the sum of the hit and misplacement ratios for multiple memberships is also not 1, as some multiple-membership series may be incorrectly classified as single-membership series and included in the calculation of the single-membership hit and misplacement ratios described above.

6.2 Single Membership Case

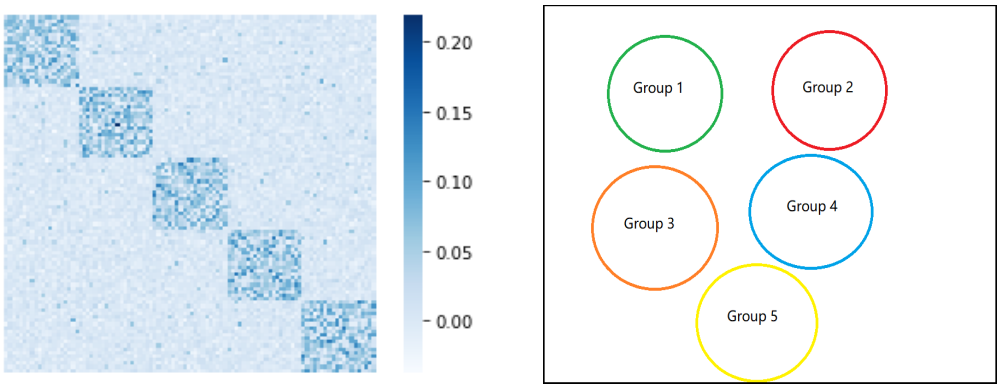
In this study, $\vec{\pi}_i$ are set up as follows:

$\{\vec{\pi}_i\}_{i=1}^{20} = [1,0,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=21}^{40} = [0,1,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=41}^{60} = [0,0,1,0,0]^T$, $\{\vec{\pi}_i\}_{i=61}^{80} = [0,0,0,1,0]^T$, $\{\vec{\pi}_i\}_{i=81}^{100} = [0,0,0,0,1]^T$.

This setup implies that all $\vec{\pi}_i$ vectors are degenerate, meaning all series exhibit single membership. The left panel of Fig. 4 displays the heatmap of the estimated symmetrized VAR coefficients, $\hat{\Phi}^S$. As observed, under this setup, it is evident that there are no overlapping blocks. The right panel of Fig. 4 depicts the block structure using a Venn diagram, further confirming the absence of overlap between groups.

The results are summarized in Table 1. Since there are no multiple-membership series in this setup, the multiple-membership hit ratio is left empty. However, because it is still possible for single-membership series to be misclassified as multiple-membership series, we

report the misplacement ratio for multiple membership. In this case, the denominator for the misplacement ratio (multiple membership) is the total number of series, as all series are single-membership, and the number of multiple-membership series is zero. From Table 1, we observe that the algorithm performs well when the ratio $(b_{i,i}/b_{i,j})$ and the sample size T are sufficiently large. The hit ratio is high, with the trade-off that a small number of series are misclassified as multiple-membership (explaining why the misplacement ratio is not 0). These findings are consistent with the theoretical expectations.



(a) Heatmap of adjacency matrix for estimated symmertized VAR matrices

(b) Illustration for overlapping structure

Figure 4: Panel (a) displays the heatmap of the estimated symmetrized VAR matrices on the panel sampled from the MMSB-VAR(1) model. Panel (b) presents a Venn diagram illustrating the overlapping structure. As observed, all series are classified into five distinct groups without any overlap.

Table 1: **Accuracy Table for First Simulation with 100 single-membership series.**

	n=100					
b_{11},b_{22}/b_{12},b_{21}	2000	5000	8000	10000	15000	20000
Hit Ratio (Single Membership)						
0.25/0.01	55.31%	77.26%	83.33.0%	85.75.0%	88.24%	89.51.0%
0.5/0.01	65.1%	91.57.0%	97.24.0%	98.19%	99.23%	99.47%
0.25/0.05	18.0%	32.12%	40.04%	43.67.0%	47.41.0%	49.79.0%
Misplacement Ratio (Single Membership)						
0.25/0.01	1.17%	0.1%	0.05%	0.026%	0.012%	0.018%
0.5/0.01	0.09%	0.0%	0.0%	0.0%	0.0%	0.0%
0.25/0.05	11.42%	6.77%	4.45%	3.7%	2.69%	2.36%
Hit Ratio (Multiple Memberships).						
0.25/0.01	*none*	*none*	*none*	*none*	*none*	*none*
0.5/0.01	*none*	*none*	*none*	*none*	*none*	*none*
0.25/0.05	*none*	*none*	*none*	*none*	*none*	*none*
Misplacement Ratio (Multiple Memberships)						
0.25/0.01	55.42%	24.89.0%	17.77%	15.1%	12.27%	10.95%
0.5/0.01	40.19%	8.43%	2.76%	1.8%	0.77%	0.53%
0.25/0.05	100.0%	87.39%	75.97%	72.03%	65.63%	62.15.0%

Since all series have single membership, the hit ratio for multiple memberships is left empty. However, the misplacement ratio for multiple memberships is still reported, as it is possible for single-membership series to be misclassified as multiple-membership series.

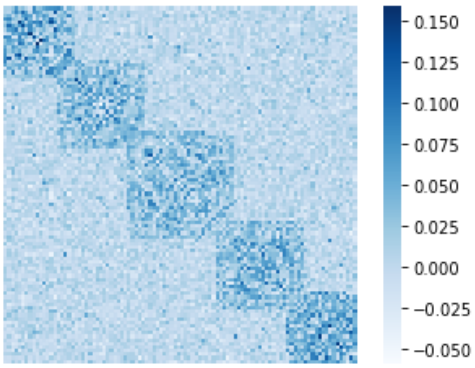
6.3 Multiple Membership Case

In this study, $\vec{\pi}$ are set up as follows:

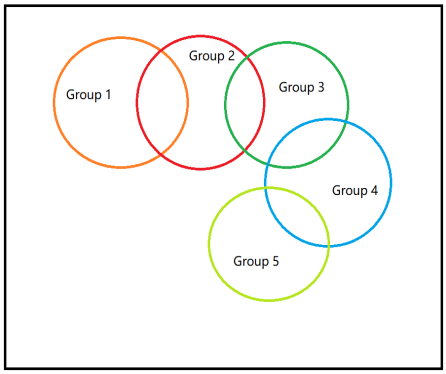
$\{\vec{\pi}_i\}_{i=1}^{20} = [1,0,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=21}^{25} = [0.5,0.5,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=26}^{35} = [0,1,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=36}^{40} = [0,0.5,0.5,0,0]^T$, $\{\vec{\pi}_i\}_{i=41}^{60} = [0,0,1,0,0]^T$, $\{\vec{\pi}_i\}_{i=61}^{65} = [0,0,0.5,0.5,0]^T$, $\{\vec{\pi}_i\}_{i=66}^{80} = [0,0,0,1,0]^T$, $\{\vec{\pi}_i\}_{i=81}^{85} = [0,0,0,0.5,0.5]^T$ and $\{\vec{\pi}_i\}_{i=86}^{100} = [0,0,0,0,1]^T$.

In this case, the dataset consists of 80 single-membership series and 20 multiple-membership

series. The left panel of Fig 5 shows the heatmap of the estimated symmetrized VAR coefficients, $\hat{\Phi}^S$ from the MMSB-VAR(1) with setup defined above. In the right panel of Fig 5, we observe that group 1 overlaps with group 2, which overlaps with group 3, which overlaps with group 4, and group 4, in turn, overlaps with group 5.



(a) Heatmap of adjacency matrix for estimated symmertized VAR matrices



(b) Illustration for overlapping structure

Figure 5: Panel (a) shows the heatmap of the estimated symmetrized VAR matrices the panel sampled from the MMSB-VAR(1). Panel (b) presents a Venn diagram illustrating the overlapping structure. In this study, we observe that group 1 overlaps with group 2, group 2 overlaps with group 3, group 3 overlaps with group 4, and group 4 overlaps with group 5.

The results obtained by applying the algorithm to the panel sampled from the MMSB-VAR(1) model under the setup defined above are summarized in Table 2. The algorithm performs well when the ratio $(b_{i,i}/b_{i,j})$ and the sample size T are sufficiently large, which is consistent with the theoretical analysis.

Table 2: **Accuracy Table for Second Simulation with 20 multiple-membership series and 80 single-membership series.**

	n=100					
b_{11},b_{22}/b_{12},b_{21}	2000	5000	8000	10000	15000	20000
Hit Ratio (Single Membership)						
0.25/0.01	33.52%	59.64%	71.94%	75.94%	80.51%	82.8%
0.5/0.01	44.35%	82.85%	94.09%	96.08%	98.36%	98.95%
0.25/0.05	15.84%	23.47%	29.3%	32.56%	36.28%	38.37%
Misplacement Ratio (Single Membership)						
0.25/0.01	12.29%	8.2%	5.61%	4.8%	3.86%	3.52%
0.5/0.01	7.06%	1.29%	0.33%	0.31%	0.07%	0.06%
0.25/0.05	17.43%	16.2%	14.48%	13.86%	13.18%	12.65%
Hit Ratio (Multiple Membership)						
0.25/0.01	78.71%	82.5%	85.08%	87.42%	88.8%	89.75%
0.5/0.01	86.2%	95.82%	98.66%	99.33%	99.74%	99.78%
0.25/0.05	73.07%	74.55%	76.94%	77.56%	78.02%	79.11%
Misplacement Ratio (Multiple Membership)						
0.25/0.01	100%	100%	100%	100%	86.2%	74.38%
0.5/0.01	100%	77.86%	26.44%	16.87%	7.37%	4.59%
0.25/0.05	100%	100%	100%	100%	100%	100%

