Autoregressive Networks

Binyan Jiang
Department of Applied Mathematics, Hong Kong Polytechnic University
Hong Kong, by.jiang@polyu.edu.hk

Jialiang Li
Department of Statistics and Applied Probability, National University of Singapore
Singapore, stalj@nus.edu.sg

Qiwei Yao
Department of Statistics, London School of Economics, London, WC2A 2AE
United Kingdom, q.yao@lse.ac.uk

8 October 2020

Abstract

We propose a first-order autoregressive model for dynamic network processes in which edges change over time while nodes remain unchanged. The model depicts the dynamic changes explicitly. It also facilitates simple and efficient statistical inference such as the maximum likelihood estimators which are proved to be (uniformly) consistent and asymptotically normal. The model diagnostic checking can be carried out easily using a permutation test. The proposed model can apply to any Erdös-Renyi network processes with various underlying structures. As an illustration, an autoregressive stochastic block model has been investigated in depth, which characterizes the latent communities by the transition probabilities over time. This leads to a more effective spectral clustering algorithm for identifying the latent communities. Inference for a change-point is incorporated into the autoregressive stochastic block model to cater for possible structure changes. The developed asymptotic theory as well as the simulation study affirm the performance of the proposed methods. Application with three real data sets illustrates both relevance and usefulness of the proposed models.

Keywords: AR(1) networks; Change-point; Dynamic stochastic block model; Erdös-Renyi network; Hamming distance; Maximum likelihood estimation; Spectral clustering algorithm; Yule-Walker equation.

1 Introduction

Understanding and being able to model the network changes over time are of immense importance for, e.g., monitoring anomalies in internet traffic networks, predicting demand and setting pricing in electricity supply networks, managing natural resources in environmental readings in sensor networks, and understanding how news and opinion propagates in online social networks. Unfortunately most existing statistical inference methods for network data are confined to static networks, though a substantial proportion of real networks are dynamic in nature. In spite of the existence of a large body of literature on dynamic networks (see below), the development of the foundation for dynamic network models is still in its infancy, and the available modelling and inference tools are sparse (Kolaczyk, 2017). As for dealing with dynamic changes of networks, most available techniques are based on the evolution analysis of network snapshots over time without really modelling the changes directly (Aggarwal and Subbian, 2014; Donnat and Holmes, 2018). Although this reflects the fact that most networks change slowly over time, it provides little insight on the dynamics underlying the changes and is almost powerless for future prediction. Speedily increasing availability of large network data recorded over time also calls for more tools to reveal underlying dynamic structures more explicitly.

In this paper we propose a first-order autoregressive (i.e. AR(1)) model for dynamic network processes of which the edges changes over time while the nodes are unchanged. The autoregressive equation depicts the changes over time explicitly. The measures for the underlying dynamic structure such as autocorrelation coefficients, the Hamming distance can be explicitly evaluated. The AR(1) network model also facilitates the maximum likelihood estimation for the parameters in the model in a simple and direct manner. Some uniform error rates and the asymptotic normality for the maximum likelihood estimators are established with the number of nodes diverging. Model diagnostic checking can be easily performed in terms of a permutation test. Illustration with real network data indicates convincingly that the proposed AR(1) model is practically relevant and fruitful.

Our setting can apply to any Erdös-Renyi network processes with various underlying structures, which we illustrate through an AR(1) stochastic block model. With an explicitly defined autoregressive structure, the latent communities are characterized by the transition probabilities over time, instead of the (static) connection probabilities – the approach adopted from static stochastic block models but widely used in the existing literation on dynamic stochastic block models; see Pensky (2019) and the references therein. This new structure also paves the way for a new spectral clustering algorithm which identifies the latent communities more effectively – a phenomenon corroborated by both the asymptotic theory and the simulation results. To cater

for possible structure changes of underlying processes, we incorporate a change-point detection mechanism in the AR(1) stochastic block modeling. Again the change-point is estimated by the maximum likelihood method.

