Community Detection in Partial Correlation Network Models

Christian Brownlees[†,◇] Guðmundur Stefán Guðmundsson[*] Gábor Lugosi[†,‡,◇]

July 15, 2020

Abstract

We introduce a class of partial correlation network models with a community structure for large panels of time series. In the model, the series are partitioned into latent groups such that correlation is higher within groups than between them. We then propose an algorithm that allows one to detect the communities using the eigenvectors of the sample covariance matrix. We study the properties of the procedure and establish its consistency. The methodology is used to study real activity clustering in the U.S.

Keywords: Community Detection, Graphical Models, Partial Correlation Networks, Random Graphs, Spectral Clustering

JEL: C3, C33, C55

[†] Department of Economics and Business, Universitat Pompeu Fabra and Barcelona GSE;
e-mail: christian.brownlees@upf.edu; gabor.lugosi@upf.edu.
[*] CREATES and Department of Economics and Business Economics, Aarhus University
e-mail: gsgudmundsson@econ.au.dk.
[‡] ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain
[◇] Barcelona Graduate School of Economics
Christian Brownlees and Gábor Lugosi acknowledge financial support from the Spanish Ministry of Economy and Competitiveness, Grant PGC2018-101643-B-I00 and from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075). G.L. was also supported by the grant "High-dimensional problems in structured probabilistic models - Ayudas Fundación BBVA a Equipos de Investigación Científica 2017"; and Google Focused Award "Algorithms and Learning for AI". Guðmundur Stefán Guðmundsson acknowledges financial support from the Spanish Ministry of Education, Culture and Sports (Grant FPU14/05022). We have benefited from discussions with Majid Al-Sadoon, Banu Demir Pakel, Jin-Chuan Duan, Geert Mesters, Fabrizio Lillo, Cavit Pakel, Nicolas Verzelen, Pedro Souza, Michael Wolf and Piotr Zwiernik.

1 Introduction

Network analysis has become an active field of research in time series econometrics (Hautsch, Schaumburg, and Schienle, 2015; Diebold and Yılmaz, 2014). Contributions in this area focus on developing methods to learn the interdependence structure of large multivariate systems. This is often achieved by assuming sparsity in the dependence structure of the system and relying on sparse estimation methods (Kock and Callot, 2015; Medeiros and Mendes, 2016).

In the analysis of real-world networks it is often found that the vertices are partitioned into groups such that the concentration of linkages depends on whether they belong to the same group or not. This empirical regularity is called *community structure* or *clustering*. A popular model used to study this type of networks is the Stochastic Block Model (SBM) (Holland, Laskey, and Leinhardt, 1983). The SBM is a random graph, that is, a graph in which the vertices are fixed and the edges are determined randomly and independently by Bernoulli trials (Chung and Lu, 2006). In the simplest version of the model, the vertices are partitioned into groups and the probability of an edge between two vertices is p if they belong to the same group and q otherwise. It is typically assumed that $p > q$ so the vertices are more likely to be linked within the same group. This feature is referred to as *homophily* and has been frequently observed empirically (McPherson, Smith-Lovin, and Cook, 2001).

In this work we introduce a class of partial correlation network models with a community structure for large panels of time series. In the model, the series are partitioned into latent groups such that correlation is higher within groups than between them. This is achieved by letting the partial correlation structure of the n series in the panel be determined by a latent graph defined over n vertices: The i-th and j-th time series have zero partial correlation if and only if vertices i and j are not joined by an edge. The graph that determines the partial correlation structure is a generalised version of the SBM.

A natural question that arises in this setting is how to detect the communities of the model from a sample of T observations. An extensive literature (Fortunato, 2010) deals with the problem of community detection when the network structure of the data

is observable. An example of a popular community detection algorithm is spectral clustering, which uses the eigenvectors of the graph Laplacian to detect community structure. In our framework, community detection is more challenging as the network structure of the data is not observed. Motivated by spectral clustering methods (Ng, Jordan, and Weiss, 2001; von Luxburg, 2007), we propose a community detection procedure called *Blockbuster* in which spectral clustering is applied to the sample covariance matrix of the observations. In particular, the algorithm applies k-means clustering to a matrix whose columns are the rescaled eigenvectors corresponding to the k largest eigenvalues of the sample covariance matrix. The k-means partition of the rows of the rescaled eigenvector matrix is the estimate of the community partition.

We establish that Blockbuster consistently detects the communities when the number of observations and the dimension of the panel are sufficiently large. More precisely, our key result establishes a bound on the fraction of vertices that the algorithm misclusters in a similar fashion to Rohe, Chatterjee, and Yu (2011) and shows that it is close to zero with high probability when n and T are large enough, provided that n/T is small.

We consider an extension of our model in which the time series in the panel are influenced by a set of common factors. Panels of economic and financial time series typically exhibit evidence of a factor structure and an extensive literature has developed around factor models (Forni, Hallin, Lippi, and Reichlin, 2000; Bai and Ng, 2013). We introduce a variant of our algorithm that consistently detects communities in this setting.

Moreover, we introduce a regularised covariance estimator based on the algorithm, motivated by the block covariance structure of the model introduced in this work. The estimator takes a block-structured form where the diagonal blocks are the sample covariance matrices of each community, while thresholding is applied to the off-diagonal blocks. The analysis of the theoretical properties of this estimator is beyond the scope of the paper, but we use it in an out-of-sample forecast validation exercise.

A natural application of our methodology is to study business cycle synchronisation (Hamilton and Owyang, 2012; Leiva-Leon, 2017), where it is of interest to find regions that co-move closely over the business cycle (Francis, Owyang, and Savascin, 2017). We

apply our methodology to analyse a dataset constructed by Hamilton and Owyang (2012), comprising quarterly employment growth rates at the state-level for the U.S. from 1956-Q2 to 2007-Q4. Blockbuster delivers a meaningful partition of the states. In particular, the U.S. communities bear close resemblance to previously published results by Hamilton and Owyang (2012). We also carry out an out-of-sample validation exercise and show that the Blockbuster covariance regularisation procedure improves covariance prediction compared to a number of alternative procedures.

This work is related to several different strands of the literature. First, the literature on estimation of sparse high-dimensional network models, see for example Demirer, Diebold, Liu, and Yilmaz (2018), Barigozzi and Brownlees (2019). Second, the literature on SBMs and community detection, see amongst others Abbe, Bandeira, and Hall (2016), Sarkar and Bickel (2015), Arias-Castro and Verzelen (2014) and Jin (2015). Our main contribution with respect to this literature consists of extending the theory of SBMs to a setting in which the random graph is not observed, under more general assumptions that are better tailored for economic applications. Third, the literature on identifying latent group structures in econometrics. Some examples of papers in this area are Hamilton and Owyang (2012), Ando and Bai (2016), Francis *et al.* (2017), Su, Shi, and Phillips (2016), Bonhomme and Manresa (2015). Fourth, the literature on identification of network effects in panel models, see, among others de Paula (2017). Last, this paper is related to Gudmundsson (2018) where the community detection problem is studied in the context of large dimensional Vector Autoregressions.

Section 2 introduces the methodology. Proofs of the main theoretical results are in the Appendix, together with further empirical results. In Section 3 we conduct a simulation study and Section 4 is the empirical application. Concluding remarks follow in Section 5.

2 Methodology

We introduce a model for an n-dimensional random vector $\boldsymbol{Y}_t = (Y_{1t}, \ldots, Y_{nt})'$. The key feature of the model is that the interdependence structure of the components of $\boldsymbol{Y}_t$ is

determined by a latent random graph $\mathcal{G}$ that is endowed with a community structure.

A number of definitions relating to graphs and random graphs have to be laid out before introducing the model. Throughout the paper we use the term graph to refer to an undirected weighted graph. An undirected weighted graph is a triplet $(\mathcal{V}, \mathcal{E}, \mathcal{W})$ where $\mathcal{V} = \{1, \ldots, n\}$ is the vertex set, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the edge set and $\mathcal{W}$ is the set of edge weights. For an edge weight $w_{ij} \in \mathcal{W}$, we have $w_{ij} > 0$ if $(i, j) \in \mathcal{E}$ and $w_{ij} = 0$ otherwise. As the graph is undirected, we have $w_{ij} = w_{ji}$. In order to represent the structure of a graph it is useful to introduce the adjacency, degree and Laplacian matrices. The adjacency matrix $\mathbf{A}$ of a graph is defined as an $n \times n$ matrix with the (i, j)-th element $[\mathbf{A}]_{ij}$ equal to w_{ij}. Let d_i denote the degree of vertex i, that is, $d_i = \sum_{j=1}^{n} [\mathbf{A}]_{ij}$. We define the degree matrix $\mathbf{D}$ as an $n \times n$ diagonal matrix with element $[\mathbf{D}]_{ii} = d_i$. Finally, the symmetric degree-normalised Laplacian is defined as $\mathbf{L} = \mathbf{I}_n - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Note that both the adjacency matrix and the Laplacian are symmetric. Finally, in this paper a *random graph* is a graph in which the vertex set $\mathcal{V}$ is fixed whereas the existence of an edge in $\mathcal{E}$ is determined by a Bernoulli trial, independently of all other edges.

We assume that the latent network $\mathcal{G}$ is generated by a generalised version of the popular stochastic block model. The stochastic block model (Holland *et al.*, 1983) is an extension of the Erdős-Rényi random graph in which the vertex set $\mathcal{V}$ is partitioned into k subsets $\mathcal{V}_1, \ldots \mathcal{V}_k$, typically referred to as communities. An edge is present between vertices i and j with probability p_s if both vertices belong to $\mathcal{V}_s$ and probability q_{sv} if they belong to $\mathcal{V}_s$ and $\mathcal{V}_v$, respectively, with $s \neq v$. Figure 1 gives an example of a stochastic block model with $n = 50$, $k = 2$, $p_s = p = 0.25$ and $q_{sv} = q = 0.01$. In the stochastic block model, all vertices within a given community have the same expected degree and edges are unweighted. In our framework this turns out to be a rather restrictive assumption. We therefore assume that $\mathcal{G}$ is generated by a weighted and degree-corrected stochastic block model (Karrer and Newman, 2011), an extension of the stochastic block model that allows for a general degree distribution as well as weighted edges. We call this the Generalised Stochastic Block Model (GSBM) and formally define it as follows.