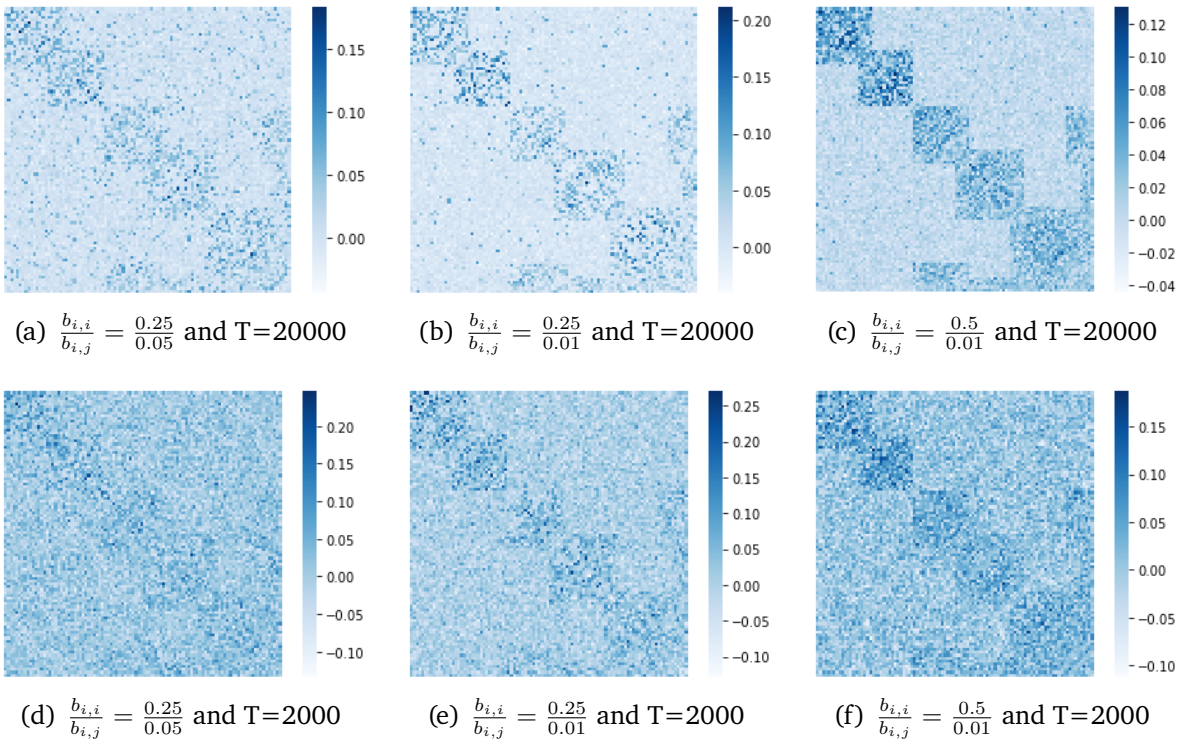
The table reports the accuracy of the algorithm.

6.4 Discussion of the Simulation Study

First, when all $\vec{\pi}_i$'s are degenerate, MMSB-VAR(p) simply reduces to SB-VAR(p) (Guðmundsson and Brownlees, 2021), Section 6.2 demonstrates that the two-step classification algorithm performs well under this circumstance. Furthermore, when some $\vec{\pi}_i$ vectors are non-degenerate, Section 6.3 shows that the algorithm also achieves effective results. Overall, we conclude that the proposed algorithm performs reliably on panels where all time series exhibit single memberships as well as on panels where some time series have multi-

ple memberships.

Second, the accuracy remains high when the in-group specific probability, $\{b_{i,j}\}_{i=j}$, is substantially larger than the between-group specific probability, $\{b_{i,j}\}_{i\neq j}$, as observed across all three rows. Accuracy decreases with smaller T, a lower ratio $b_{i,i}/b_{i,j}$. This finding aligns with the results shown in Fig.6.



(a) $\frac{b_{i,i}}{b_{i,j}} = \frac{0.25}{0.05}$ and T=20000 (b) $\frac{b_{i,i}}{b_{i,j}} = \frac{0.25}{0.01}$ and T=20000 (c) $\frac{b_{i,i}}{b_{i,j}} = \frac{0.5}{0.01}$ and T=20000

(d) $\frac{b_{i,i}}{b_{i,j}} = \frac{0.25}{0.05}$ and T=2000 (e) $\frac{b_{i,i}}{b_{i,j}} = \frac{0.25}{0.01}$ and T=2000 (f) $\frac{b_{i,i}}{b_{i,j}} = \frac{0.5}{0.01}$ and T=2000

Figure 6: The comparisons between heatmaps of adjacency matrices derived from the estimated symmetrized VAR coefficients under varying $b_{i,i}/b_{i,j}$ ratio and sample size T reveal a clear trend. Specifically, as the ratio $b_{i,i}/b_{i,j}$ increases and the sample size T becomes larger, the recovery of block structures in the graphs improves significantly.

7 Empirical Analysis

The algorithm presented in this work assumes that the number of communities K is known beforehand. However, in empirical applications, this assumption often does not hold. To address this limitation, we propose a heuristic method to estimate the number of communities directly from the data. The method is an extension of non-backtracking

method (Le and Levina, 2015; Li et al., 2022). A non-backtracking matrix is defined as $B_{nb} = \begin{bmatrix} 0 & D-I \\ -I & A \end{bmatrix}$ where A is the adjacency matrix, I is the identity matrix and D is a diagonal matrix with each diagonal entry representing the degree of the corresponding node in the graph. The number of communities is estimated as the number of eigenvalues satisfying the condition $|\lambda_i| > \|B_{nb}\|^{1/2}$. This criterion is based on the theoretical result that the largest k eigenvectors of B_{nb} will be well-separated from the radius $\|B_{nb}\|^{1/2}$.

We design a heuristic criterion for determining the number of communities based on the non-backtracking method. First, for a pre-specified value of k, we partition the time series into k groups by applying k-means algorithm to the symmetrized autoregressive matrix, $\hat{\Phi}^S$. Next, we reorder the time series based on the partitioning results and estimate a new symmetrized autoregressive matrix, $\hat{\Phi}^S$, from the re-ordered panel. Using a pre-specified threshold, γ_k, we construct a new binary adjacency matrix from the updated $\hat{\Phi}^S$ by setting all elements greater than γ_k to 1 and all other elements to 0. Finally, we construct $\hat{B}_{nb}$ from the binary adjacency matrix and estimate a $\hat{k}$ by applying the non-backtracking method to $\hat{B}_{nb}$. We then compare the estimated $\hat{k}$ with the initially chosen value of k used to construct $\hat{\Phi}^S$. If $\hat{k} = k$, we increment k by one and repeat the procedure described above. Starting from $k = 2$, the process continues until $\hat{k} \neq k$. We conclude that that the number of communities is $k-1$ representing the largest value of k such that $\hat{k} = k$. The details of this method can be found in the online Appendix B.3.

7.1 Public Opinion about Ukraine-Russia War

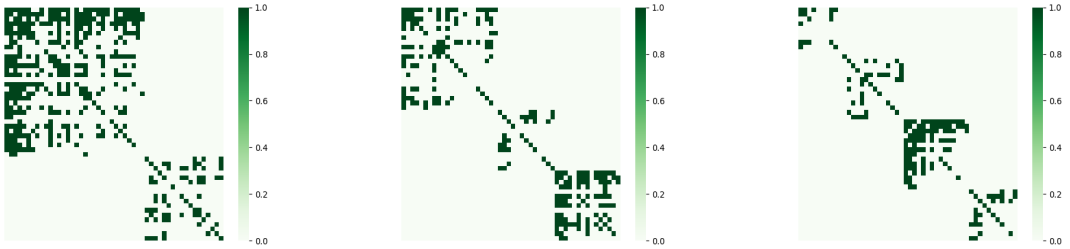
We study group structures in a panel of word counts constructed from a dataset of Reddit comments. Reddit is an American social news aggregation, content rating, and forum-based social network. The dataset consists of comments following posts related to the Russia-Ukraine war, providing a rich source of public discourse on the conflict. We obtain the dataset from Kaggle, a data science competition platform and online community for data

scientists and machine learning practitioners, operated by Google. Kaggle facilitates dataset sharing, model exploration, and collaborative development in a web-based data science environment. This dataset offers a valuable collection of public opinions and discussions from Reddit forums regarding the Russia-Ukraine war, comprising daily updated user comments that capture diverse perspectives, sentiments, and concerns surrounding the conflict.

The dataset is daily updated and consists of comments posted on Reddit since November 14, 2014. However, the number of comments was relatively low in the early stages and surged significantly following the outbreak of the full-scale war in 2022. The raw dataset contains over 4 million comments. For this study, we specifically use comments between July 2024 and September 2024, resulting in approximately 1.15 million comments after filtering. We preprocess the raw dataset using standard natural language processing (NLP) techniques, including the removal of HTML tags and special characters, tokenization, lowercasing, stopword removal, lemmatization, and the elimination of missing data. From this dataset, we construct a panel of word counts by selecting the top 50 words based on their frequency across the entire dataset. We then generate a time series of word counts for each of these 50 words, aggregated at 10-minute intervals from July 11, 2024, at 00:00:00 to September 23, 2024, at 23.59:59. This results in a panel with dimensions of $N = 50$ and sample size of $T = 10,796$.

We apply the two-step VAR algorithm to the constructed word count panel. The lag order of the VAR model is set to 1, and the number of groups, $\hat{k}$, is determined to be 3, as identified by the community number selection criterion described above. As shown in Fig.7, the largest k for which the set-up value equals the estimated $\hat{k}$ from the non-backtracking method is 3. The largest eigenvalue of the estimated VAR coefficient is 0.828. Fig.8 provides a summary of the algorithm's output. The left panel of Fig.8 hows heatmaps of the symmetrized autoregressive matrix of the VAR for the case of three communities, with the series reordered according to the estimated group partitions. The right panel of Fig.8 displays word clouds generated from the panel, where words are color-coded based

on their group partitions. The word cloud visualizes the 50 words in the panel, with font sizes proportional to their frequency counts across the entire dataset.

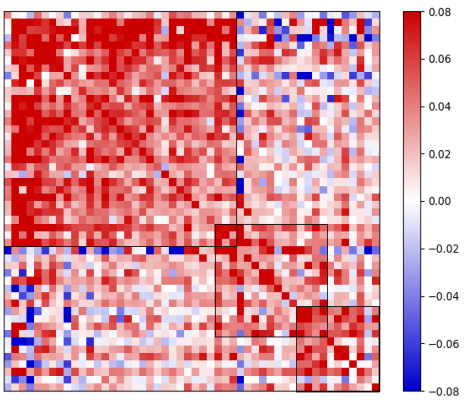


(a) Recovered adjacency for k=2. (b) Recovered adjacency for k=3. (c) Recovered adjacency for k=4.