Theoretical developments for dynamic stochastic block models in literatures were typically based on the assumption that networks observed at different times are independent; see Pensky (2019); Bhattacharjee et al. (2020) and references therein. The autoregressive structure considered in this paper brings the extra complexity due to serial dependence. By establishing the α-mixing property with exponentially decaying coefficients for the AR(1) network processes, we are able to show that the proposed spectral clustering algorithm leads to a consistent recovery of the latent community structure. On the other hand, an extra challenge in detecting a change point in dynamic stochastic block network process is that the estimation for latent community structures before and after a possible change-point is typically not consistent during the search for the change-point. To overcome this obstacle, we introduce a truncation technique which breaks the searching interval into two parts such that the error bounds for the estimated change-point can be established.

The literature on dynamic network is large, across mathematics, computer science, engineer, statistics, biology, genetics and social sciences. We can only list a small selection of more statistics-oriented papers here. Fu et al. (2009) proposed a state space mixed membership stochastic block model (with a logistic normal prior). Hanneke et al. (2010) proposed an exponential random-graph model which specifies the conditional distribution of a network given its lagged values as a separable exponential function containing an implicit normalized constant. The inference for the model was conducted by an MCMC method. Krivitsky and Handcock (2014) further developed some separable exponential models. Durante et al. (2016) assumed that the elements of adjacency matrix at each time are conditionally independent Bernoulli random variables given two latent processes, and the conditional probabilities are the functions of those two processes. They further developed two so-called locally adaptive dynamic inference methods based on MCMC. Crane et al. (2016) studied the limit properties of Markovian, exchangeable and càdlàg (i.e. every edge remains in each state which it visits for a positive amount of time) dynamic network. Matias and Miele (2017) proposed a variational EM-algorithm for a dynamic stochastic block network model. Pensky (2019) studied the theoretical properties (such as the minimax lower bounds for the risk) of dynamic stochastic block model, assuming 'smooth' connectivity probabilities. The literature on change-point detection in dynamic networks include Yang et al. (2011); Wang et al. (2018); Wilson et al. (2019); Zhao et al. (2019); Bhattacharjee et al. (2020); Zhu et al. (2020a). While autoregressive models have been used in dynamic networks for modelling continuous responses observed from nodes (Zhu et al., 2017, 2019; Chen et al., 2020; Zhu et al., 2020b), to our best

knowledge no attempts have been made on modelling the dynamics of adjacency matrices in an autoregressive manner. Kang et al. (2017) uses dynamic network as a tool to model non-stationary vector autoregressive processes.

The rest of the paper is organized as follows. A general framework of AR(1) networks is laid out in Section 2. Theoretical properties of the model such as stationarity, Yule-Walker equation and Hamming distance are established. The maximum likelihood estimation for the parameters and the associated asymptotic properties are developed. Section 2 ends with an easy-to-use permutation test for the model diagnostics. Section 3 deals with AR(1) stochastic block models. The asymptotic theory is developed for the new spectral clustering algorithm based on the transition probabilities. Further extension of both the inference method and the asymptotic theory to the setting with a change-point is established. Simulation results are reported in Section 4, and the illustration with three real dynamic network data sets is presented in Section 5. All technical proofs are relegated to an Appendix.

2 Autoregressive network models

2.1 AR(1) models

We introduce a first-order autoregressive AR(1) dynamic network process $\{\mathbf{X}_t, t = 0, 1, 2, \cdots\}$ on the p nodes, denoted by $\{1, \cdots, p\}$, where $\mathbf{X}_t \equiv (X_{i,j}^t)$ denotes the $p \times p$ adjacency matrix at time t, and the p nodes are unchanged over time. We also assume that all networks are Erdös-Renyi in the sense that $X_{i,j}^t$, $(i,j) \in \mathcal{J}$, are independent and take values either 1 or 0, where $\mathcal{J} = \{(i,j) : 1 \le i \le j \le p\}$ for undirected networks, $\mathcal{J} = \{(i,j) : 1 \le i < j \le p\}$ for undirected networks without selfloops, $\mathcal{J} = \{(i,j) : 1 \le i, j \le p\}$ for directed networks, and $\mathcal{J} = \{(i,j) : 1 \le i \ne j \le p\}$ for directed networks without selfloops. Note that an edge from node i to j is indicated by $X_{i,j} = 1$, and no edge is denoted by $X_{i,j} = 0$. For undirected networks, $X_{i,j}^t = X_{j,i}^t$.