Definition 1 (Generalised Stochastic Block Model)**.** *Let* $\mathbf{Z}$ *be the* $n \times k$ *community membership matrix, such that* $[\mathbf{Z}]_{is} = 1$ *if vertex* i *belongs to community* s *and zero otherwise. Each vertex belongs to exactly one community, so for all* i, $\sum_s [\mathbf{Z}]_{is} = 1$*. Let* $\mathbf{B}$ *be the symmetric* $k \times k$ *matrix of community-specific edge probabilities and let* $\mathbf{\Theta}$ *be the* $n \times n$ *diagonal matrix of non-negative, fixed and unknown vertex-specific probability weights. Let* $\mathbf{W}$ *be an* $n \times n$ *symmetric matrix where each element* $[\mathbf{W}]_{ij}$ *is a random variable supported on the interval* $[\underline{w}, \overline{w}]$ *with* $0 < \underline{w} \leq \overline{w}$*, and expected value* μ*.*

In a Generalised Stochastic Block Model the probability of an edge between vertices i *and* j *that belong to communities* s *and* v*, respectively, is* $[\mathbf{\Theta}]_{ii}[\mathbf{B}]_{sv}[\mathbf{\Theta}]_{jj} \leq 1$ *and all edges are independent. Furthermore, each edge* (i, j) *is associated with a weight* $[\mathbf{W}]_{ij}$ *which is drawn independently of all other weights, and all edges. It is convenient to write*

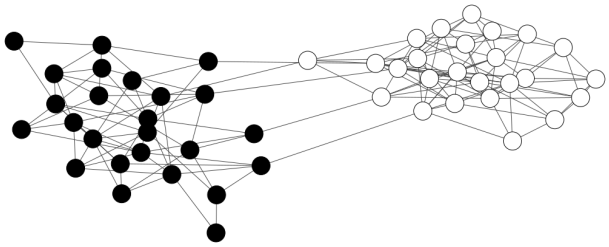
$$\mathcal{G} \sim \text{GSBM}\left(\mathbf{Z}, \mathbf{B}, \mathbf{\Theta}, \mathbf{W}\right), \tag{1}$$

to indicate that a random graph $\mathcal{G}$ *is a Generalised Stochastic Block Model.*

Notice that the matrix $\mathbf{Z}$ defines the community partition $\mathcal{V}^k = \{\mathcal{V}_1, \ldots \mathcal{V}_k\}$ of the vertex set $\mathcal{V}$.[1] We let $n_s = |\mathcal{V}_s|$ denote the size of community s. We may assume that the elements of $\mathcal{V}_1$ are the first n_1 elements of $\mathcal{V}$, the elements of $\mathcal{V}_2$ the next n_2, and so on. The probability of an edge between vertices i and j that belong, respectively, to communities s and v is $[\mathbf{\Theta}]_{ii}[\mathbf{B}]_{sv}[\mathbf{\Theta}]_{jj}$. The matrix $\mathbf{B}$ contains the community-specific component of the edge probabilities whereas the matrix $\mathbf{\Theta}$ contains the vertex-specific component. This allows vertices that belong to the same community to have different expected degrees. As $\mathbf{\Theta}$ and $\mathbf{B}$ are only unique up to a multiplicative constant, we normalise $[\mathbf{\Theta}]_{ii}$ to sum to one within communities, that is, $\sum_{i \in \mathcal{V}_s} [\mathbf{\Theta}]_{ii} = 1$ for all $s = 1, \ldots, k$ (Karrer and Newman, 2011). Last, the random matrix $\mathbf{W}$ contains the weights associated with the edges. We allow entries of $\mathbf{W}$ to be heterogeneous yet with the same expected value μ.

[1] We assume with no loss of generality that there are no empty communities.

Figure 1: The Stochastic Block Model



The figure shows a realisation of a stochastic block model with $n = 50$, $k = 2$, $p = 0.25$ and $q = 0.01$.

In this work the interdependence structure of the random vector $\boldsymbol{Y}_t$ ($t \in \{1, \ldots, T\}$) is characterised by partial correlations. We assume, without loss of generality, that for all t, $\boldsymbol{Y}_t$ is mean zero with covariance matrix $\boldsymbol{\Sigma} = \mathbb{E}[\boldsymbol{Y}_t \boldsymbol{Y}_t']$. The partial correlation between Y_{it} and Y_{jt} measures the linear dependence between the two variables after partialling out the influence of the remaining variables in the panel. We formally define it as $\rho^{ij} = \text{Corr}(e_{it}, e_{jt})$, where e_{it} and e_{jt} are the prediction errors of the best linear predictors of Y_{it} and Y_{jt}, respectively, based on $\{Y_{st} : 1 \leq s \neq i, j \leq n\}$. It is well known that the linear partial dependence structure of the system is embedded in the inverse of the covariance matrix $\mathbf{K} = \boldsymbol{\Sigma}^{-1}$ (Dempster, 1972), which we refer to as the *precision* matrix hereafter. In fact we have that the elements of the precision matrix $[\mathbf{K}]_{ij}$ are related to the partial correlations through the identity

$$\rho^{ij} = -\frac{[\mathbf{K}]_{ij}}{\sqrt{[\mathbf{K}]_{ii}[\mathbf{K}]_{jj}}}, \quad i \neq j \ .$$

The derivation of this identity may be found, for example, in Pourahmadi (2013, Section 5.2). The precision matrix and the partial correlations of $\boldsymbol{Y}_t$ thus share the same sparsity structure: The (i, j)-th element of $\mathbf{K}$ is zero if and only if Y_{it} and Y_{jt} have zero partial correlation. If $\boldsymbol{Y}_t$ is Gaussian, a zero partial correlation between Y_{it} and Y_{jt} implies their conditional independence given the remaining components of $\boldsymbol{Y}_t$.

We introduce a model in which the partial correlation structure of the random vector $\boldsymbol{Y}_t$ is determined by a latent GSBM $\mathcal{G}$. We assume that the precision matrix $\mathbf{K}$ is a

function of the Laplacian $\mathbf{L}$ of the latent graph $\mathcal{G}$ through the equation

$$\mathbf{K} = \frac{1}{\sigma^2}\mathbf{I}_n + \frac{\phi}{\sigma^2}\mathbf{L}, \tag{2}$$

where $\sigma^2 > 0$ and $\phi \geq 0$ are the network variance and dependence parameters.

A number of comments on the specification in (2) are in order. First, it follows from the properties of the Laplacian that $\mathbf{K}$ is guaranteed to be symmetric and positive definite. Note that while it may be natural to make the precision matrix equal to (an appropriately rescaled version of) the Laplacian only, this would deliver a precision matrix that is not positive definite, as the Laplacian always has at least one zero eigenvalue. Adding the identity matrix in (2) is a natural way to ensure that all the eigenvalues of $\mathbf{K}$ are positive while preserving the eigenvector structure implied by $\mathbf{L}$. Second, the model is specified in a way such that the graph $\mathcal{G}$ determines the partial correlation structure of the vector $\boldsymbol{Y}_t$. In fact, we have that ρ^{ij} is non-zero if and only if i and j are joined by an edge in $\mathcal{G}$. Third, it is instructive to report the partial correlations and regression representation implied by the model. It follows from the basic properties of the precision matrix (Pourahmadi, 2013, Section 5.2) that the partial correlation coefficient for variables i and j is

$$\rho^{ij} = \frac{\phi}{1+\phi}\frac{w_{ij}}{\sqrt{d_i d_j}},$$

and that the regression representation of the components of $\boldsymbol{Y}_t$ is given by

$$Y_{it} = \sum_{j \in N(i)} \beta_{ij} Y_{jt} + \epsilon_{it}, \qquad \beta_{ij} = \frac{\phi}{1+\phi}\frac{w_{ij}}{\sqrt{d_i d_j}}, \qquad \mathrm{Var}(\epsilon_{it}) = \frac{\sigma^2}{(1+\phi)}, \tag{3}$$

where $N(i)$ is the set of vertices that have edges with i and we disallow self-loops for clarity. The formula in (3) implies that the realisations of the i-th component of $\boldsymbol{Y}_t$ can be represented as a weighted average of its neighbours in $\mathcal{G}$ plus an idiosyncratic shock. We remark that while specifying a precise functional form for the relation between $\mathcal{G}$ and $\mathbf{K}$ is clearly quite restrictive, the specification allows for a fair amount of flexibility in that each regression coefficient β_{ij} is a function of the pair specific weight w_{ij}. In

the equation of the β_{ij} coefficients we have that the pair specific weight w_{ij} is rescaled by the square root of the product of the degrees of i and j. Note that in general some appropriate form of rescaling is required when the weights w_{ij} are arbitrarily chosen in order to ensure that the covariance matrix implied by the model is positive definite. We point out that in our framework we impose the weights to be positive, which implies that the partial correlations of the model are non-negative. It is possible to relax this, however, the fraction of negative weights would have to be appropriately controlled to ensure the positive definiteness of $\mathbf{K}$. We leave this for future research. Barigozzi and Brownlees (2019) and Brownlees, Nualart, and Sun (2018) document that the fraction of negative partial correlations estimated in network models for economic panels is negligible, so this assumption has some empirical justification. Last, we also point out that the model in (3) may be interpreted as a spatial lag model with spatial autoregressive coefficient $\phi/(1+\phi) \in [0,1)$ and spatial weight matrix $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$.[2]

We conclude with the formal definition of our partial correlation network model.

Definition 2 (Stochastic Block Partial Correlation Model)**.** *Let* $\mathcal{G} \sim \text{GSBM}(\mathbf{Z}, \mathbf{B}, \mathbf{\Theta}, \mathbf{W})$ *be a Generalised Stochastic Block Model as in Definition* 1*. Let* $\mathbf{K}$ *be the* $n \times n$ *precision matrix corresponding to the random graph* $\mathcal{G}$*, defined as in* (2)*. In a Stochastic Block Partial Correlation Model, the* n*-dimensional stationary stochastic process* $\{\boldsymbol{Y}_t\}$ *is such that, for all* t*,* $\boldsymbol{Y}_t$ *has mean zero and covariance matrix* $\mathbf{\Sigma} = \mathbf{K}^{-1}$*.*

2.1 The Blockbuster Algorithm

Suppose we observe a sample $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T$ of observations from the model in Definition 2 arranged into the $T \times n$ matrix $\mathbf{Y} = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T]'$, and let k be the number of communities. The community structure of the model is assumed to be unknown, although we assume that k is known. We adapt spectral clustering techniques (Ng *et al.*, 2001) to detect the communities of the model from the sample in a procedure we call Blockbuster. The

[2] A spatial lag (or spatial autoregressive) model (Anselin, 2007) for a random vector $\boldsymbol{Y}_t$ is given by

$$\boldsymbol{Y}_t = \rho W \boldsymbol{Y}_{t-1} + \epsilon_t \ ,$$

where ρ is called spatial autoregressive coefficient and W is called spatial weight matrix.

proposed algorithm detects the communities using the rescaled eigenvectors of the sample covariance matrix. An important feature of the algorithm is that it allows one to detect the communities without estimating the network structure of the data.