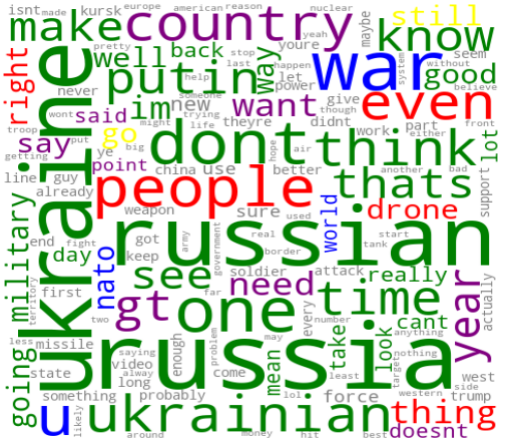
Figure 7: Recovered adjacency from panel for k=2, 3 and 4. $\hat{B}_{nb}$ is calculated based on three recovered adjacency. After applying non-backtracking method, we have that $\hat{k} = 2$ for $k = 2$, $\hat{k} = 3$ for $k = 3$ and $\hat{k} = 2$ for $k = 4$.

We describe the output of the procedure for $\hat{k} = 3$ and then comment on the results. The two-step VAR algorithm partitions the series in the panel into distinct groups characterized by weak spillover effects between them. The resulting three clusters consist of one large cluster and two medium-sized clusters, with minor overlaps between certain groups. Specifically, one overlap occurs between the largest cluster and one of the medium clusters, while another occurs between the two medium clusters. The largest non-overlapping cluster predominantly includes words associated with the two main sides in the war: Ukraine and Russia. Words in this cluster include Russia, Ukraine, Russian, Ukrainian, Putin, and military. The second non-overlapping cluster contains words referring to a third party that, while not directly involved in the conflict, plays a pivotal role. Words in this cluster include US, NATO, world, and war. The final non-overlapping cluster encompasses words that are militarily neutral. The words include right, people, and drone. Additionally, two overlapping clusters comprise neutral, commonly used words. The first overlapping cluster is shared between the third-party cluster and the militarily neutral cluster, including words such as say, country, said, point, year, want, and need. The second overlapping cluster is shared between the main conflict cluster and the militarily neutral cluster, including words

such as go, can't, and still.



(a) Heatmap of adjacency matrix for estimated symmertized VAR matrices in which the series are ordered by the estimated group partition.



(b) Word cloud plot. Green, blue and red refer to three non-overlapping clusters. Purple and yellow refer to two overlapping clusters. Grey are words in dataset except top 50 words used in study.

Figure 8: Reddit comment group clustering for k=3.

Words such as Ukraine, Russia, and Putin, associated with the two main parties in the war, are classified into a single cluster. This indicates that, on Reddit forums, users tend to leave comments that simultaneously reference Ukraine and Russia, with very few comments focusing exclusively on one side. Another cluster consists of words associated with key third parties in the war, including NATO, U.S., and war. This suggests that many users independently discuss the involvement of these third parties in their comments. It is unsurprising that neutral words such as go, say, country, and point are placed into overlapping clusters that belong to multiple groups, as these are common words with no specific connection to the Ukraine-Russia war. Overall, the two-step VAR classification method effectively partitions words associated with key participants in the Ukraine-Russia war and those related to the broader context of the conflict into distinct groups. This analysis demonstrates that words are naturally grouped around common thematic clusters, and these topics can be accurately recovered using our proposed group detection methodology.

8 Conclusion

In this work, we introduce a novel class of vector autoregressive models, the Mixed Membership Stochastic Block vector autoregression (MMSB-VAR(p)). This model is designed to capture multiple membership structures in panel data, offering a flexible framework for analyzing complex dependencies among time series. We propose a two-step algorithm to infer the latent group structure from observational data and analyze its theoretical properties. In particular, we establish the consistency of the proposed algorithm under the condition that the number of time observations (T) is sufficiently large, while the ratio of variables to observations (N/T) remains moderate.

References

Abbe, Emmanuel, Afonso S Bandeira and Georgina Hall. 2015. "Exact recovery in the stochastic block model." *IEEE Transactions on information theory* 62(1):471–487.

Ahn, Yong-Yeol, James P Bagrow and Sune Lehmann. 2010. "Link communities reveal multiscale complexity in networks." *nature* 466(7307):761–764.

Airoldi, Edo M, David Blei, Stephen Fienberg and Eric Xing. 2008. "Mixed membership stochastic blockmodels." *Advances in neural information processing systems* 21.

Ando, Tomohiro and Jushan Bai. 2016. "Panel data models with grouped factor structure under unknown group membership." *Journal of Applied Econometrics* 31(1):163–191.

Barigozzi, Matteo and Christian Brownlees. 2019. "Nets: Network estimation for time series." *Journal of Applied Econometrics* 34(3):347–364.

Bhatia, Rajendra. 2007. *Perturbation bounds for matrix eigenvalues*. SIAM.

Billio, Monica, Mila Getmansky, Andrew W Lo and Loriana Pelizzon. 2012. "Econometric measures of connectedness and systemic risk in the finance and insurance sectors." *Journal of financial economics* 104(3):535–559.

Blei, David M, Andrew Y Ng and Michael I Jordan. 2003. "Latent dirichlet allocation." *Journal of machine Learning research* 3(Jan):993–1022.

Bonhomme, Stéphane and Elena Manresa. 2015. "Grouped patterns of heterogeneity in panel data." *Econometrica* 83(3):1147–1184.

Brownlees, Christian, Guðmundur Stefán Guðmundsson and Gábor Lugosi. 2022. "Community detection in partial correlation network models." *Journal of Business & Economic Statistics* 40(1):216–226.

Chen, Cathy Yi-Hsuan, Wolfgang Karl Härdle and Yegor Klochkov. 2022. "Sonic: Social network analysis with influencers and communities." *Journal of Econometrics* 228(2):177–220.

Chen, Elynn Y, Jianqing Fan and Xuening Zhu. 2023. "Community network auto-regression for high-dimensional time series." *Journal of Econometrics* 235(2):1239–1256.

Demirer, Mert, Francis X Diebold, Laura Liu and Kamil Yilmaz. 2018. "Estimating global bank network connectedness." *Journal of Applied Econometrics* 33(1):1–15.

Diebold, Francis X and Kamil Yılmaz. 2014. "On the network topology of variance decompositions: Measuring the connectedness of financial firms." *Journal of econometrics* 182(1):119–134.

Erosheva, Elena A. 2003. "Bayesian estimation of the grade of membership model." *Bayesian statistics* 7:501–510.

Fan, Jianqing, Yuan Liao and Martina Mincheva. 2013. "Large covariance estimation by thresholding principal orthogonal complements." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 75(4):603–680.

Guðmundsson, Guðmundur Stefán and Christian Brownlees. 2021. "Detecting groups in large vector autoregressions." *Journal of Econometrics* 225(1):2–26.

Härdle, Wolfgang Karl, Weining Wang and Lining Yu. 2016. "Tenet: Tail-event driven network risk." *Journal of Econometrics* 192(2):499–513.

Holland, Paul W, Kathryn Blackmond Laskey and Samuel Leinhardt. 1983. "Stochastic blockmodels: First steps." *Social networks* 5(2):109–137.

Jin, Jiashun, Zheng Tracy Ke and Shengming Luo. 2024. "Mixed membership estimation for social networks." *Journal of Econometrics* 239(2):105369.

Joseph, Antony and Bin Yu. 2016. "Impact of regularization on spectral clustering.".

Karrer, Brian and Mark EJ Newman. 2011. "Stochastic blockmodels and community structure in networks." *Physical review E* 83(1):016107.

Koller, Daphne and Nir Friedman. 2009. *Probabilistic graphical models: principles and techniques*. MIT press.

Latouche, Pierre, Etienne Birmelé and Christophe Ambroise. 2009. "Overlapping stochastic block models." *arXiv preprint arXiv:0910.2098* .

Le, Can M and Elizaveta Levina. 2015. "Estimating the number of communities in networks by spectral methods." *arXiv preprint arXiv:1507.00827* .

Lei, Jing and Alessandro Rinaldo. 2015. "Consistency of spectral clustering in stochastic block models." *The Annals of Statistics* pp. 215–237.

Li, Tianxi, Lihua Lei, Sharmodeep Bhattacharyya, Koen Van den Berge, Purnamrita Sarkar, Peter J Bickel and Elizaveta Levina. 2022. "Hierarchical community detection by recursive partitioning." *Journal of the American Statistical Association* 117(538):951–968.

Merlevède, Florence, Magda Peligrad and Emmanuel Rio. 2011. "A Bernstein type inequality and moderate deviations for weakly dependent sequences." *Probability Theory and Related Fields* 151:435–474.

Ng, Andrew, Michael Jordan and Yair Weiss. 2001. "On spectral clustering: Analysis and an algorithm." *Advances in neural information processing systems* 14.

Palla, Gergely, Imre Derényi, Illés Farkas and Tamás Vicsek. 2005. "Uncovering the overlapping community structure of complex networks in nature and society." *nature* 435(7043):814–818.

Qin, Tai and Karl Rohe. 2013. "Regularized spectral clustering under the degree-corrected stochastic blockmodel." *Advances in neural information processing systems* 26.

Rohe, Karl, Sourav Chatterjee and Bin Yu. 2011. "Spectral clustering and the high-dimensional stochastic blockmodel.".

Rohe, Karl and Tai Qin. 2013. "The blessing of transitivity in sparse and stochastic net-

works." *arXiv preprint arXiv:1307.2302* .

Sarkar, Purnamrita and Peter J Bickel. 2015. "Role of normalization in spectral clustering for stochastic blockmodels.".

Su, Liangjun, Zhentao Shi and Peter CB Phillips. 2016. "Identifying latent structures in panel data." *Econometrica* 84(6):2215–2264.

Von Luxburg, Ulrike. 2007. "A tutorial on spectral clustering." *Statistics and computing* 17:395–416.

Yang, Jaewon and Jure Leskovec. 2013. Overlapping community detection at scale: a non-negative matrix factorization approach. In *Proceedings of the sixth ACM international conference on Web search and data mining*. pp. 587–596.

Zhang, Yingying, Huixia Judy Wang and Zhongyi Zhu. 2019. "Quantile-regression-based clustering for panel data." *Journal of Econometrics* 213(1):54–67.

Zhu, Xuening, Rui Pan, Guodong Li, Yuewen Liu and Hansheng Wang. 2017. "Network vector autoregression.".