Definition 1. *For $t \ge 1$,*

$$X_{i,j}^t \;=\; X_{i,j}^{t-1}\, I(\varepsilon_{i,j}^t = 0) \;+\; I(\varepsilon_{i,j}^t = 1), \tag{2.1}$$

where $I(\cdot)$ denotes the indicator function, innovations $\varepsilon_{i,j}^t$, $(i,j) \in \mathcal{J}$, are independent, and

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

In the above expression, $\alpha_{i,j}^t, \beta_{i,j}^t$ are non-negative constants, and $\alpha_{i,j}^t + \beta_{i,j}^t \le 1$.

Equation (2.1) is an analogue of the noisy network model of Chang et al. (2020c). The innovation (or noise) $\varepsilon_{i,j}^t$ is 'added' via the two indicator functions to ensure that $X_{i,j}^t$ is still binary. Obviously $\{\mathbf{X}_t, t = 0, 1, 2, \cdots\}$ is a Markovian chain, and

$$P(X_{i,j}^t = 1 | X_{i,j}^{t-1} = 0) = \alpha_{i,j}^t, \quad P(X_{i,j}^t = 0 | X_{i,j}^{t-1} = 1) = \beta_{i,j}^t, \tag{2.3}$$

or collectively,

$$\begin{aligned} P(\mathbf{X}_t | \mathbf{X}_{t-1}, \cdots, \mathbf{X}_0) \;&=\; P(\mathbf{X}_t | \mathbf{X}_{t-1}) \;=\; \prod_{(i,j)\in\mathcal{J}} P(X_{i,j}^t | X_{i,j}^{t-1}) \qquad (2.4) \\ = \prod_{(i,j)\in\mathcal{J}} &(\alpha_{i,j}^t)^{X_{i,j}^t(1-X_{i,j}^{t-1})}(1-\alpha_{i,j}^t)^{(1-X_{i,j}^t)(1-X_{i,j}^{t-1})}(\beta_{i,j}^t)^{(1-X_{i,j}^t)X_{i,j}^{t-1}}(1-\beta_{i,j}^t)^{X_{i,j}^t X_{i,j}^{t-1}}. \end{aligned}$$

It is clear that the smaller $\alpha_{i,j}^t$ is, more likely no-edge status at time $t-1$ (i.e. $X_{i,j}^{t-1} = 0$) will be retained at time t (i.e. $X_{i,j}^t = 0$); and the smaller $\beta_{i,j}^t$ is, more likely an edge at time $t-1$ (i.e. $X_{i,j}^{t-1} = 1$) will be retained at time t (i.e. $X_{i,j}^t = 1$). For most slowly changing networks (such as social networks), we expect $\alpha_{i,j}^t$ and $\beta_{i,j}^t$ to be small.

2.2 Stationarity

Note that $\{\mathbf{X}_t\}$ is a homogeneous Markov chain if

$$\alpha_{i,j}^t \equiv \alpha_{i,j} \quad \text{and} \quad \beta_{i,j}^t \equiv \beta_{i,j} \quad \text{for all } t \geq 1 \text{ and } (i,j) \in \mathcal{J}. \tag{2.5}$$

Specify the distribution of the initial network $\mathbf{X}_0 = (X_{i,j}^0)$ as follows:

$$P(X_{i,j}^0 = 1) = \pi_{i,j} = 1 - P(X_{i,j}^0 = 0), \tag{2.6}$$

where $\pi_{i,j} \in (0,1)$, $(i,j) \in \mathcal{J}$, are constants. Theorem 1 below identifies the condition under which the network process $\{\mathbf{X}_t,\, t = 0, 1, 2, \cdots\}$ is strictly stationary.