We provide a description of Blockbuster in Algorithm 1. Given the sample and the number of communities k, first construct the $n \times k$ matrix $\widehat{\mathbf{U}}$ of the eigenvectors corresponding to the k largest eigenvalues of the sample covariance matrix $\widehat{\mathbf{\Sigma}} = (1/T)\mathbf{Y}'\mathbf{Y}$.[3] We refer to this matrix as the matrix of partitioning eigenvectors. Then form the matrix $\widehat{\mathbf{X}}$ by normalising the rows of $\widehat{\mathbf{U}}$ to unit length, that is, $\widehat{\mathbf{X}} = \widehat{\mathbf{N}}\widehat{\mathbf{U}}$ where $\widehat{\mathbf{N}}$ is an $n \times n$ diagonal matrix with its i-th element $[\widehat{\mathbf{N}}]_{ii} = 1/\left\|[\widehat{\mathbf{U}}]_{i\bullet}\right\|$ and $\left\|[\widehat{\mathbf{U}}]_{i\bullet}\right\|$ is the Euclidean norm of $[\widehat{\mathbf{U}}]_{i\bullet}$, the i-th row of $\widehat{\mathbf{U}}$. The algorithm then applies k-means clustering to the rows of $\widehat{\mathbf{X}}$. The k-means algorithm partitions a set of data points into k clusters by solving

$$\min_{\{\boldsymbol{m}_1,\ldots,\boldsymbol{m}_k\}} \sum_{i=1}^{n} \min_{s} \|[\widehat{\mathbf{X}}]_{i\bullet} - \boldsymbol{m}_s\|^2. \tag{4}$$

Let $\boldsymbol{m}_s^* \in \mathbb{R}^k$ be the vectors that solve the problem.[4][5] These vectors are called the centroids of the clusters returned by k-means. Each row $[\widehat{\mathbf{X}}]_{i\bullet}$ is then assigned to the cluster corresponding to the centroid closest to it. This yields a partition of the vertex set $\widehat{\mathcal{V}}^k = \{\widehat{\mathcal{V}}_1, \ldots, \widehat{\mathcal{V}}_k\}$, which is our estimator of the community partition $\mathcal{V}^k$.

2.2 Theory

In this section we show that the fraction of vertices that are incorrectly clustered by Blockbuster is close to zero with probability close to one when the cross-sectional dimension n and the number of observations T are large enough, provided n/T is small. The proofs of this section are collected in the Appendix A.

We introduce additional notation to be used throughout. Let $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_F$ denote the spectral and Frobenius norms of the $n \times n$ matrix $\mathbf{A}$, respectively. We denote the

[3] If the k-th and $k+1$-th are tied, we may take both.

[4] The solution is not necessarily unique. We let $\boldsymbol{m}_s^*$ be some set of vectors that achieve the minimum.

[5] The exact solution is NP hard and the standard approximation to it is prone to local minima, which may be arbitrarily far from the solution. We use k-means++ which guarantees an approximation within a factor of $O(\log k)$ of the optimum. For simplicity of discussion, we assume that the minimum is computed. All theoretical findings remain true for k-means++.

Algorithm 1 The Blockbuster Algorithm

INPUT: Sample $\boldsymbol{Y}_t$ for $t = 1, \ldots, T$, number of communities k.

PROCEDURE:

1. Compute the sample covariance matrix $\widehat{\boldsymbol{\Sigma}}$.
2. Construct the $[n \times k]$ eigenvector matrix $\widehat{\mathbf{U}}$ such that its columns are the eigenvectors corresponding to the k largest eigenvalues of $\widehat{\boldsymbol{\Sigma}}$.
3. Standardise each row of $\widehat{\mathbf{U}}$ by its norm and denote the row-normalised eigenvector matrix by $\widehat{\mathbf{X}}$, so that $[\widehat{\mathbf{X}}]_{ij} = [\widehat{\mathbf{U}}]_{ij} / \left\| [\widehat{\mathbf{U}}]_{i\bullet} \right\|$.
4. Apply the k-means algorithm to the rows of $\widehat{\mathbf{X}}$.

OUTPUT: The k-means partition $\widehat{\mathcal{V}}^k = \{\widehat{\mathcal{V}}_1, \ldots, \widehat{\mathcal{V}}_k\}$.

i-th smallest eigenvalue of $\mathbf{A}$ as $\lambda_i(\mathbf{A})$. Notice that we use the convention $\lambda_1(\mathbf{A}) \leq \cdots \leq \lambda_n(\mathbf{A})$. We refer to the eigenvectors corresponding to the k largest (smallest) eigenvalues of $\mathbf{A}$ as its k top (bottom) eigenvectors, counting multiplicities. By $f(n) = O(g(n))$ we mean that there exist a real number $M > 0$ and a real number n_0 such that for all $n \geq n_0$ we have $|f(n)| \leq M|g(n)|$. Similarly, $f(n) = \Omega(g(n))$ means $|f(n)| \geq M|g(n)|$ for all $n \geq n_0$ and $M > 0$. We refer to an event that happens with probability approaching one as an event that happens with high probability.

Define $\mathcal{A} = \mathbb{E}[\mathbf{A}]$ as the population adjacency matrix of the graph $\mathcal{G}$. Notice that if $\mathcal{G} \sim \text{GSBM}(\mathbf{Z}, \mathbf{B}, \boldsymbol{\Theta}, \mathbf{W})$ we may decompose the population adjacency matrix as $\mathcal{A} = \mu\boldsymbol{\Theta}\mathbf{Z}\mathbf{B}\mathbf{Z}'\boldsymbol{\Theta}$. We also define the population degree matrix $\mathcal{D}$ as the diagonal matrix with $[\mathcal{D}]_{ii} = \sum_{j=1}^{n} [\mathcal{A}]_{ij}$, the population normalised Laplacian $\mathcal{L} = \mathbf{I}_n - \mathcal{D}^{-1/2}\mathcal{A}\mathcal{D}^{-1/2}$ and the population precision matrix $\mathcal{K} = \frac{1}{\sigma^2}\mathbf{I}_n + \frac{\phi}{\sigma^2}\mathcal{L}$, analogously to $\mathbf{D}$, $\mathbf{L}$ and $\mathbf{K}$. Let $\mathbf{U}$ and $\mathcal{U}$ be the $n \times k$ matrices of the bottom k eigenvectors of $\mathbf{K}$ and $\mathcal{K}$, respectively. Define $\mathbf{N}$ and $\mathcal{N}$ as $n \times n$ diagonal matrices with $[\mathbf{N}]_{ii} = \|[\mathbf{U}]_{i\bullet}\|^{-1}$ and $[\mathcal{N}]_{ii} = \|[\mathcal{U}]_{i\bullet}\|^{-1}$, respectively. Last, $\mathbf{X} = \mathbf{N}\mathbf{U}$ and $\mathcal{X} = \mathcal{N}\mathcal{U}$ are the row-normalised counterparts of $\mathbf{U}$ and $\mathcal{U}$.

In our framework only the time series $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T$ are observed and the community detection is based on the eigenvectors of the sample precision matrix $\widehat{\mathbf{K}} = \widehat{\boldsymbol{\Sigma}}^{-1}$. Notice that in our framework two layers of randomness are present. First, the precision matrix $\mathbf{K}$ differs from the population precision matrix $\mathcal{K}$ because of the randomness of the graph $\mathcal{G}$.

Second, the sample precision matrix $\widehat{\mathbf{K}}$ differs from $\mathbf{K}$ because of sampling uncertainty of the data $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T$. It is important to remind readers familiar with network data inference that in our framework the random graph $\mathcal{G}$ and thus $\mathbf{K}$ are unobserved.

We begin by establishing that the population precision matrix $\mathcal{K}$ allows one to trivially retrieve the communities of the random graph $\mathcal{G}$. To establish this result we require the following assumptions on the random graph.

Assumption 1. *We assume that* (i) $\mathbf{B}$ *is symmetric and positive definite and* (ii) *all the elements of* $\mathbf{B}$ *are proportional to the common rate variable* ρ_n*, with* $\rho_n = \Omega\left(\log(n)/n\right)$.

The first assumption implies the invertibility of $\mathbf{B}$, so that all k communities are distinguishable.[6] As an example, consider the simple version of the model where $p_s = p$ and $q_{sv} = q$ for all $s \neq v$. Then this assumption simply requires $p > q$. It is natural to characterise the behaviour of a random graph when n grows as a function of the edge probabilities in $\mathbf{B}$. The second assumption requires the probabilities to be at least of the order $\log(n)/n$, which is termed the semi-sparse regime by Sarkar and Bickel (2015). This is the sparsest regime where exact recovery of the communities is possible (Abbe *et al.*, 2016), even with the network fully observed.

The following lemma is an extension of Lemma 3.3 from Qin and Rohe (2013).

Lemma 1. *Let* $\mathcal{G} \sim \mathrm{GSBM}\left(\mathbf{Z}, \mathbf{B}, \boldsymbol{\Theta}, \mathbf{W}\right)$ *be a Generalised Stochastic Block Model as in Definition* 1*. Let* $\mathcal{K}$ *be the population precision matrix,* $\mathcal{U}$ *the matrix of its bottom* k *eigenvectors and* $\mathcal{X}$ *the row-normalised counterpart of* $\mathcal{U}$*. Suppose Assumption* 1 *holds. Then* $\lambda_i(\mathcal{K}) = (1+\phi)/\sigma^2$ *for all* $i = k+1, \ldots, n$ *and* $\lambda_i\left(\mathcal{K}\right) \in [1/\sigma^2, (1+\phi)/\sigma^2)$ *for* $i = 1, \ldots, k$*. Furthermore, there exists a* $k \times k$ *orthonormal matrix* $\mathbf{V}$ *such that* $\mathcal{X} = \mathbf{ZV}$.