Zhu, Xuening, Weining Wang, Hansheng Wang and Wolfgang Karl Härdle. 2019. "Network quantile autoregression." *Journal of econometrics* 212(1):345–358.

Appendix A. Proof of Results

We begin by introducing additional notation. We define the population adjacency matrix as $\mathbb{E}[A_l] = \mathcal{A}$ for $l = 1, \ldots, p$ and the population normalized matrix $\mathcal{D}$, where $[\mathcal{D}]_{ii} = \sum_{j=1}^{n} \Big([\mathcal{A}]_{ij} + [\mathcal{A}]_{ji} \Big)$. The population autoregressive coefficient is defined as $\Phi_l = \phi_l \mathcal{D}^{-1/2} \mathcal{A} \mathcal{D}^{-1/2}$ and the symmertised autoregressive coefficient is given by $\Phi^S = \sum_{l=1}^{p} \Phi_l + \Phi_l'$

Proof of Lemma 1. Assuming that $\mathbb{P}\big(\boldsymbol{\pi}|\vec{\alpha}\big) = 1$, we take $\boldsymbol{\pi}$ as given and have $\mathbb{E}(A_l) = \mathcal{A} \approx \boldsymbol{\pi} B \boldsymbol{\pi}'$. Then,

$$\begin{aligned}
\Phi^S &= \sum_{l=1}^{p} \Big(\phi_l \Big) \mathcal{D}^{-1/2} \Big(\mathcal{A} + \mathcal{A}' \Big) \mathcal{D}^{-1/2} \\
&= 2p \Big(\boldsymbol{\pi} B_L \boldsymbol{\pi}' \Big) \sum_{l=1}^{p} \Big(\phi_l \Big) \; where \; B_L = \mathcal{D}^{-1/2} B \mathcal{D}^{-1/2}.
\end{aligned}$$

We now focus on $\boldsymbol{\pi} B_L \boldsymbol{\pi}'$. Define $\Delta = diag\Big(\sqrt{S_1}, \cdots, \sqrt{S_k} \Big)$ where S_i for $i = 1, \ldots, k$ represents the sum of the stickiness of all nodes belongs to k groups. Then,

$$\begin{aligned}
\Phi^S &\approx \boldsymbol{\pi} \Delta^{-1} \Delta B_L \Delta \Delta^{-1} \boldsymbol{\pi}' \\
&= \big(\boldsymbol{\pi} \Delta^{-1} \big) \Delta B_L \Delta \big(\boldsymbol{\pi} \Delta^{-1} \big)'
\end{aligned}$$

It's straightforward to see that $\boldsymbol{\pi} \Delta^{-1}$ is orthonormal if all π_i in $\boldsymbol{\pi}$ are standard unit vectors. Let $U_L \Lambda U_L'$ be the egien-decomposition of $\Delta B_L \Delta'$, such that $U_L \Lambda U_L' = \Delta B_L \Delta'$, which implies that

$$\begin{aligned}
\Phi^S &\approx \big(\boldsymbol{\pi} \Delta^{-1} \big) \Delta B_L \Delta \big(\boldsymbol{\pi} \Delta^{-1} \big)' \\
&= \big(\boldsymbol{\pi} \Delta^{-1} \big) U_L \Lambda U_L' \big(\boldsymbol{\pi} \Delta^{-1} \big)' \\
&= \big(\boldsymbol{\pi} \Delta^{-1} U_L \big) \Lambda \big(\boldsymbol{\pi} \Delta^{-1} U_L \big)'
\end{aligned}$$

Defining the eigen-decomposition of Φ^S as $U \Lambda U'$ such that $\Phi^S = U \Lambda U'$, it's straightforward to have $U \approx \boldsymbol{\pi} \big(\Delta^{-1} U_L \big)$ which implies that U is a quasi-rotation of $\boldsymbol{\pi}$ with the quasi-rotation matrix given by $\Delta^{-1} U_L$. This establishes the first statement.

Next, let $R = \Delta^{-1} U_L$, since

$det(R) = det(\Delta^{-1}) det(U_L) > 1$,

It follows that R^{-1} exists, thus proving the second statement. $\square$

Proof of Lemma 2. We aim to bound $\|\hat{\Phi}^S - \Phi^S\|$ where $\hat{\Phi}^S$ and Φ^S represent the estimated and population symmertised VAR coefficients respectively. Following the strategy outlined in Guðmundsson and Brownlees (2021), we establish the following inequality:

$$\|\hat{\Phi}^S - \overline{\Phi}^S\| \leq \|\hat{\Phi}^S - \Phi^S\| + \|\overline{\Phi}^S - \Phi^S\|,$$

where $\Phi^S = \sum_{l=1}^{p} \left(\Phi_l + \Phi_l'\right)$ represents the sample symmertised VAR coefficient. We prove that $\|\hat{\Phi}^S - \overline{\Phi}^S\|$ is bounded by applying Lemma A.1 and Lemma A.2, which are defined below. □

Lemma A.1. *Given the MMSB-VAR(p) that satisfy Assumption 2, where the pairwise probability matrix B satisfies Assumption 1, let $\overline{\Phi}^S$ and Φ^S denote the sample and population symmertised autoregressive matrices. Then, the following conclusion is derived:*

$$\|\overline{\Phi}^S - \Phi^S\| \leq O_p\left(\sqrt{\frac{\alpha_0 log(N)}{\underline{\alpha} N B_N}}\right).$$

Proof of Lemma A.1. I prove this by applying the conclusion of Lemma A.1 from Guðmundsson and Brownlees (2021) which states:

Let $\overline{d}_{min} = \min_i d_i$ denote the minimum expected degree of the graph and define $v = 2\max_{ij} \overline{w}_{ij} + \max_{ij}(\overline{w}_{ij}^2/\underline{w}_{ij}^2)$, where $\overline{w}$ and $\underline{w}$ represent the lower and upper bounds for sampling the edge weights respectively. Then, for a constant $c > 0$, there exists another constant $C > 0$ independent of number of nodes, N, and the edge probabilities. Moreover, if $\overline{d}_{\min} \geq Clog(N)$ and for all $n^{-c} \leq \delta \leq 1/2,$ we have:

$$\mathbb{P}\Big(\|L - \mathcal{L}\| \leq 14\sqrt{\frac{vlog(2N/\delta)}{\overline{d}_{min}}}\Big) \geq 1 - \delta.$$

Our goal is to prove that the current model setup satisfies the conditions required for Lemma A.1 to hold. Since

$$\|\overline{\Phi}^S - \Phi^S\| = \sum_{l=1}^{p} \phi_l \|L_l^S - \mathcal{L}_l^S\|.$$

Note that $\mathcal{A}_l = \mathbb{E}(\phi) diag(\vec{\alpha}/\alpha_0)\{\vec{\pi}_i\}_{i=1}^N (B + B^T)\big(\{\vec{\pi}_i\}_{i=1}^N\}\big)^T diag(\vec{\alpha}/\alpha_0)$, then, we have

$$\big[D_l^S\big]_{ii} = \sum_{j=1}^{N} \big[\mathcal{A}_l^S\big]_{ij} = \mathbb{E}(\phi)\vec{\pi}_i diag(\vec{\alpha}/\alpha_0) \sum_{j=1}^{N} diag(\vec{\alpha}/\alpha_0)(B + B^T)\vec{\pi}_j^T,$$

then, by Assumption 2, define $\underline{\alpha} = \min_i \alpha_i$, the minimum degree should be

$$d_{min} = \min_i \big[d_l^S\big]_{ii} = \Omega\big((\underline{\alpha}/\alpha_0)^2 N B_N\big) = \Omega\big((\underline{\alpha}/\alpha_0)^2 log(N)\big)$$

Hence, if we pick a $\delta = N^{-c}$ such that $N^{-c} \leq 1/2$, we have

$$\mathbb{P}\Big(\|L_l^S - \mathcal{L}_l^S\| \geq C\sqrt{\frac{\alpha_0^2 log(N)}{\underline{\alpha}^2 N B_N}}\Big) \leq N^{-c},$$

for c and N sufficiently large. $\square$

Lemma A.2. *Given the MMSB-VAR(p) model as defined in Definition 2 and satisfying Assumption 3, let $\hat{\Phi}^S$ denote the estimated symmertised autoregressive matrix and Φ^S the sample symmertised autoregressive matrix. Then, with high probability, we have $\|\hat{\Phi}^S - \Phi^S\| \leq O_p\left(\frac{N^{3/2}}{\sqrt{T}}\lambda_{max}(\Phi^S)\right)$.*

Proof of Lemma A.2. I prove the Lemma by applying the Lemma A.2 from (Guðmundsson and Brownlees, 2021), which states that, given a $T \times N$ row de-meaned matrix Y satisfying (iii), (iv) and (v) of Assumption 3, let Σ_Y and $\hat{\Sigma}_Y = \frac{1}{T}Y'Y$ denote the population covariance and sample covariance of Y respectively.

If T=$\Omega(n^{2/\gamma-1})$, then

$$\mathbb{P}\left(||\hat{\Sigma}_Y - \Sigma_Y|| \leq C\sqrt{\frac{N}{T}}||\Sigma_Y||\right) \geq 1 - 3e^{-N}$$

and

$$\sqrt{\lambda_1(\Sigma_Y)}\left(1 - C\sqrt{\frac{N}{T}}\right) \leq \sqrt{\lambda_1(\hat{\Sigma}_Y)} \leq \sqrt{\lambda_N(\hat{\Sigma}_Y)} \leq \sqrt{\lambda_N(\Sigma_Y)}\left(1 + C\sqrt{\frac{N}{T}}\right),$$

with probability at least $1 - 3e^{-N}$.

Given the Lemma above, the first step is to prove that the OLS estimator exists and converges. Let Y be a $T \times N$ row de-meaned matrix. Define the sample covariance as $\hat{\Sigma}_X = \frac{1}{T}\sum_{t=p}^{T} X_t X_t'$, the p-th order autocovariances as $\hat{\Sigma}_{XY} = \frac{1}{T}\sum_{t=p}^{T} X_t Y_t'$ and the OLS estimator as $\hat{\Phi} = \hat{\Sigma}_X^{-1}\hat{\Sigma}_{XY}'$, provided that the inverse exists.

By applying the lemma A.2 from (Guðmundsson and Brownlees, 2021) to $\hat{\Sigma}_X$, we have

$$\sqrt{\lambda_1(\Sigma_X)}\left(1 - C\sqrt{\frac{N}{T}}\right) \leq \sqrt{\lambda_1(\hat{\Sigma}_X)},$$

$$\lambda_1(\Sigma_X)\left(1 - C\sqrt{\frac{N}{T}}\right)^2 \leq \lambda_1(\hat{\Sigma}_X),$$

$$\lambda_1(\Sigma_X) - O_p\left(\sqrt{\frac{N}{T}}\right) \leq \lambda_1(\hat{\Sigma}_X).$$

where $\lambda_1(\Sigma_X)$ and $\lambda_1(\hat{\Sigma}_X)$ denote the smallest eigenvalues of Σ_X and $\hat{\Sigma}_X$ respectively. Then, with high probability, the OLS estimator $\hat{\Phi}$ exists when T is sufficiently large relative to N.

Next, We need to bound $\|\hat{\Phi} - \Phi\|$ and $\|\hat{\Phi}^S - \Phi^S\|$.

By applying the Lemma SM.1 in our supplementary material,

$$||\hat{\Phi} - \Phi|| = ||\hat{\Sigma}_X^{-1}\hat{\Sigma}'_{YX} - \Sigma_X^{-1}\Sigma'_{YX}|| \leq ||\hat{\Sigma}_X^{-1}||\,||\hat{\Sigma}_{YX} - \Sigma_{YX}|| + ||\hat{\Sigma}_{YX}||\,||\hat{\Sigma}_X^{-1} - \Sigma_X^{-1}||. \quad (10)$$

Then, I firstly focus on $||\hat{\Sigma}_{YX} - \Sigma_{YX}||$ and $||\hat{\Sigma}_X^{-1} - \Sigma_X^{-1}||$ and come to $||\hat{\Sigma}_X^{-1}||$ and $\|\hat{\Sigma}_{YX}\|$ later.

By applying the Lemma A.2 in (Guðmundsson and Brownlees, 2021), we have

$$||\hat{\Sigma}_{YX} - \Sigma_{YX}|| = O_p\Big(\sqrt{\frac{N}{T}}||\Sigma_{YX}||\Big) \quad (11)$$

and

$$||\hat{\Sigma}_X - \Sigma_X|| = O_p\Big(\sqrt{\frac{N}{T}}||\Sigma_X||\Big).$$

By applying Lemma OA.8 in (Guðmundsson and Brownlees, 2021), we have

$$||\hat{\Sigma}_X^{-1} - \Sigma_X^{-1}|| \leq ||\hat{\Sigma}_X^{-1}||\,||\hat{\Sigma}_X - \Sigma_X||\,||\Sigma_X^{-1}|| = O_p\Big(||\Sigma_X^{-1}||\sqrt{\frac{N}{T}}||\Sigma_X||\Big). \quad (12)$$

Combing (10), (11) and (12), we have

$$\begin{aligned} ||\hat{\Phi} - \Phi|| &= O_p\Big(||\Sigma_X^{-1}||\sqrt{\frac{N}{T}}||\Sigma_X|| + ||\Sigma_X^{-1}||\,||\Sigma_{YX}||\sqrt{\frac{N}{T}}||\Sigma_X||\Big) \\ &= O_p\Big(||\Sigma_X^{-1}||\sqrt{\frac{N}{T}}||\Sigma_{YX}||\,||\Sigma_X||\Big). \end{aligned} \quad (13)$$

When the concentration matrix is invertible, we have $||\Sigma_{YX}|| = ||\Sigma_X\Phi|| \leq ||\Sigma_X||\,||\Phi||$, and Lemma OA.13 in (Guðmundsson and Brownlees, 2021) shows that $||\Sigma_X^{-1}||$ and $||\Sigma_X||$ are bounded by absolute constants. Then, (13) becomes

$$||\hat{\Phi} - \Phi|| = O_p\Big(\sqrt{\frac{N}{T}}||\Phi||\Big).$$

From Lemma SM.2 in my supplementary material, we have $||\Phi_l|| \leq \phi_l N\lambda_{max}(\Phi_l)$ for

$l = 1, \ldots, p$ and from Lemma OA.11 in (Guðmundsson and Brownlees, 2021), we have $||\Phi|| \leq \sum_{l=1}^{p} ||\Phi_l||$.

Furthermore, $||\hat{\Phi}^S - \Phi^S|| \leq \sum_{l=1}^{p} ||\hat{\Phi}_l^S - \Phi_l^S|| \leq \sum_{l=1}^{p} ||\hat{\Phi}_l - \Phi_l||$ by triangle inequality.

Finally, Lemma OA.11 in (Guðmundsson and Brownlees, 2021) gives that $||\hat{\Phi}_l - \Phi_l|| \leq ||\hat{\Phi} - \Phi||$ for all $l = 1, \ldots, p$. □

Proof of Lemma 3. This is a straightforward extension of Lemma 4 in (Guðmundsson and Brownlees, 2021), closely following the approach in (Sarkar and Bickel, 2015), while utilizing the result from (Rohe, Chatterjee and Yu, 2011), the Davis-Kahan theorem and the fact that $\lambda_{max}(\Phi^S) \leq 2$. □

Proof of Lemma 4. This is a straightforward extension of the proof of Lemma 3.2 in (Rohe, Chatterjee and Yu, 2011) to the row-normalized matrix as well as the proof of Theorem 4.4 in (Rohe and Qin, 2013). □

Proof of Theorem 1. This is a straightforward extension of the proof of Theorem 1 in (Guðmundsson and Brownlees, 2021), closely following the approach in (Rohe, Chatterjee and Yu, 2011).

Please note that implementing the k-means algorithm is equivalent to minimizing the following objective function: $\min_{\{m_1,\ldots,m_k\} \in \mathcal{R}^k} \sum_{i=1}^{N} \min_s \|x_i - m_s\|_2^2$.

I define the estimated and population centroids as follows:

$\hat{\mathcal{C}}_i = \underset{m_i \in M}{\arg\min} ||\hat{x}_i - m_i||^2$ and $\mathcal{C}_i = \underset{m_i \in M}{\arg\min} ||\mathbf{X}_i\mathcal{O} - m_i||^2$ where $\mathcal{O}$ represents a $k \times k$ orthonormal rotation matrix as defined in Lemma 3 and $M = \{m_1, ..., m_k\}$.

Let $\mathcal{M}$ denote the set of misclustered series, $\hat{\mathcal{C}}$ the estimated centroid and $\mathcal{C}$ the population centroid, then

$\mathcal{M} = \left\{i : ||\hat{\mathcal{C}}_i - \mathcal{C}_i|| \geq \sqrt{1/2}\right\}$ and $\|\hat{X} - \hat{\mathcal{C}}|| \leq ||\hat{X} - \mathcal{C}||$.

It establishes that if the difference between the population and estimated centroids of series i exceeds $\sqrt{1/2}$, then this series is considered misclustered. Consequently, $|\mathcal{M}|$ is defined as the total number of misclustered series. Bounding $\frac{|\mathcal{M}|}{N}$ is equivalent to bounding the proportion of misclustered series. If this ratio is sufficiently small, the consistency of the first step is established.

Let $\mathcal{M}$ be the set of misclustered series. We address the problem by adapting it the non-overlapping case, since the misclustered ratio in the overlapping case is always bounded above by that of the non-overlapping case. We define $\mathcal{M}^*$ as the set of misclustered series in the non-overlapping case, where we designate one of the groups that a multiple-membership node belongs to as the true group and the others as incorrect ones. Then, it is evident that

(1) $\frac{|\mathcal{M}|}{N} = \frac{|\mathcal{M}^*|}{N}$ if all variables belong to one group exactly.
(2) $\frac{|\mathcal{M}|}{N} \leq \frac{|\mathcal{M}^*|}{N}$ if at least one variable belong to more than one group.
It is evident that $|\mathcal{M}|$ is bounded above by $|\mathcal{M}^*|$, we focus on $|\mathcal{M}^*|$.
By triangle inequality, $||\hat{\mathcal{C}} - \mathcal{C}|| \leq ||\hat{\mathcal{C}} - \hat{X}|| + ||\mathcal{C} - \hat{X}|| \leq 2||\mathcal{C} - \hat{X}||$.
By Lemma 2 and 3, we bound $|\mathcal{M}^*|$ as follows:
$|\mathcal{M}^*| = \sum_{i \in \mathcal{M}^*} 1 \leq 2 \sum_{i \in \mathcal{M}^*} ||\hat{\mathcal{C}}_i - \mathcal{C}_i||^2 \leq 2||\hat{\mathcal{C}} - \mathcal{C}||^2 \leq 8||\mathcal{C} - \hat{X}||^2$.
And since $||\mathcal{C} - \hat{X}|| = O_p\left(\sqrt{\frac{\alpha_0 log(N)}{\underline{\alpha} B_N}} + \frac{N^2}{\sqrt{T}}\right)$, then we have $|\mathcal{M}^*| = O_p\left(\frac{\alpha_0 log(N)}{\underline{\alpha} B_N} + \frac{N^4}{T}\right)$,
which implies that $\frac{|\mathcal{M}|}{N} \leq \frac{|\mathcal{M}^*|}{N} = O_p\left(\frac{\alpha_0}{\underline{\alpha}} \frac{log(N)}{N B_N} + \frac{N^3}{T}\right)$. $\square$

Proof of Lemma 4. Given reconstructed sub-matrix A_{group} of $\mathbb{E}(\Phi^S)$, obtained from the partition results in step 1 and satisfying Assumption 1,2,3 and 4, let λ and $\vec{v}$ be any eigenvalue and eigenvector of A_{group}. then, we have $A_{group}\vec{v} = \lambda\vec{v}$ and for specific vertex j,

$$\sum_{i=j} a_{ij} v_i = \lambda v_j,$$

where v_i for $i = 1, \ldots, N$ is the element of eigenvector $\vec{v}$ corresponding to eigenvalue λ of A. Then,

$$\begin{aligned}
\lambda &= \frac{\sum_{i=j} a_{ij} v_i}{v_j} \\
&= \sum_{i=j} a_{ij} \frac{v_i}{v_j} \\
&\leq \sum_{i=j} a_{ij} \frac{v_j}{v_j} \; where\; v_{majx}\; is\; the\; maximum\; in\; \vec{v}\; without\; loss\; of\; generality \\
&= d_j,\; where\; d_j = \sum_{i=j} a_{ij}\; is\; the\; largest\; degree. \\
&\leq d_{max}\; where\; d_{max}\; is\; the\; largest\; degree.
\end{aligned}$$

On the other hand, since all elements in A_{group} are identical and non-negative under the given assumptions, we have $rank(A_{group}) = 1$, meaning there's only one non-zero eigenvalue. Furthermore, because all rows in A_{group} are identical, the degrees of all variables in A_{group} are equal, implying that $d_i = d_j$ for $i \neq j$.