In particular, the lemma implies $[\mathcal{X}]_{i\bullet} = [\mathcal{X}]_{j\bullet}$ if and only if $[\mathbf{Z}]_{i\bullet} = [\mathbf{Z}]_{j\bullet}$, so the rows corresponding to two vertices that belong to the same community are equal in $\mathcal{X}$. Hence there are only k different rows in $\mathcal{X}$ and if k-means was to be applied then it would trivially recover the partition $\mathcal{V}^k$ from $\mathcal{X}$ by selecting each of the different rows as centroids.

[6] In fact, we can easily allow $\mathbf{B}$ to be negative definite. This corresponds to a bipartite graph with a heterophilic community structure (Rohe *et al.*, 2011), but we do not pursue this here.

We proceed by showing that the sample row-normalised eigenvector matrix $\widehat{\mathbf{X}}$ associated with $\widehat{\mathbf{K}}$ is concentrated around its population counterpart $\mathcal{X}$. To this end, we require appropriate mixing and distributional assumptions. We derive the result conditional on the random graph $\mathcal{G}$. This is justified by the fact that observing a single random graph suffices in random graph concentration results like those of Oliveira (2009), provided that n is large enough. We formulate our assumptions on the sequence of isotropic random vectors $\mathbf{\Sigma}^{-1/2}\boldsymbol{Y}_t$.[7] Let $\mathcal{B}_{-\infty}^{r}$ and $\mathcal{B}_{r+m}^{\infty}$ be the σ-algebras generated by $\{\mathbf{\Sigma}^{-1/2}\boldsymbol{Y}_t : -\infty \leq t \leq r\}$ and $\{\mathbf{\Sigma}^{-1/2}\boldsymbol{Y}_t : r+m \leq t \leq \infty\}$, respectively. Define the α-mixing coefficients of the process as $\alpha(m) = \sup_r \sup_{A\in\mathcal{B}_{-\infty}^{r}, B\in\mathcal{B}_{r+m}^{\infty}} |\mathbb{P}(A\cap B|\mathcal{G}) - \mathbb{P}(A|\mathcal{G})\,\mathbb{P}(B|\mathcal{G})|$. We assume that the process $\{\boldsymbol{Y}_t\}$ satisfies the following assumptions.

Assumption 2. *Let $\{\boldsymbol{Y}_t\}$ be a zero-mean stationary process with covariance matrix $\mathbf{\Sigma} = \mathrm{E}[\boldsymbol{Y}_t\boldsymbol{Y}_t']$. We assume that (i) $\{\mathbf{\Sigma}^{-1/2}\boldsymbol{Y}_t\}$ is strongly mixing with mixing coefficients satisfying $\alpha(m) \leq e^{-c_1 m^{\gamma_1}}$ for any positive integer m, where $\gamma_1, c_1 > 0$ are constants; (ii) for any vector $\boldsymbol{x}$ with $\|\boldsymbol{x}\| = 1$ and for any $s > 0$, $\sup_{t>0}\mathbb{P}\left(\left|\boldsymbol{x}'\mathbf{\Sigma}^{-1/2}\boldsymbol{Y}_t\right| > s\middle|\mathcal{G}\right) \leq c_3 e^{-(s/c_2)^{\gamma_2}}$, where $\gamma_2, c_2, c_3 > 0$ are constants; and (iii) $\gamma < 1$ where $1/\gamma = 1/\gamma_1 + 1/\gamma_2$.*

The assumptions allow us to apply the concentration inequality of Merlevède, Peligrad, and Rio (2011) to the sample covariance matrix $\widehat{\mathbf{\Sigma}}$, which plays a key role in the following theorem. The second assumption imposes generalised exponential tails on the distribution of the isotropic vectors $\mathbf{\Sigma}^{-1/2}\boldsymbol{Y}_t$. In particular, this implies that the elements of the vector have generalised exponential tails similar to what Fan, Liao, and Mincheva (2013) assume.

Theorem 1 establishes the concentration of the sample row-normalised eigenvectors.

Theorem 1 (Concentration)**.** *Let $\boldsymbol{Y}_t$ for $t = 1, \ldots, T$ be observations from an SBPCM as in Definition 2. Let $\widehat{\mathbf{U}}$ the matrix of the bottom k eigenvectors of the sample precision matrix $\widehat{\mathbf{K}}$ and $\widehat{\mathbf{X}}$ its row-normalised counterpart. Let $\mathcal{U}$ and $\mathcal{X}$ be as in Lemma 1. Suppose Assumptions 1 and 2 hold. If $T = \Omega\left(n^{2/\gamma-1}\right)$, there exists a $k\times k$ orthonormal rotation matrix $\mathcal{O}$, that depends on $\widehat{\mathbf{X}}$ and $\mathcal{X}$, such that with, high probability,*

$$\left\|\widehat{\mathbf{X}} - \mathcal{X}\mathcal{O}\right\| = O\left(\frac{n}{\sqrt{T}} + \sqrt{\frac{\log n}{\rho_n}}\right).$$

[7] Notice that $\{\mathbf{\Sigma}^{-1/2}\boldsymbol{Y}_t\}$ is a sequence of isotropic random vectors as $\mathbb{E}\left[\mathbf{\Sigma}^{-1/2}\boldsymbol{Y}_t\boldsymbol{Y}_t'\mathbf{\Sigma}^{-1/2}|\mathcal{G}\right] = \mathbf{I}_n$.

A number of remarks are in order. First, the theorem establishes the concentration of the sample row-normalised eigenvectors $\widehat{\mathbf{X}}$ around their population analogue $\mathcal{X}$, up to a rotation. The rotation is required because $\widehat{\mathbf{K}} = \mathcal{K}$ does not necessarily imply $\widehat{\mathbf{U}} = \mathcal{U}$. In particular, $\mathcal{O}$ is a $k \times k$ orthonormal rotation matrix obtained from the singular value decomposition of $\mathcal{U}'\widehat{\mathbf{U}}$. Second, the number of observations T is required to be larger than the cross-sectional dimension n. The more fat-tailed and dependent the data are (γ closer to 1), the larger T has to be. Last, we remark that the concentration rate of the theorem does not depend on the ϕ and σ^2 parameters of the GSBM.

To bound the fraction of misclustered vertices, we follow the strategy of Rohe *et al.* (2011). We begin by noting that the k-means objective function of (4) can be written as

$$\min_{\{\boldsymbol{m}_1,\ldots,\boldsymbol{m}_k\}} \sum_{i=1}^{n} \min_{s} \|[\widehat{\mathbf{X}}]_{i\bullet} - \boldsymbol{m}_s\|^2 = \min_{\mathbf{M}\in\mathcal{M}(n,k)} \left\|\widehat{\mathbf{X}} - \mathbf{M}\right\|_F^2,$$

where $\mathcal{M}(n,k) = \{\mathbf{M} \in \mathbb{R}^{n\times k} : \mathbf{M} \text{ has no more than } k \text{ different rows}\}$. Define

$$\widehat{\mathbf{C}} = \operatorname*{argmin}_{\mathbf{M}\in\mathcal{M}(n,k)} \left\|\widehat{\mathbf{X}} - \mathbf{M}\right\|_F^2, \tag{5}$$

as the estimated centroid matrix. Its i-th row $[\widehat{\mathbf{C}}]_{i\bullet}$ is equal to the k-means centroid that is closest to row i of the eigenvector matrix $\widehat{\mathbf{X}}$, so that $[\widehat{\mathbf{C}}]_{i\bullet} \in \{\boldsymbol{m}_1^*, \ldots, \boldsymbol{m}_k^*\}$. It is clear that the k-means centroid matrix has no more than k different rows. We similarly define the population centroid matrix as

$$\mathcal{C} = \operatorname*{argmin}_{\mathbf{M}\in\mathcal{M}(n,k)} \|\mathcal{X}\mathcal{O} - \mathbf{M}\|_F^2. \tag{6}$$

We adopt the same definition of misclustered vertices as Rohe *et al.* (2011) and say that vertex i is correctly clustered if $[\widehat{\mathbf{C}}]_{i\bullet}$ is closer to $[\mathcal{C}]_{i\bullet}$ than any other population centroid $[\mathcal{C}]_{j\bullet}$ for $j \neq i$. The next lemma provides a condition that implies that vertex i is correctly clustered. This statement is established in Theorem 4.4 in Qin and Rohe (2013) (see also Rohe *et al.* (2011)).

Lemma 2. *Let* $\widehat{\mathbf{C}}$ *be the estimated centroid matrix from* (5) *and* $\mathcal{C}$ *the population centroid*

matrix from (6). *Then* $\left\|[\widehat{\mathbf{C}}]_{i\bullet} - [\mathcal{C}]_{i\bullet}\right\| < \left\|[\widehat{\mathbf{C}}]_{i\bullet} - [\mathcal{C}]_{j\bullet}\right\|$ *for all* $j \neq i$ *holds for all* i *that satisfy the condition* $\left\|[\widehat{\mathbf{C}}]_{i\bullet} - [\mathcal{C}]_{i\bullet}\right\| < \sqrt{1/2}$.

The lemma justifies bounding the number of misclustered vertices by the number of nodes that do not satisfy the condition $\left\|[\widehat{\mathbf{C}}]_{i\bullet} - [\mathcal{C}]_{i\bullet}\right\| < \sqrt{1/2}$, which is sufficient for vertex i to be correctly clustered. Define the set

$$\mathscr{M} = \left\{i : \left\|[\widehat{\mathbf{C}}]_{i\bullet} - [\mathcal{C}]_{i\bullet}\right\| \geq \sqrt{1/2}\right\}. \tag{7}$$

The set $\mathscr{M}$ hence contains the set of all misclustered nodes and its cardinality $|\mathscr{M}|$ is thus an upper bound of the number of misclustered vertices. Finally, Theorem 2 bounds the fraction $|\mathscr{M}|/n$. The theorem is in the spirit of Theorem 3.1 from Rohe *et al.* (2011) and Theorem 4.4 of Qin and Rohe (2013). The theorem shows that the fraction of misclassified vertices is close to zero with high probability when $\log n/(n\rho_n)$ and n/T are close to zero.