Furthermore, for A_{group}, we have $\sum \lambda_i = trace(a_{ii})$ which implies that the only non-zero eigenvalue is positive.

Overall, we obtain the bound $0 \leq \lambda \leq d_i$ for $i = 1, \ldots, \tilde{N}$ where $\tilde{N}$ denotes the number of nodes in a specific group. $\square$

Proof of Lemma 5. We aim to prove that $|\hat{\lambda}^*_{group} - \overline{\lambda}_{group}|$ is bounded, where $\hat{\lambda}^*_{group}$ and $\overline{\lambda}_{group}$ represent the eigenvalues of reconstructed adjacency of $\hat{\Phi}^S$ and $\mathbb{E}(\Phi^S)$ respectively according to the partition results in step 1. To achieve this, We prove that $|\hat{\lambda}_i - \overline{\lambda}_i|$ is bounded above, where $\hat{\lambda}_i$ and $\overline{\lambda}_i$ represent the eigenvalues of $\hat{\Phi}^S$ and $\mathbb{E}(\Phi^S)$ respectively. Since, by Lemma SM.4, any sub-matrix is always bounded above by its original matrix, proving the bound for the entire adjacency matrix applies to all sub-matrix.

We apply Weyl's inequality from Theorem 2.1 in Bhatia (2007), which states the following:

Given the estimated symmertised autoregressive coefficient $\hat{\Phi}^S$ and its population counterpart $\overline{\Phi}^S$, let $\hat{\lambda}_i$ and $\overline{\lambda}_i$ denote the eigenvalues of $\hat{\Phi}^S\hat{\Phi}^S$ and $\overline{\Phi}^S\overline{\Phi}^S$ respectively. Assuming the eigenvalues are ordered as $\hat{\lambda}_1 \leq \cdots \leq \hat{\lambda}_N$ and $\overline{\lambda}_1 \leq \cdots \leq \overline{\lambda}_N$, we obtain the bound:

$$\max_i |\hat{\lambda}_i - \overline{\lambda}_i| \leq \|\hat{\Phi}^S\hat{\Phi}^S - \overline{\Phi}^S\overline{\Phi}^S\|$$

We sequentially examine both the left-hand and right-hand sides of the inequality.

First, we examine $\max_i |\hat{\lambda}_i - \overline{\lambda}_i|$. Let $\hat{\lambda}^*_i$ and $\overline{\lambda}^*_i$ denote the estimated and population eigenvalues of $\hat{\Phi}^S$ and $\overline{\Phi}^S$. We have $(\overline{\lambda}^*_i)^2 = \overline{\lambda}_i$ and $(\hat{\lambda}^*_i)^2 = \hat{\lambda}_i$ since, for any symmetric matrix A with eigenvalue λ and eigenvector v, it holds that $A^2v = \lambda^2 v$.

$$\begin{aligned}
\max_i |\hat{\lambda}_i - \overline{\lambda}_i| &\geq |\hat{\lambda}_i - \overline{\lambda}_i| \\
&= |(\hat{\lambda}^*_i)^2 - (\overline{\lambda}^*_i)^2| \\
&= |(\hat{\lambda}^*_i - \overline{\lambda}^*_i)^2 - 2\overline{\lambda}^*_i + 2\hat{\lambda}^*_i\overline{\lambda}^*_i| \\
&= |(\hat{\lambda}^*_i - \overline{\lambda}^*_i)^2 - 2\overline{\lambda}^*_i(1 - \hat{\lambda}^*_i)| \\
&\geq |\hat{\lambda}^*_i - \overline{\lambda}^*_i|^2 - 2|\overline{\lambda}^*_i(1 - \hat{\lambda}^*_i)| \; by \; triangle \; inequality \\
&\geq |\hat{\lambda}^*_i - \overline{\lambda}^*_i|^2 - 4 \; since \; \hat{\lambda}^*_i, \overline{\lambda}^*_i \in [-1, 1].
\end{aligned}$$

It should be noted that $\hat{\lambda}^*_i, \overline{\lambda}^*_i \in [-1, 1]$ come from Lemma SM.3 in supplementary material.

Next, we examine $\|\hat{\Phi}^S\hat{\Phi}^S - \overline{\Phi}^S\overline{\Phi}^S\|$.

By Lemma SM.1 in supplementary material, we have

$$\|\hat{\Phi}^S\hat{\Phi}^S - \overline{\Phi}^S\overline{\Phi}^S\| \leq \|\overline{\Phi}^S\|\|\overline{\Phi}^S - \hat{\Phi}^S\| + \|\hat{\Phi}^S\|\|\hat{\Phi}^S - \overline{\Phi}^S\|$$

We have $\|\hat{\Phi}^S - \overline{\Phi}^S\| \leq O_p\left(\sqrt{\frac{\alpha_0 log(N)}{\underline{\alpha} N B_N}} + \frac{N^{3/2}}{\sqrt{T}}\lambda_{max}(\Phi^S)\right)$ by Lemma 2 in this work.

The matrix $\overline{\Phi}^S$ is always bounded above by a constant, since $\overline{\Phi}^S = \mathbb{E}(\Phi^S) \approx \boldsymbol{\pi} B \boldsymbol{\pi}'$ and it is straightforward to see that $\overline{\Phi}^S$ is bounded above by the norm of a matrix in which all elements are equal to the largest element in B.

As for $\hat{\Phi}^S$, we have $\hat{\Phi} = \hat{\Sigma}_X^{-1}\hat{\Sigma}'_{XY}$ by the property of OLS. Furthermore, $\hat{\Sigma}_X^{-1}$ and $\hat{\Sigma}'_{XY}$ are close to Σ_X^{-1} and Σ'_{XY} when T is sufficiently large. Furthermore, if concentration matrix is invertible, then $\Sigma'_{XY} = \Sigma_X \Phi$ which implies that $\hat{\Phi} = \Phi$ when T is sufficiently large. By Lemma OA.12 in
Guðmundsson and Brownlees (2021), $\|\Phi_l\| \leq \phi_l$ where ϕ_l is an auxiliary scalar that ensures the stability of the MMSB-VAR(p). Additionally, by Lemma OA.11 in Guðmundsson and Brownlees (2021), we have $\|\Phi\| \leq \sum_{l=1}^{p} \|\Phi_l\| \leq \sum_{l=1}^{p} \phi_l$ implying that $\hat{\Phi}^S$ is also bounded above.

Finally, we combine two parts and have

$$|\hat{\lambda}_i^* - \overline{\lambda}_i^*|^2 \leq \max_i |\hat{\lambda}_i - \overline{\lambda}_i| \leq \|\hat{\Phi}^S \hat{\Phi}^S - \overline{\Phi}^S \overline{\Phi}^S\| \leq O_p\left(\sqrt{\frac{\alpha_0 log(N)}{\underline{\alpha} N B_N}} + \frac{N^{3/2}}{\sqrt{T}} \lambda_{max}(\Phi^S)\right),$$

which implies that

$$\sum_{i=1}^{N} |\hat{\lambda}_i^* - \overline{\lambda}_i^*| \leq O_p\left(\sqrt{\sqrt{\frac{\alpha_0 log(N)}{\underline{\alpha} N B_N}} + \frac{N^{3/2}}{\sqrt{T}} \lambda_{max}(\Phi^S)}\right),$$

and, therefore,

$$\sum_{l=1}^{p}\sum_{i=1}^{N} |\hat{\lambda}_{i,l}^* - \overline{\lambda}_{i,l}^*| \leq O_p\left(\sqrt{\sqrt{\frac{\alpha_0 log(N)}{\underline{\alpha} N B_N}} + \frac{N^{3/2}}{\sqrt{T}} \lambda_{max}(\Phi^S)}\right). \quad \square$$

Proof of Theorem 2. It is similar to the proof in Theorem 1. Let $\mathcal{N}$ be the set of mis-clustered variables, $\hat{\boldsymbol{\lambda}}$ the estimated eigenvalue vector in which the elements are $\hat{\lambda}_i$ for $i = 1, \ldots, N$ and $\overline{\boldsymbol{\lambda}}$ the population eigenvalue vector in which the elements are $\overline{\lambda}_i$ for $i = 1, \ldots, N$, if we define $\mathcal{N}$ as $\mathcal{N} = \left\{j : \|\hat{\lambda} - \overline{\lambda}\|^2 \geq \sqrt{\sqrt{\frac{\alpha_0 log(N)}{\underline{\alpha} N B_N}} + \frac{N^{3/2}}{\sqrt{T}}}\right\}$. Then, we can bound $\mathcal{N}$ as follows:

$|\mathcal{N}| = \sum_{j \in \mathcal{N}} 1 \leq \sum_{j \in \mathcal{N}} \|\hat{\lambda} - \overline{\lambda}\|^2 \leq \sqrt{\sqrt{\frac{\alpha_0 log(N)}{\underline{\alpha} N B_N}} + \frac{N^{3/2}}{\sqrt{T}}}$.

Then,

$\frac{|\mathcal{N}|}{N_{total} - N} \leq \frac{|\mathcal{N}|}{cN - N} \leq \frac{1}{c-1}\sqrt{\sqrt{\frac{\alpha_0 log(N)}{\underline{\alpha} N^5 B_N}} + \frac{N^{1/2}}{\sqrt{T}}}$. $\square$

Appendix B. Supplementary Material

Appendix B.1 Supplementary Proof

Lemma SM.1. *Let A_1 and A_2 be $n \times n$ matrices and B_1 and B_2 $n \times n$ matrices, then* $||A_2B_2 - A_1B_1|| \leq ||A_1||\ ||B_1 - B_2|| + ||B_2||\ ||A_2 - A_1||$.

Proof of Lemma SM.1. $||A_2B_2 - A_1B_1|| = ||A_1(B_1 - B_2) + B_2(A_2 - A_1)||$
$\leq ||A_1(B_1 - B_2)|| + ||B_2(A_2 - A_1)||$, by triangle inequality
$\leq ||A_1||\ ||(B_1 - B_2)|| + ||B_2||\ ||(A_2 - A_1)||$, by sub-multiplicativity of norms. □

Lemma SM.2. *The spectral norm of Φ_l from the MMSB-VAR(p) satisfies* $||\Phi_l|| \leq ||\Phi_l + \Phi_l^{'}|| \leq \phi_l N \lambda_{max}(\Phi_l^S)$.

Proof of Lemma SM.2. $||\Phi_l|| \leq ||\Phi_l + \Phi_l^{'}|| = \phi_l||D_l^{-1/2}(A_l^{'} + A_l)D_l^{-1/2}|| \leq \phi_l||U\Lambda U^{'}|| \leq \phi_l||\mathbf{1}\Lambda\mathbf{1}^{'}|| \leq \phi_l N \lambda_{max}(\Phi_l^S)$. □

Lemma SM.3. *Given a symmertised autoregressive coefficient Φ^S of the MMSB-VAR(p) in Definition 2 and define its Laplacian as $L^S = I - \Phi^S$. Furthermore, define $\lambda_1 \leq \cdots \leq \lambda_N$ and $\alpha_1 \leq \cdots \leq \alpha_N$ as the eigenvalues of Φ^S and L^S respectively from smallest to largest, then*

$$-1 \leq \lambda_1 \leq \cdots \leq \lambda_N \leq 1 \; and \; 0 \leq \alpha_1 \leq \cdots \leq \alpha_N \leq 2.$$

Proof of Lemma SM.3. Given that $L^S = I - \Phi^S = D^{-1/2}(D - A)D^{-1/2} = D^{-1/2}(L_G)D^{-1/2}$, where $A \approx \boldsymbol{\pi} B \boldsymbol{\pi}'$ and $L_G = D - A$,
it is well known that 0 is one of the eigenvalues of L_G. Therefore, we assume $\vec{v}_0$ is the eigenvector corresponding to 0 eigenvalue of L_G. We want to show that 0 is also an eigenvalue of L_S. Then,

$$L^S(D^{1/2}\vec{v}_0) = D^{-1/2}L_G D^{-1/2}D^{1/2}\vec{v}_0 = D^{-1/2}(L_G\vec{v}_0) = 0.$$

It shows that $D^{-1/2}\vec{v}_0$ is eigenvector of L^S corresponding to 0.

Then, we show that 0 is also the smallest eigenvalue of L^S by proving that L^S is positive semi-definite.

Assume that $x \in \mathbb{R}^N$, then

$$
\begin{aligned}
xL^S x' &= x(I - \Phi^S)x' \\
&= \sum_{i \in N} x_i^2 - \sum_{i,j \in N} \frac{x_i x_j w_{ij}}{\sqrt{d_i d_j}} \\
&= \frac{1}{2}\Big[\sum_{i \in N} x_i^2 + \sum_{j \in N} x_j^2 - \sum_{i,j \in N} \frac{2x_i x_j w_{ij}}{\sqrt{d_i d_j}} \Big] \\
&= \frac{1}{2}\Big[\sum_{i,j \in N} \frac{w_{ij}}{\sqrt{d_i d_j}} (x_i - x_j)^2 \Big] \geq 0
\end{aligned}
$$

Therefore, L^S is positive semi-definite and $\alpha_1 \geq 0$.
On the other hand, since $x(I - \Phi^S)x' \geq 0$, we have

$$
\frac{x\Phi^S x'}{xx'} \leq 1
$$

This Rayleigh quotient gives us the upper bound such that $\lambda_N \leq 1$.
Simialrly, we can show that $I + \Phi^S$ is also positive semi-definite, then

$$
\begin{aligned}
x(I + \Phi^S)x' &\geq 0 \\
xx' &\geq -x\Phi^S x' \\
\frac{x\Phi^S x'}{xx'} &\geq -1
\end{aligned}
$$

This Rayleigh quotient gives us the lower bound such that $\lambda_1 \geq -1$.
Similarly, we can also prove that $\alpha_N \leq 2$. $\square$

Lemma SM.4. *Given a $N \times N$ symmetric matrix A and $M \times M$ symmetric matrix A_1 which is a sub-matrix of A, where $M \leq N$, then*

$$
\|A\| \geq \|A_1\|.
$$

Proof of Lemma SM.4. Without loss of generality, assume $A = \begin{bmatrix} A_1 & A_3 \\ A_2 & A_4 \end{bmatrix}$, then,

$$
\begin{aligned}
\|A_1\| &= \sup_{\|x\|=1} \|A_1 x\| \\
&\leq \sup_{\|x\|=1} \|\begin{pmatrix} A_1 & A_2 \end{pmatrix}' x\| \\
&= \sup_{\|x\|=1} \|A\begin{pmatrix} x & 0 \end{pmatrix}'\| \\
&\leq \sup_{\|u\|=1} \|Au\| \\
&= \|A\| \quad \square
\end{aligned}
$$

Appendix B.2 Tables and Figures

Table 3: Groups for Empirical Study

Category			Words		
$Top\ 50$	russia	ukraine	russian	like	war
	one	dont	people	us	get
	even	ukrainian	think	country	time
	putin	know	gt	thats	make
	year	see	thing	could	want
	need	im	good	well	drone
	way	still	military	going	say
	go	right	nato	really	take
	back	look	day	cant	doesnt
	mean	said	lot	world	point
$Top\ 51-100$	support	attack	even	want	line
	support	attack	line	soldier	every
	let	isnt	probably	work	give
	weapon	western	theyre	part	end
	since	side	trump	didnt	anything
	around	use	got	west	maybe
	already	keep	video	guy	force
	kursk	sure	used	better	shit
	long	never	state	yes	missile
	nothing	actually	enough	power	youre
	something	new	come	china	first

Table 4: Partitions for Ukraine-Russia War Comment Dataset

Group Number			Words		
$Group\ 1$	right	people	even	thing	drone
$Group\ 2$	us	world	war	nato	
$Group\ 3$	russia	ukraine	russian	ukrainian	putin
	like	one	dont	get	think
	time	know	thats	make	see
	could	im	good	well	way
	military	going	really	take	back
	look	day	cant	mean	lot
$Overlap\ Group\ 1\ and\ 2$	say	doesnt	country	said	point
	gt	year	want	need	
$Overlap\ Group\ 1\ and\ 3$	go	cant	still		

Appendix B.3 Determining the Number of Communities

The algorithm presented in this paper requires specifying the number of communities. Therefore, we introduce a heuristic method to assist with this selection. However, a detailed theoretical analysis is beyond the scope of this work. This method is an extension of the non-backtracking approach proposed by Le and Levina (Le and Levina, 2015; Li et al., 2022). The overlapping problem is treated approximately as a non-overlapping block detection problem, aiming to identify the number of significant diagonal blocks in the adjacency matrix.

Input: multiple time series, Y_t for $t = 1, \cdots, T$.

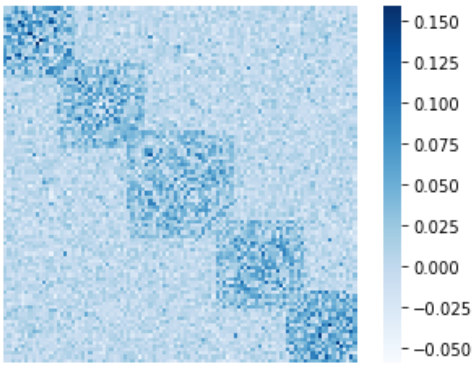
1. Estimate the autoregressive coefficients $\hat{\Phi}_l$ for $l = 1, \cdots, p$ using OLS.
2. Symmetrize the VAR matrices: $\hat{\phi}_l^S = \hat{\phi}_l + \hat{\phi}_l^T$ for $l = 1, \cdots, p$.
3. Try different k from 2 onwards, apply k-means to $\hat{\phi}_l^S$ and reorder series based on the results.
4. With a pre-specified threshold γ_k, get A_k by setting all elements of $\hat{\phi}_l^S > \gamma_k$ to 1; otherwise, to 0.
5. Define $B_{nb,k} = \begin{bmatrix} 0 & D_k - I \\ -I & A_k \end{bmatrix}$ where A_k is the result from last step, D_k is the degrees of A_k and I is identity matrix.
6. Find the number of eigenvalues, $\hat{k}$, such that $|\lambda_i| > \|B_{nb}\|^{1/2}$ where λ_i is i-th eigenvalue of $\|B_{nb}\|^{1/2}$.
7. Check if or not the pre-specified k is equal to $\hat{k}$ and find the maximal $\hat{k}$ at which $\hat{k} \neq k$.

Output: return $\hat{k} - 1$, the estimated number of communities.

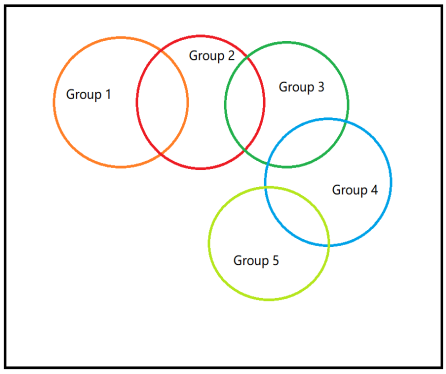
Appendix B.4 Further Simulation Study

Appendix B.4.1 Second Overlapping Case

$\vec{\pi}_i$'s are set up as follows:
$\{\vec{\pi}_i\}_{i=1}^{18} = [1,0,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=19}^{20} = [0.5,0.5,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=21}^{38} = [0,1,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=39}^{40} = [0,0.5,0.5,0,0]^T$, $\{\vec{\pi}_i\}_{i=41}^{60} = [0,0,1,0,0]^T$, $\{\vec{\pi}_i\}_{i=61}^{62} = [0,0,0.5,0.5,0]^T$, $\{\vec{\pi}_i\}_{i=63}^{80} = [0,0,0,1,0]^T$, $\{\vec{\pi}_i\}_{i=81}^{82} = [0,0,0,0.5,0.5]^T$ and $\{\vec{\pi}_i\}_{i=83}^{100} = [0,0,0,0,1]^T$.