Theorem 2 (Misclustered Vertices)**.** *Consider an SBPCM as in Definition* 2 *and let* $\mathscr{M}$ *be as in* (7). *Suppose Assumptions* 1 *and* 2 *hold. If* $T = \Omega\left(n^{2/\gamma-1}\right)$, *we have with high probability*

$$\frac{|\mathscr{M}|}{n} = O\left(\frac{n}{T} + \frac{\log n}{n\rho_n}\right).$$

2.3 Discussion

A number of comments on the SBPCM of Definition 2 and the Blockbuster algorithm are in order. The model bears some similarities to a factor model with a block structure, that is, a model in which the variables are generated by a factor model and each variable loads on exactly one factor (Hallin and Liska, 2011). However, it is typically assumed in such models that the correlation across blocks is negligible so that the covariance matrix is approximately diagonal. Assume, for ease of exposition, that $p_s = p$ and $q_{sv} = q$. Then, a block factor model corresponds to q close to or equal to zero. On the contrary, we only require $p > q$ in our framework implying that there may be high correlation between blocks. Furthermore, notice that a block factor model is associated with a partial correlation network in which each community is a clique, that is, a sub-graph

where every possible pair of vertices is joined by an edge. Our model can replicate such a structure with $p = 1$ and $q = 0$, but it also allows for much sparser structures.

Alternative approaches to detect communities when the graph structure is unknown may be considered. One possibility is to estimate the partial correlation network and then apply spectral clustering to the estimated network Laplacian. The estimation of the graphical structure of the data is typically carried out using LASSO techniques, which requires the selection of a tuning parameter that determines the sparsity of the estimated network (Meinshausen and Bühlmann, 2006). Note that Meinshausen (2008) shows that LASSO may not recover the sparsity pattern consistently for some classes of graphical models. A highlight of our approach is that it allows us to learn the community structure without estimating the network.

The parameters of the GSBM may be estimated by pseudo-likelihood methods (Karrer and Newman, 2011; Amini, Chen, Bickel, and Levina, 2013) if one assumes that the graph is known. In our setting the graph is unobserved and hence these methods ought to be appropriately adapted. A possible strategy is to first use the GLASSO algorithm (Friedman, Hastie, and Tibshirani, 2011) to recover the network structure of the graph.

We point out that the methodology we develop is designed for stationary strongly-mixing data. The objective of our analysis is to establish that the Blockbuster algorithm performs well in a time series setting, irrespective of the specific time series model that may have generated the data. That being said, we point out that knowledge of the specific time series structure of the data may be used to improve the algorithm.

Finally, it should be noted that the sample covariance estimator in Blockbuster may be substituted with other covariance matrix estimators, for example a shrinkage estimator. By resorting to a shrinkage estimator for the covariance, one may obtain a better rate of convergence.

2.4 Community Detection in the Presence of Common Factors

We introduce an extension of the SBPCM in which the components of $\boldsymbol{Y}_t$ are influenced by common factors, and an algorithm that consistently detects communities in this setting.

Let $\boldsymbol{F}_t$ be an R-dimensional vector of common factors with mean zero and covariance matrix $\mathbf{I}_R$ and let $\boldsymbol{q}_r$ be n-dimensional fixed vectors of factor loadings for $r = 1, \ldots, R$. We assume that the random vector $\boldsymbol{Y}_t$ is generated as

$$\boldsymbol{Y}_t = \mathbf{Q}\boldsymbol{F}_t + \boldsymbol{\epsilon}_t, \tag{8}$$

where $\mathbf{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_R]$ is an $n \times R$ matrix of factor loadings and $\boldsymbol{\epsilon}_t$ is generated by an SBPCM as in Definition 2, with $\mathbb{E}[\boldsymbol{\epsilon}_t | \boldsymbol{F}_t] = \mathbf{0}$ and covariance matrix $\mathrm{Cov}(\boldsymbol{\epsilon}_t | \boldsymbol{F}_t) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. We may assume without loss of generality that $\|\boldsymbol{q}_1\| \geq \|\boldsymbol{q}_2\| \geq \ldots \geq \|\boldsymbol{q}_R\|$. We define the SBPCM with common factors as follows.

Definition 3. *Let $\mathcal{G} \sim \mathrm{GSBM}(\mathbf{Z}, \mathbf{B}, \boldsymbol{\Theta}, \mathbf{W})$ be a Generalised Stochastic Block Model as in Definition* 1. *Let $\mathbf{K}_{\boldsymbol{\epsilon}}$ be the $n \times n$ precision matrix corresponding to the random graph $\mathcal{G}$, defined as in* (2). *In an SBPCM with common factors, the n-dimensional stationary stochastic process $\{\boldsymbol{Y}_t\}$ is such that, for all t, $\boldsymbol{Y}_t$, defined as in* (8), *has mean zero and covariance matrix*

$$\boldsymbol{\Sigma} = \mathbf{K}_{\boldsymbol{\epsilon}}^{-1} + \sum_{r=1}^{R} \boldsymbol{q}_r \boldsymbol{q}_r'. \tag{9}$$

Consider a sample drawn from the model in Definition 3 satisfying Assumption 2, arranged into a $T \times n$ matrix $\mathbf{Y} = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T]'$. Under appropriate conditions, with high probability, the top R eigenvectors of $\widehat{\boldsymbol{\Sigma}} = (1/T)\mathbf{Y}'\mathbf{Y}$ correspond to the R factors and the next k eigenvectors are the partitioning eigenvectors. This motivates us to propose the following extension of our algorithm. Given $\mathbf{Y}$ and k, construct the $n \times k$ matrix $\widehat{\mathbf{U}}$ such that it contains the eigenvectors corresponding to the $(R+1)$-th largest to the $(R+k)$-th largest eigenvalues of $\widehat{\boldsymbol{\Sigma}}$. Then follow steps 3 and 4 of Algorithm 1.

It is convenient to state the results of this section in terms of the precision matrices rather than covariances. Let $\mathcal{K}_{\boldsymbol{\epsilon}}$ be the population precision matrix of $\boldsymbol{\epsilon}$ from (8) and $\mathcal{K}$ be the population precision matrix of $\boldsymbol{Y}$. Let $\boldsymbol{u}_i(\mathcal{K})$ and $\boldsymbol{u}_i(\mathcal{K}_{\boldsymbol{\epsilon}})$ for $i = 1, \ldots, n$ be their eigenvectors. Define $\mathcal{U}_{\boldsymbol{\epsilon}}$ as the $n \times k$ matrix of the bottom k eigenvectors of $\mathcal{K}_{\boldsymbol{\epsilon}}$ and $\mathcal{U}$ as the matrix of the eigenvectors corresponding to the $(R+1)$-th smallest to the $(R+k)$-th smallest eigenvalues of $\mathcal{K}$, so that $[\mathcal{U}]_{\bullet i} = \boldsymbol{u}_{R+i}(\mathcal{K})$ for $i = 1, \ldots, k$, where $[\mathcal{U}]_{\bullet i}$ refers to

the i-th column of $\mathcal{U}$. We impose the following assumptions.

Assumption 3. *Let the process* $\{\boldsymbol{Y}_t\}$ *be an SBPCM as in* Definition 3*, satisfying Assumption 2. We assume additionally that (i)* $\boldsymbol{q}_r'\boldsymbol{q}_v = 0$ *for all* $r, v = 1, \ldots, R$*,* $r \neq v$*; (ii)* $\boldsymbol{q}_r'[\mathcal{U}_\epsilon]_{\bullet i} = 0$ *for all* $r = 1, \ldots, R$ *and* $i = 1, \ldots, k$*; and (iii)* $\|\boldsymbol{q}_R\|^2 > \frac{\sigma^2\phi}{1+\phi}$.

The first assumption requires the factor loadings to be mutually orthogonal. The second implies that the factor loading vectors carry no information on the community structure nor the degrees of the underlying graph $\mathcal{G}$. We remark that this assumption is made for simplicity and may be violated in practice. In the Online Appendix we show through simulations that the procedure performs well even when weak correlation between the factors and communities is present. The third assumption guarantees that the factors are strong enough to be dominant in the spectrum of $\mathcal{K}$. Notice that this assumption is trivially satisfied for large enough n when the factors are strong and the norms diverge to infinity, as in for example Bai and Ng (2013). It is not restrictive in the case of weak factors either, where the norms of the factor loading vectors are bounded, see for example Onatski (2012).[8]

The following is an extension of Theorem 2 which allows for the presence of factors.

Theorem 3 (Misclustered Vertices II)**.** *Consider an SBPCM with common factors as in Definition* 3*. Let* $\mathscr{M}$ *be defined analogously to* (7)*. Suppose Assumptions* 1*,* 2 *and* 3 *hold. If* $T = \Omega\left(n^{2/\gamma-1}\right)$*, we have with high probability*

$$\frac{|\mathscr{M}|}{n} = O\left(\frac{n}{T}\|\boldsymbol{\Sigma}\|^2 + \frac{\log n}{n\rho_n}\right).$$

2.5 Community Structure and Covariance Estimation

The presence of a community structure in a panel suggests a natural covariance estimation strategy. The covariance matrix may be estimated by a block covariance matrix estimator where the diagonal blocks contain the sample covariance matrices of each community, while thresholding is applied to the off-diagonal blocks (Pourahmadi, 2013, Chapter 6). We provide a description of an estimator based on this idea in Algorithm 2.

[8]For further discussion on strong and weak factors, see Chudik, Pesaran, and Tosetti (2011).

Notice that we may choose a different thresholding parameter for each block of the covariance matrix in our estimator. This is a reasonable compromise between universal thresholding (where a single threshold level is chosen for all entries of the matrix) and adaptive thresholding (where a different threshold level is chosen for each entry) for panels exhibiting a block covariance structure. In the empirical application we choose the threshold parameters using an adaptation of the cross-validation criterion of Bickel and Levina (2008). We emphasise that the criterion does not ensure positive definiteness of the resulting covariance estimator. However, this was not a concern in our empirical exercise for reasonable ranges of the thresholding parameters.

Algorithm 2 Blockbuster Covariance Estimator

INPUT: Sample $\boldsymbol{Y}_t$ for $t = 1, \ldots, T$, number of communities k, threshold parameters λ_{ij} for $1 \leq i \neq j \leq k$.

PROCEDURE:

1. Run the Blockbuster algorithm and obtain an estimate of the community partition of the panel $\widehat{\mathcal{V}}^k = \{\widehat{\mathcal{V}}_1, \ldots, \widehat{\mathcal{V}}_k\}$.
2. Re-order the series in the panel so that the first $\hat{n}_1 = |\widehat{\mathcal{V}}_1|$ series are the ones in community $\widehat{\mathcal{V}}_1$, the following $\hat{n}_2$ series are the ones in $\widehat{\mathcal{V}}_2$, and so on.
3. Let $\widehat{\boldsymbol{\Sigma}}_{sv}$ denote the $\hat{n}_s \times \hat{n}_v$ sample covariance matrix of the series in community $\widehat{\mathcal{V}}_s$ with the series in community $\widehat{\mathcal{V}}_v$.
4. The Blockbuster covariance estimator $\widehat{\boldsymbol{\Sigma}}_B$ is defined as

$$\widehat{\boldsymbol{\Sigma}}_{B\,sv} = \begin{cases} \widehat{\boldsymbol{\Sigma}}_{ss} & s = v, \\ T_{\lambda_{sv}}(\widehat{\boldsymbol{\Sigma}}_{sv}) & \text{otherwise}, \end{cases}$$

   where $\widehat{\boldsymbol{\Sigma}}_{B\,sv}$ denotes the (s, v)-th $\hat{n}_s \times \hat{n}_v$ block of $\widehat{\boldsymbol{\Sigma}}_B$, and $T_{\lambda_{sv}}\left(\widehat{\boldsymbol{\Sigma}}_{sv}\right)$ returns $\widehat{\boldsymbol{\Sigma}}_{sv}$ with the elements that are less than λ_{sv} in absolute value set to zero.

OUTPUT: Return the Blockbuster covariance $\widehat{\boldsymbol{\Sigma}}_B$.

In case there are common factors present in the panel, as in the model in Definition 3, we may employ a regularisation approach similar to POET (Fan *et al.*, 2013). Let $\lambda_i(\widehat{\boldsymbol{\Sigma}})$ and $\boldsymbol{u}_i(\widehat{\boldsymbol{\Sigma}})$ denote the i-th eigenvalue and eigenvector of the sample covariance matrix.

Then we can regularise the sample covariance of the data using

$$\widehat{\mathbf{\Sigma}}_B = \sum_{i=n-R+1}^{n} \lambda_i(\widehat{\mathbf{\Sigma}})\boldsymbol{u}_i(\widehat{\mathbf{\Sigma}})\boldsymbol{u}_i'(\widehat{\mathbf{\Sigma}}) + \widehat{\mathbf{R}}_B,$$

where R is the number of factors and $\widehat{\mathbf{R}}_B$ is the Blockbuster covariance estimator applied to the residual matrix $\widehat{\mathbf{R}} = \sum_{i=1}^{n-R} \lambda_i(\widehat{\mathbf{\Sigma}})\boldsymbol{u}_i(\widehat{\mathbf{\Sigma}})\boldsymbol{u}_i'(\widehat{\mathbf{\Sigma}})$. Notice that we apply block-by-block thresholding to the off-diagonal blocks of the residual covariance matrix whereas the standard POET estimator applies universal thresholding to the off-diagonal elements. We use this estimator to validate the community partition estimated by Blockbuster out-of-sample in the empirical application.[9]

3 Simulation Study

In this section we investigate the finite-sample properties of the Blockbuster algorithm with a simulation study. We draw repeated samples of size T from an n-variate diagonal VAR(1) with coefficients drawn uniformly from the interval [0,0.9] and i.i.d. innovations drawn from a Gaussian SBPCM. In particular we rely on the SBPCM with common factors, with $R = 1$ factors and $k = 5$ communities. The edge probabilities are set to $p_s = p = c_p(\log n)^{1.01}/n$ for all $s = 1, \ldots, k$ and $q_{vr} = q = c_q(\log n)^{1.01}/n$ for all $v, r = 1, \ldots, k$, $v \neq r$. We calibrate c_p and c_q so that when $n = 100$ the (p, q) pair is equal to $(0.25, 0.01)$, $(0.25, 0.05)$ or $(0.50, 0.01)$. Note that as the edge probabilities are functions of n, varying n changes the probabilities. We draw $[\mathbf{\Theta}]_{ii}$ from a power law distribution $f(x) = \alpha x_m^\alpha / x^{\alpha+1}$ for $x \geq x_m$ with $x_m = 0.75$ and $\alpha = 2.5$ and the edge-weights w_{ij} from the interval $[0.3, 1]$ uniformly. The entries of the factor loading vector $\boldsymbol{q}$ are generated from a standard normal.

We apply the Blockbuster algorithm to each sample to recover the communities. To measure the quality of the Blockbuster partition $\widehat{\mathcal{V}}^k$ we compare it to $\mathcal{V}^k$ by calculating the hit ratio as the fraction of correctly classified vertices.[10] We repeat the Monte Carlo

[9] The properties of the estimator have not been studied and we leave this question for future research.

[10] Note that $\widehat{\mathcal{V}}^k$ only estimates $\mathcal{V}^k$ up to a permutation. As k is low in our simulations, we calculate the hit percentage for every possible permutation and select the maximum as the final hit ratio.

experiment 1 000 times for different values of n (50, 100 and 200) and T (50, 100, 200, 500 and 1 000). The results are summarised in Table 1. Panel A of Table 1 shows the results with $\phi = 5$, whereas panel B shows the results when $\phi = 50$. The first two rows of panel A in Table 1 show that Blockbuster performs quite well when $n = 50$. If the probability of within-community edges is high as in the second row, the algorithm has a 82% hit ratio even with a small sample of $T = 50$. Comparing the second to third rows, we see that results improve when p is large relative to q. It is also clear that the results with $n = 100$ and $n = 200$ are only better than those with $n = 50$ when the sample size is larger, as in the last column of Panel B. For larger n, more samples are required to consistently estimate the covariance matrix and thus the community partition as Theorem 2 suggests. Additional simulation evidence is reported in the Online Appendix. There we show that pre-filtering the time series using an AR(1) before applying Blockbuster improves detection and that Blockbuster performs satisfactorily in the presence of weak correlation between the factors and communities.

Table 1: Hit Ratio of Blockbuster

	Panel A: $\phi = 5$					Panel B: $\phi = 50$				
	$T = 50$	100	200	500	1 000	50	100	200	500	1 000
p/q	$n = 50$					$n = 50$				
0.25/0.01	65.5%	82.1%	90.6%	94.0%	94.7%	81.7%	91.2%	94.4%	94.6%	93.2%
0.50/0.01	82.4%	96.9%	99.7%	99.9%	100.0%	97.0%	99.9%	99.8%	99.9%	99.9%
0.25/0.05	40.9%	43.5%	50.5%	57.8%	61.8%	44.3%	48.0%	52.7%	62.7%	61.6%
	$n = 100$					$n = 100$				
0.25/0.01	49.0%	65.8%	83.0%	92.7%	94.6%	75.1%	88.1%	92.1%	95.0%	94.4%
0.50/0.01	63.6%	87.3%	97.6%	99.7%	99.9%	94.8%	98.9%	99.8%	99.9%	100.0%
0.25/0.05	33.8%	35.8%	38.1%	44.8%	52.5%	35.1%	38.6%	43.9%	53.0%	59.3%
	$n = 200$					$n = 200$				
0.25/0.01	37.2%	46.6%	67.4%	86.4%	94.6%	59.6%	80.3%	89.8%	93.9%	95.7%
0.50/0.01	46.1%	67.7%	88.0%	98.6%	99.7%	82.6%	94.7%	98.8%	99.9%	99.9%
0.25/0.05	29.2%	29.9%	31.1%	35.9%	42.9%	30.2%	31.5%	33.7%	41.6%	51.8%

The table reports hit ratios for Blockbuster algorithm. The probabilities in the rows correspond to $n = 100$.

4 Empirical Application

We apply our methodology to a panel of state-level real activity growth for the U.S. Our goal is to group the states into communities characterised by a high degree of

interdependence. We consider a dataset constructed by Hamilton and Owyang (2012). It consists of quarterly growth rates of payroll employment for the U.S. states (excluding Alaska and Hawaii) from the second quarter of 1956 to the fourth of 2007, which results in a panel of $n = 48$ time series over $T = 207$ periods. The data are seasonally adjusted and annualised.[11] See Hamilton and Owyang (2012) for further details.

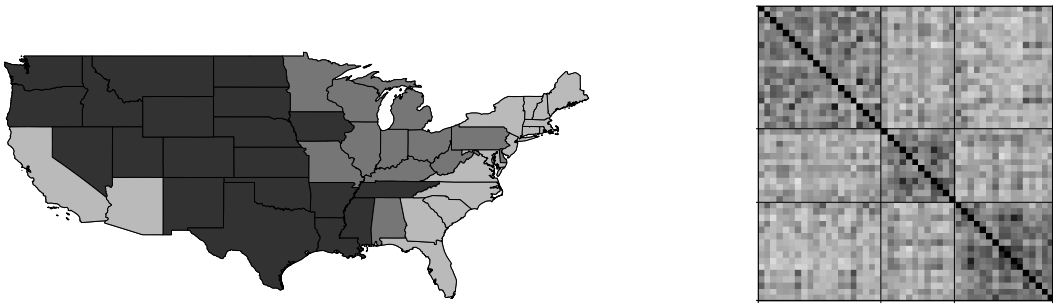
Hamilton and Owyang (2012) use their dataset to study business cycle synchronicity. We summarise their findings for the purpose of comparison, even if their research question differs from the present one. They introduce a Bayesian methodology based on a Markov-switching model to cluster the states into communities with similar business cycle timing. They also propose a cross-validation procedure to estimate the number of communities and find evidence of three clusters. We focus on the results where they use a set of exogenous state level characteristics in addition to the employment growth rate. They find (see right column of figure 3 in Hamilton and Owyang, 2012) that the states are partitioned, roughly speaking, into a cluster of oil-producing and agricultural states, a cluster containing several East Coast states together with California, Arizona and Colorado, and a cluster containing the remaining states. Note that our results are not directly comparable with theirs, as we provide point estimates while they provide community membership posterior probability. Moreover, they use more information than us as they also take advantage of exogenous explanatory variables. It is also worth pointing out that in their results some states have a low posterior probability of belonging to any cluster, whereas in our algorithm every state is assigned to a community.

We show the results of applying Blockbuster to the entire sample in Figure 2. A scree plot suggests one common factor, so we apply the extension of our algorithm with $R = 1$. The number of communities is set to three as in Hamilton and Owyang (2012). The left panel of the figure shows that our results bear interesting similarities to their findings. The dark grey community roughly matches the first cluster of Hamilton and Owyang (2012) and contains oil-producing and agricultural states. The light grey cluster contains East Coast states, California and Arizona, which roughly corresponds to their third

[11] We winsorise the data when the growth rates are larger than 20% in absolute value. There are only three such observations and they all belong to West Virginia. See also Hamilton and Owyang (2012).

cluster. Finally, the grey cluster contains Mid West states together with Alabama. Notice that the communities estimated by Blockbuster mostly form geographically contiguous areas even though no spatial information is given to the algorithm.[12] The right panel of Figure 2 shows a heatmap of the correlation matrix of the panel conditional on the factor where the series are ordered according to their community membership.[13] Series in the same community are positively correlated in the vast majority of cases and the intra-community correlation is larger than the inter-community correlation. We also calculate the proportion of variance explained by the principal components. The first principal component explains 50% of the total variance. The proportion of the variance explained by the principal components associated with the communities is sizeable and explains 16% of the total variation. We remark that the Online Appendix contains a number of robustness checks. In particular, it shows that results are relatively stable to the choice of k and the estimation window. Moreover, we show that an alternative community detection procedure based on applying spectral clustering to the partial correlation network estimate obtained through the GLASSO algorithm delivers similar results.