(a) Heatmap of adjacency matrix for estimated symmertized VAR matrices



(b) Illustration for overlapping structure

Figure 9: (a) is the heatmap of the estimated symmetrized VAR matrices based on samples simulated from the MMSB-VAR model. (b) is a Venn diagram illustrating the overlapping group structure. In this study, we observe the following overlap pattern: group 1 overlaps with group 2, group 2 overlaps with group 3, group 3 overlaps with group 4, and group 4 overlaps with group 5.

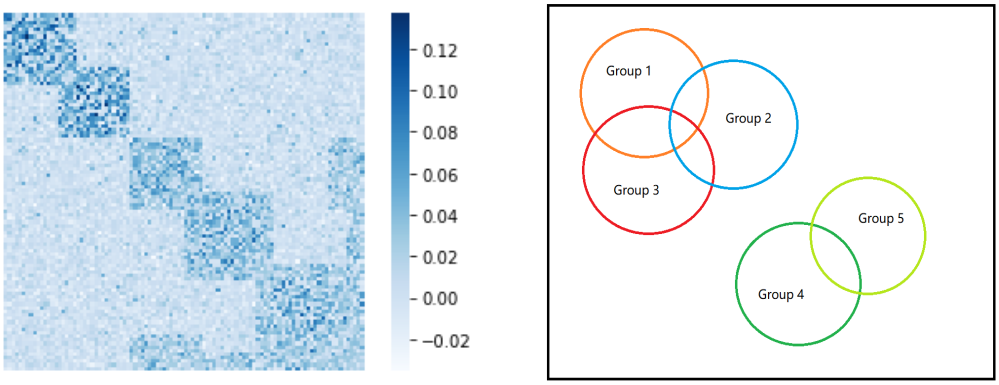
Table 5: **Accuracy Table for First Simulation with 8 single-membership series and 92 multiple-membership series.**

	n=100					
b_{11},b_{22}/b_{12},b_{21}	2000	5000	8000	10000	15000	20000
Hit Ratio (Single Membership)						
0.25/0.01	41.69%	70.92%	80.87%	83.82%	86.42%	88.27%
0.5/0.01	55.49%	89.6%	96.53%	97.89%	99.09%	99.4%
0.25/0.05	16.28%	26.07%	32.59%	35.1%	38.79%	41.44%
Misplacement Ratio (Non-overlapping Part)						
0.25/0.01	5.69%	2.07%	1.18%	0.93%	0.79%	0.56%
0.5/0.01	1.91%	0.14%	0.02%	0.0%	0.0%	0.0%
0.25/0.05	13.21%	10.56%	8.41%	8%	7.16%	6.6%
Hit Ratio (Multiple Membership)						
0.25/0.01	79.65%	86.05%	91.28%	92.23%	93.38%	94.6%
0.5/0.01	87.88%	98.88%	99.83%	99.95%	99.98%	100%
0.25/0.05	77.15%	76.03%	77.95%	77.45%	77.48%	78.93%
Misplacement Ratio (Multiple Membership)						
0.25/0.01	100%	100%	100%	100%	100%	100%
0.5/0.01	100%	100%	41.15%	24.73%	10.5%	6.85%
0.25/0.05	100%	100%	100%	100%	100%	100%

The table reports the accuracy of the algorithm brought forwards in section 2.3 applied to the dataset simulated from MMSM-VAR defined in definition 2.

Appendix B.4.2 Third Overlapping Case

$\vec{\pi}_i$'s are set up as follows:
$\{\vec{\pi}_i\}_{i=1}^{15} = [1,0,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=16}^{20} = [0.5,0.5,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=21}^{35} = [0,1,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=36}^{50} = [0,0,1,0,0]^T$, $\{\vec{\pi}_i\}_{i=51}^{55} = [0,0,0.5,0.5,0]^T$, $\{\vec{\pi}_i\}_{i=56}^{70} = [0,0,0,1,0]^T$, $\{\vec{\pi}_i\}_{i=71}^{75} = [0,0,0.5,0.5,0]^T$, $\{\vec{\pi}_i\}_{i=76}^{90} = [0,0,0,0,1]^T$, $\{\vec{\pi}_i\}_{i=91}^{95} = [0,0,0.5,0,0.5]^T$ and $\{\vec{\pi}_i\}_{i=96}^{100} = [0,0,0.33,0.33,0.33]^T$.



(a) Heatmap of adjacency matrix for estimated symmertized VAR matrices

(b) Illustration for overlapping structure

Figure 10: (a) is the heatmap of the estimated symmetrized VAR matrices based on samples simulated from the MMSB-VAR model. (b) is a Venn diagram illustrating the overlapping structure. In this study, we observe that the first two groups exclusively overlap with each other, while the last three groups exclusively overlap among themselves.

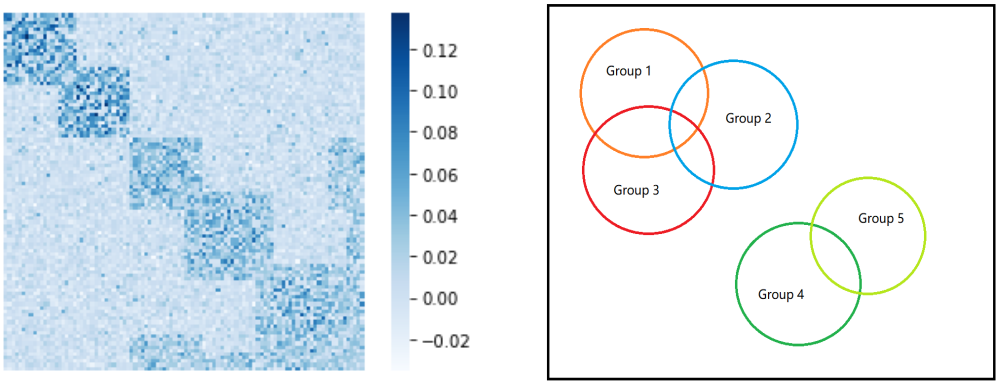
Table 6: **Accuracy Table for First Simulation with 25 single-membership series and 75 multiple-membership series.**

	n=100					
b_{11},b_{22}/b_{12},b_{21}	2000	5000	8000	10000	15000	20000
Hit Ratio (Single Membership)						
0.25/0.01	28.29%	52.49%	63.71%	68.67%	72.25%	77.72%
0.5/0.01	38.1%	74.85%	91.1%	94.65%	98.14%	98.78%
0.25/0.05	15.59%	21.88%	27.15%	29.25%	33.5%	35.41%
Misplacement Ratio (Single Membership)						
0.25/0.01	16.45%	11.84%	9.49%	8.48%	7.09%	6.28%
0.5/0.01	11.05%	3.83%	1.2%	0.67%	0.22%	0.17%
0.25/0.05	19.34%	19.31%	18.09%	17.38%	16.6%	16.21%
Hit Ratio (Multiple Membership)						
0.25/0.01	59.97%	63.94%	67.62%	69.36%	72.67%	73.5%
0.5/0.01	67.78%	79.98%	85.73%	86.07%	88.26%	89.82%
0.25/0.05	58.19%	56.79%	58.61%	59.32%	60.36%	60.02%
Misplacement Ratio (Multiple Membership)						
0.25/0.01	100%	100%	89.7%	77.26%	60.6%	53.76%
0.5/0.01	100%	65.68%	23.41%	13.94%	5.28%	3.21%
0.25/0.05	100%	100%	100%	100%	100%	100%

The table reports the accuracy of the algorithm brought forwards in section 2.3 applied to the dataset simulated from MMSM-VAR defined in definition 2.

Appendix B.4.3 Fourth Overlapping Case

$\vec{\pi}_i$'s are set up as follows:
$\{\vec{\pi}_i\}_{i=1}^{20} = [1,0,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=21}^{25} = [0.5,0.5,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=26}^{45} = [0,1,0,0,0]^T$, $\{\vec{\pi}_i\}_{i=46}^{60} = [0,0,1,0,0]^T$, $\{\vec{\pi}_i\}_{i=61}^{62} = [0,0,0.5,0.5,0]^T$, $\{\vec{\pi}_i\}_{i=63}^{78} = [0,0,0,1,0]^T$, $\{\vec{\pi}_i\}_{i=79}^{80} = [0,0,0.5,0.5,0]^T$, $\{\vec{\pi}_i\}_{i=81}^{96} = [0,0,0,0,1]^T$, $\{\vec{\pi}_i\}_{i=97}^{98} = [0,0,0.5,0,0.5]^T$ and $\{\vec{\pi}_i\}_{i=99}^{100} = [0,0,0.33,0.33,0.33]^T$.



(a) Heatmap of adjacency matrix for estimated symmertized VAR matrices

(b) Illustration for overlapping structure

Figure 11: (a) is the heatmap of the estimated symmetrized VAR matrices based on samples simulated from the MMSB-VAR model. (b) is a Venn diagram illustrating the overlapping structure. In this study, we observe that the first two groups exclusively overlap with each other, while the last three groups exclusively overlap among themselves.

Table 7: **Accuracy Table for First Simulation with 13 single-membership series and 87 multiple-membership series.**

	n=100					
b_{11},b_{22}/b_{12},b_{21}	2000	5000	8000	10000	15000	20000
Hit Ratio (Single Membership)						
0.25/0.01	37.25%	66.45%	77.57%	80.43%	83.63%	86.23%
0.5/0.01	49.67%	86.43%	95.07%	97.25%	98.84%	99.12%
0.25/0.05	16.27%	25.3%	30.75%	33.27%	38.4%	40.25%
Misplacement Ratio (Single Membership)						
0.25/0.01	8.14%	3.58%	2.38%	1.99%	1.65%	1.27%
0.5/0.01	3.46%	0.29%	0.06%	0.03%	0.0%	0.0%
0.25/0.05	14.88%	12.38%	11.29%	10.54%	8.75%	8.84%
Hit Ratio (Multiple Membership)						
0.25/0.01	70.67%	75.48%	78.64%	80.41%	82.27%	83.84%
0.5/0.01	79.98%	88.58%	90.95%	91.23%	91.58%	92.13%
0.25/0.05	69.15%	68.75%	68.91%	68.5%	69.1%	69.91%
Misplacement Ratio (Multiple Membership)						
0.25/0.01	100%	100%	100%	100%	90.44%	77.13%
0.5/0.01	100%	77.34%	27.05%	15.09%	6.21%	4.77%
0.25/0.05	100%	100%	100%	100%	100%	100%

The table reports the accuracy of the algorithm brought forwards in section 2.3 applied to the dataset simulated from MMSM-VAR defined in definition 2.