Figure 2: U.S. Real Activity Clustering ($k = 3$)



The figure displays the results of Blockbuster with $k = 3$. The left panel shows the communities on a map and the right panel a heatmap of the correlations (conditional on the factor) ordered by the Blockbuster partition. The top group in the heatmap corresponds to the dark grey community on the map, the middle to the grey and the bottom to the light grey.

We carry out an out-of-sample validation exercise to assess the performance of the community detection algorithm on the basis of the covariance regularisation procedure

[12] We also try setting $k = 2$ and $k = 4$, the results of which may be found in the Online Appendix.
[13] In the Online Appendix we report the average correlations by block.

from Algorithm 2. We split the sample in half and label the first half in-sample and the second half out-of-sample.[14] We estimate the Blockbuster covariance matrix in the in-sample period and then use it to predict the sample covariance matrix of the panel in the out-of-sample period. We apply the Blockbuster covariance estimator accounting for one factor. We measure the precision of the forecast on the basis of the Kullback-Leibler (KL) loss $L = \text{tr}\left(\widehat{\mathbf{\Sigma}}_C \mathbf{\Sigma}_B^{-1}\right) - \log\left(\left|\widehat{\mathbf{\Sigma}}_C \mathbf{\Sigma}_B^{-1}\right|\right) - n$, where $\widehat{\mathbf{\Sigma}}_C$ is a candidate estimator and $\mathbf{\Sigma}_B$ is a benchmark to compare against. We use the sample covariance matrix of the out-of-sample period as the benchmark. We run the procedure with the number of communities ranging from two to four. We compare the performance of the Blockbuster covariance estimator with the (standard) sample covariance, the POET covariance estimator (Fan *et al.*, 2013) and the (linear) Ledoit and Wolf (LW) shrinkage covariance estimator (Ledoit and Wolf, 2004). Table 2 reports the relative KL loss gains over the sample covariance, POET and LW estimators. The Blockbuster estimator performs favourably relative to these alternatives. The Online Appendix contains forecasting results using a set of alternative loss functions. Broadly speaking, the Blockbuster estimator performs satisfactorily but we do document some degree of heterogeneity in the performance depending on the loss function of interest.

Table 2: U.S. Real Activity Kullback-Leibler Losses

k	SCM	LW	POET
2	5.735%	13.436%	4.313%
3	3.127%	11.041%	1.521%
4	1.537%	9.581%	1.950%

The table reports the relative Kullback-Leibler loss gains of the Blockbuster covariance estimator over the sample covariance, the POET and the Ledoit and Wolf shrinkage estimators as a function of the number of communities k.

5 Conclusion

We introduce a class of partial correlation network models in which the underlying network structure is random and determined by a latent random graph with a community

[14] We do this to estimate the out-of-sample covariance matrix with sufficient accuracy. For example, a 75%/25% split would yield only 51 observations to estimate a 48-dimensional matrix.

structure. We then propose an algorithm called Blockbuster that uses the k-means clustering procedure on the top rescaled eigenvectors of the estimated covariance matrix to detect the community structure. We establish the consistency of the algorithm when the number of variables n and observations T are large. The methodology is applied to study real activity clustering in the U.S. using a dataset of quarterly state-level employment growth rates. Blockbuster detects a meaningful partition of the states that bears a close resemblance to previously published results. An out-of-sample validation exercise shows that covariance regularisation based on Blockbuster improves estimation accuracy.

References

Abbe, E., Bandeira, A. S., and Hall, G. (2016). Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, **62**, 471–487.

Amini, A. A., Chen, A., Bickel, P. J., and Levina, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics*, **41**, 2097–2122.

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

Anselin, L. (2007). Spatial econometrics. In *A Companion to Theoretical Econometrics*, chapter 14, pages 310–330. Wiley-Blackwell.

Arias-Castro, E. and Verzelen, N. (2014). Community detection in dense random networks. *The Annals of Statistics*, **42**, 940–969.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, **176**, 18–29.

Barigozzi, M. and Brownlees, C. (2019). NETS: Network estimation for time series. *Journal of Applied Econometrics*, **34**, 347–364.

Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, **36**, 2577–2604.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, **83**, 1147–1184.

Brownlees, C., Nualart, E., and Sun, Y. (2018). Realized networks. *Journal of Applied Econometrics*. Forthcoming.

Chudik, A., Pesaran, M. H., and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal*, **14**, C45–C90.

Chung, F. and Lu, L. (2006). *Complex Graphs and Networks*. American Mathematical Society, Providence.

de Paula, A. (2017). *Econometrics of Network Models*, volume 1 of *Econometric Society Monographs*, page 268–323. Cambridge University Press.

Demirer, M., Diebold, F. X., Liu, L., and Yilmaz, K. (2018). Estimating global bank network connectedness. *Journal of Applied Econometrics*, **33**, 1–15.

Dempster, A. P. (1972). Covariance selection. *Biometrics*, **28**, 157–175.

Diebold, F. X. and Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, **182**, 119–134.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large Covariance Estimation by Thresholding Principal Orthogonal Complements. *Journal of the Royal Statistical Society, Series B*, **75**, 603–680.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, **82**, 540–554.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, **486**, 75–174.

Francis, N., Owyang, M. T., and Savascin, O. (2017). An endogenously clustered factor approach to international business cycles. *Journal of Applied Econometrics*, **32**, 1261–1276.

Friedman, J., Hastie, T., and Tibshirani, R. (2011). Sparse Inverse Covariance Estimation with the Graphical Lasso. *Biostatistics*, **9**(3), 432–441.

Gudmundsson, G. S. (2018). Community detection in large vector autoregressions. Working paper.

Hallin, M. and Liska, R. (2011). Dynamic factors in the presence of blocks. *Journal of Econometrics*, **163**, 29–41.

Hamilton, J. D. and Owyang, M. T. (2012). The propagation of regional recessions. *The Review of Economics and Statistics*, **94**, 935–947.

Hautsch, N., Schaumburg, J., and Schienle, M. (2015). Financial network systemic risk contributions. *Review of Finance*, **19**, 685–738.

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

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

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

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics*, **186**, 325–344.

Ledoit, O. and Wolf, M. (2004). A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. *Journal of Multivariate Analysis*, **88**, 365–411.

Leiva-Leon, D. (2017). Measuring business cycles intra-synchronization in US: A regime-switching interdependence framework. *Oxford Bulletin of Economics and Statistics*, **79**, 513–545.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, **27**, 415–444.

Medeiros, M. C. and Mendes, E. F. (2016). ℓ_1-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors. *Journal of Econometrics*, **191**, 255–271.

Meinshausen, N. (2008). A note on the LASSO for Gaussian graphical model selection. *Statistics & Probability Letters*, **78**, 880–884.

Meinshausen, N. and Bühlmann, P. (2006). High dimensional graphs and variable selection with the LASSO. *The Annals of Statistics*, **34**, 1436–1462.

Merlevède, F., Peligrad, M., and Rio, E. (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, **151**, 435–474.

Ng, A. Y., Jordan, M. I., and Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, pages 849–856. MIT Press.

Oliveira, R. I. (2009). Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. Technical report.

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, **168**, 244–258.

Pourahmadi, M. (2013). *High-Dimensional Covariance Estimation*. John Wiley & Sons, Inc.

Qin, T. and Rohe, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, pages 3120–3128. Curran Associates Inc.

Rohe, K., Chatterjee, S., and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, **39**, 1878–1915.

Sarkar, P. and Bickel, P. J. (2015). Role of Normalization in Spectral Clustering for Stochastic Blockmodels. *The Annals of Statistics*, **43**, 962–990.

Su, L., Shi, Z., and Phillips, P. C. B. (2016). Identifying latent structures in panel data. *Econometrica*, **84**, 2215–2264.

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

A Proofs of Main Results

Proof of Lemma 1. See Lemmas 3.2 and 3.3 of Qin and Rohe (2013) and let $\tau = 0$. Notice that $\mathcal{K} = (1/\sigma^2)\mathbf{I}_n + (\phi/\sigma^2)\mathcal{L}$ implies that the eigenvectors of $\mathcal{K}$ and $\mathcal{L}$ are the same. □

Proof of Theorem 1. We rely on Theorems OA-1 and OA-2 that are stated and proved in the Online Appendix. As $\bar{d}_{min} = \Omega(n\rho_n)$, Theorem OA-1 gives us with high probability

$$\|\mathbf{K} - \mathcal{K}\| = O\left(\|\mathbf{L} - \mathcal{L}\|\right) = O\left(\sqrt{\frac{\log n}{n\rho_n}}\right), \tag{10}$$

and Theorem OA-2, provided that $T = \Omega\left(n^{2/\gamma - 1}\right)$, gives with high probability

$$\left\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\| = O\left(\sqrt{\frac{n}{T}}\|\mathbf{\Sigma}\|\right).$$

To obtain a statement about the precision matrix, note that we may write

$$\left\|\widehat{\mathbf{K}} - \mathbf{K}\right\| = \left\|\widehat{\mathbf{K}}\left(\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}\right)\mathbf{K}\right\| \leq \left\|\widehat{\mathbf{K}}\right\|\left\|\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}}\right\|\|\mathbf{K}\| = O\left(\sqrt{\frac{n}{T}}\right), \tag{11}$$

as $\|\mathbf{\Sigma}\|$ and $\|\mathbf{K}\|$ are bounded with high probability by (10), Weyl's inequality and Lemma 1 for large enough T. The triangle inequality along with (11) and (10) then delivers

$$\left\|\widehat{\mathbf{K}} - \mathcal{K}\right\| \leq \left\|\widehat{\mathbf{K}} - \mathbf{K}\right\| + \left\|\mathbf{K} - \mathcal{K}\right\| = O\left(\sqrt{\frac{n}{T}} + \sqrt{\frac{\log n}{n\rho_n}}\right). \tag{12}$$

We next apply the Davis-Kahan theorem to bound the angles between the subspaces of the bottom k eigenvectors of $\widehat{\mathbf{K}}$ and $\mathcal{K}$. We follow Sarkar and Bickel (2015) and the arguments of Appendix B of Rohe *et al.* (2011) closely to obtain with high probability

$$\left\|\widehat{\mathbf{U}} - \mathcal{U}\mathcal{O}\right\| = O\left(\sqrt{\frac{n}{T}} + \sqrt{\frac{\log n}{n\rho_n}}\right), \tag{13}$$

where $\mathcal{O} \equiv \mathbf{F}\mathbf{G}'$ is a $k \times k$ orthonormal rotation matrix based on the singular value decomposition $\mathcal{U}'\widehat{\mathbf{U}} = \mathbf{F}\mathbf{\Psi}\mathbf{G}'$. Notice that $\left|[\widehat{\mathbf{U}}]_{ji} - [\mathcal{U}]_{j\bullet}[\mathcal{O}]_{\bullet i}\right| \leq \left\|[\widehat{\mathbf{U}}]_{\bullet i} - \mathcal{U}[\mathcal{O}]_{\bullet i}\right\| \leq \left\|\widehat{\mathbf{U}} - \mathcal{U}\mathcal{O}\right\|$ for all $i, j = 1, \ldots, k$. As the rows of $\widehat{\mathbf{U}}$ and $\mathcal{U}$ have only k elements, we have

$$\left\|[\widehat{\mathbf{U}}]_{j\bullet} - [\mathcal{U}]_{j\bullet}\mathcal{O}\right\| = \sqrt{\sum_{i=1}^{k}\left([\widehat{\mathbf{U}}]_{ji} - [\mathcal{U}]_{j\bullet}[\mathcal{O}]_{\bullet i}\right)^2} = O\left(\left\|\widehat{\mathbf{U}} - \mathcal{U}\mathcal{O}\right\|\right),$$

which bounds the row norms of $\widehat{\mathbf{U}} - \mathcal{U}\mathcal{O}$. By the reverse triangle inequality, $[\widehat{\mathbf{N}}^{-1} - \mathcal{N}^{-1}]_{jj} = \left\|[\widehat{\mathbf{U}}]_{j\bullet}\right\| - \left\|[\mathcal{U}]_{j\bullet}\mathcal{O}\right\| \leq \left\|[\widehat{\mathbf{U}}]_{j\bullet} - [\mathcal{U}]_{j\bullet}\mathcal{O}\right\|$. Thus $\left\|\widehat{\mathbf{N}}^{-1} - \mathcal{N}^{-1}\right\| = O\left(\left\|\widehat{\mathbf{U}} - \mathcal{U}\mathcal{O}\right\|\right)$ as $\widehat{\mathbf{N}}^{-1} - \mathcal{N}^{-1}$ is a diagonal matrix. Furthermore $\left\|\widehat{\mathbf{N}} - \mathcal{N}\right\| \leq \left\|\widehat{\mathbf{N}}\right\|\left\|(\mathcal{N}^{-1} - \widehat{\mathbf{N}}^{-1})\right\|\|\mathcal{N}\| = O\left(\|\mathcal{N}\|\left\|\widehat{\mathbf{U}} - \mathcal{U}\mathcal{O}\right\|\right)$ with high probability, as $\left\|\widehat{\mathbf{N}}\right\|$ is close to $\|\mathcal{N}\|$ for large enough n. As $\|\mathcal{N}\| = 1/\min_j\|[\mathcal{U}]_{j\bullet}\|$ and row i of $\mathcal{U}$ has length

$$\|[\mathcal{U}]_{i\bullet}\| = \left(\frac{[\mathbf{\Theta}]_{ii}}{\sum_{j\in\mathcal{V}_s}[\mathbf{\Theta}]_{jj}}\right)^{1/2},$$

where s is the community that vertex i belongs to, we have $\|\mathcal{N}\| = O(\sqrt{n})$. It then follows with high probability that

$$\left\|\widehat{\mathbf{N}} - \mathcal{N}\right\| = O\left(\sqrt{n}\left\|\widehat{\mathbf{U}} - \mathcal{U}\mathcal{O}\right\|\right). \tag{14}$$

To put the pieces together, we may write

$$\left\|\widehat{\mathbf{X}} - \mathcal{X}\mathcal{O}\right\| = \left\|\widehat{\mathbf{N}}\widehat{\mathbf{U}} - \mathcal{N}\mathcal{U}\mathcal{O}\right\| \leq \left\|\widehat{\mathbf{N}} - \mathcal{N}\right\|\left\|\widehat{\mathbf{U}} - \mathcal{U}\mathcal{O}\right\| + \left\|\widehat{\mathbf{N}} - \mathcal{N}\right\|\|\mathcal{U}\mathcal{O}\| + \|\mathcal{N}\|\left\|\widehat{\mathbf{U}} - \mathcal{U}\mathcal{O}\right\|.$$

The cross term is lower order and $\|\mathcal{U}\mathcal{O}\|$ is a constant, as the matrix $\mathcal{U}\mathcal{O}$ has k columns of unit length. This taken together with (13) and (14), yields with high probability

$$\left\|\widehat{\mathbf{X}} - \mathcal{X}\mathcal{O}\right\| = O\left(\sqrt{n}\left\|\widehat{\mathbf{U}} - \mathcal{U}\mathcal{O}\right\|\right) = O\left(\frac{n}{\sqrt{T}} + \sqrt{\frac{\log n}{\rho_n}}\right).$$

□

Proof of Lemma 2. For the GSBM of Definition 1 we have $\mathcal{C} = \mathcal{X}\mathcal{O} = \mathbf{Z}\mathbf{V}\mathcal{O}$. It follows that $[\mathcal{C}]_{i\bullet} = [\mathbf{Z}]_{i\bullet}\mathbf{V}\mathcal{O}$ and we may thus equivalently define vertex i as correctly clustered if $[\widehat{\mathbf{C}}]_{i\bullet}$ is closer to $[\mathbf{Z}]_{i\bullet}\mathbf{V}\mathcal{O}$ than any other rotated population centroid $[\mathbf{Z}]_{j\bullet}\mathbf{V}\mathcal{O}$, $j \neq i$. The result then follows from the proof of Theorem 4.4 of Qin and Rohe (2013). See also the proof of Lemma 3.2 from Rohe *et al.* (2011). □

Proof of Theorem 2. We follow Rohe *et al.* (2011) closely. Note that as $\mathcal{C} \in \mathcal{M}(n,k)$, we have $\left\|\widehat{\mathbf{X}} - \widehat{\mathbf{C}}\right\|_F \leq \left\|\widehat{\mathbf{X}} - \mathcal{C}\right\|_F$, and by the triangle inequality we obtain

$$\left\|\widehat{\mathbf{C}} - \mathcal{C}\right\|_F \leq \left\|\widehat{\mathbf{C}} - \widehat{\mathbf{X}}\right\|_F + \left\|\widehat{\mathbf{X}} - \mathcal{C}\right\|_F \leq 2\left\|\widehat{\mathbf{X}} - \mathcal{C}\right\|_F.$$

Notice that $|\mathscr{M}| = \sum_{i\in\mathscr{M}} 1 \leq 2\sum_{i\in\mathscr{M}}\left\|[\widehat{\mathbf{C}}]_{i\bullet} - [\mathbf{Z}]_{i\bullet}\mathbf{V}\mathcal{O}\right\|^2 \leq 2\left\|\widehat{\mathbf{C}} - \mathcal{C}\right\|_F^2 \leq 8\left\|\widehat{\mathbf{X}} - \mathcal{X}\mathcal{O}\right\|_F^2$ where the second inequality follows from the fact that $\sum_{i=1}^n\left\|[\widehat{\mathbf{C}}]_{i\bullet} - [\mathbf{Z}]_{i\bullet}\mathbf{V}\mathcal{O}\right\|^2 = \left\|\widehat{\mathbf{C}} - \mathcal{C}\right\|_F^2$ and that the sum only includes a subset of the vertices. Theorem 1 implies

$$\left\|\widehat{\mathbf{X}} - \mathcal{X}\mathcal{O}\right\|_F = O\left(\frac{n}{\sqrt{T}} + \sqrt{\frac{\log n}{\rho_n}}\right),$$

as $\left\|\widehat{\mathbf{X}} - \mathcal{X}\mathcal{O}\right\|_F \leq \sqrt{k}\left\|\widehat{\mathbf{X}} - \mathcal{X}\mathcal{O}\right\|$ for a rank k matrix. The result follows. □

Proof of Theorem 3. We have

$$\left\|\mathbf{K}^{-1} - \mathcal{K}^{-1}\right\| = \left\|\mathbf{K}_{\boldsymbol{\epsilon}}^{-1} + \sum_{r=1}^R \boldsymbol{q}_r\boldsymbol{q}_r' - \left(\mathcal{K}_{\boldsymbol{\epsilon}}^{-1} + \sum_{r=1}^R \boldsymbol{q}_r\boldsymbol{q}_r'\right)\right\| = \left\|\mathbf{K}_{\boldsymbol{\epsilon}}^{-1} - \mathcal{K}_{\boldsymbol{\epsilon}}^{-1}\right\| = O\left(\sqrt{\frac{\log n}{n\rho_n}}\right),$$

by Theorem OA-1 and the fact hat $\|\mathcal{K}_{\boldsymbol{\epsilon}}^{-1}\|$ is bounded by Lemma 1. This implies $\|\mathbf{K} - \mathcal{K}\| = O\left(\sqrt{\frac{\log n}{n\rho_n}}\right)$, as $\|\mathcal{K}\|$ is bounded by Lemma OA-1. By Theorem OA-2 we obtain $\left\|\widehat{\mathbf{K}} - \mathbf{K}\right\| = O\left(\sqrt{\frac{n}{T}}\|\boldsymbol{\Sigma}\|\right)$, as $\|\mathbf{K}\|$ is bounded with high probability. Then we have

by the triangle inequality $\left\|\widehat{\mathbf{K}} - \mathcal{K}\right\| \leq \left\|\widehat{\mathbf{K}} - \mathbf{K}\right\| + \left\|\mathbf{K} - \mathcal{K}\right\| = O\left(\sqrt{\frac{n}{T}}\|\mathbf{\Sigma}\| + \sqrt{\frac{\log n}{n\rho_n}}\right)$. Similar arguments as in Theorems 1 and 2 then yield the result. □