Theorem 1. *Let the homogeneity condition (2.5) hold with $\alpha_{i,j} + \beta_{i,j} \in (0,1]$, and*

$$\pi_{i,j} = \alpha_{i,j}/(\alpha_{i,j} + \beta_{i,j}), \;\; (i,j) \in \mathcal{J}. \tag{2.7}$$

Then $\{\mathbf{X}_t, t = 0, 1, 2, \cdots\}$ is a strictly stationary process. Furthermore for any $(i,j), (\ell,m) \in \mathcal{J}$ and $t, s \geq 0$,

$$E(X_{i,j}^t) = \frac{\alpha_{i,j}}{\alpha_{i,j} + \beta_{i,j}}, \qquad \mathrm{Var}(X_{i,j}^t) = \frac{\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j} + \beta_{i,j})^2}, \tag{2.8}$$

$$\rho_{i,j}(|t-s|) \equiv \mathrm{Corr}(X_{i,j}^t, X_{\ell m}^s) = \begin{cases} (1-\alpha_{i,j}-\beta_{i,j})^{|t-s|} & \text{if } (i,j) = (\ell,m), \\ 0 & \text{otherwise.} \end{cases} \tag{2.9}$$

Remark 1. (i) The autocorrelation function (ACF) in (2.9) resembles that of the scalar AR(1) time series vividly: it is a power function of 'autoregressive coefficient' $(1-\alpha_{i,j}-\beta_{i,j}) = E\{I(\varepsilon_{i,j}^t = 0)\}$ (see (2.1) and (2.2)), and decays exponentially. In fact (2.1) implies a Yule-Walker equation

$$\gamma_{i,j}(k) = (1 - \alpha_{i,j} - \beta_{i,j})\gamma_{i,j}(k-1), \quad k = 1, 2, \cdots, \tag{2.10}$$

where $\gamma_{i,j}(k) = \mathrm{Cov}(X_{i,j}^{t+k}, X_{i,j}^t)$. (See Section A.1 in the Appendix for its proof.) It can also be seen from (2.4) that the smaller $\alpha_{i,j}$ and $\beta_{i,j}$ are, more likely that the network will retain its current status in the future, and, hence, the larger the autocorrelations are. When $\alpha_{i,j} + \beta_{i,j} = 1$, $\rho_{i,j}(k) = 0$ for all $k \neq 0$. Note then $X_{i,j}^t = I(\varepsilon_{i,j}^t = 1)$ carries no information on its lagged values.

(ii) As (2.7) can be written as $\pi_{i,j} = (1 + \beta_{i,j}/\alpha_{i,j})^{-1}$, the stationary marginal distribution of $\mathbf{X}_t$ depends on the ratios $\beta_{i,j}/\alpha_{i,j}$, $(i,j) \in \mathcal{J}$. Hence different stationary network processes may have the same marginal distribution.

Although the ACF in (2.9) admits very simple form, the implication of correlation coefficients for binary random variables is less clear. An alternative is to use the Hamming distance to measure the closeness of two networks, which counts the number of different edges in the two networks. See, for example, Donnat and Holmes (2018).

Definition 2. *For any two matrices* $\mathbf{A} = (A_{i,j})$ *and* $\mathbf{B} = (B_{i,j})$ *of the same size, the Hamming distance is defined as*

$$D_H(\mathbf{A}, \mathbf{B}) \ = \ \sum_{i,j} I(A_{i,j} \neq B_{i,j}).$$

Theorem 2. *Let* $\{\mathbf{X}_t, t = 0, 1, \cdots\}$ *be a stationary network process satisfying the condition of Theorem 1. Let* $d_H(|t-s|) = E\{D_H(\mathbf{X}_t, \mathbf{X}_s)\}$ *for any* $t, s \geq 0$. *Then* $d_H(0) = 0$, *and it holds for any* $k \geq 1$ *that*

$$d_H(k) \ = \ d_H(k-1) + \sum_{(i,j)\in\mathcal{J}} \frac{2\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j} + \beta_{i,j}}(1 - \alpha_{i,j} - \beta_{i,j})^{k-1} \tag{2.11}$$

$$= \sum_{(i,j)\in\mathcal{J}} \frac{2\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j} + \beta_{i,j})^2}\{1 - (1 - \alpha_{i,j} - \beta_{i,j})^k\}. \tag{2.12}$$

Theorem 2 indicates that the expected Hamming distance $d_H(d) = E\{D_H(\mathbf{X}_t, \mathbf{X}_{t+k})\}$ increases strictly, as k increases, initially from $d_H(1) = \sum \frac{2\alpha_{i,j}\beta_{i,j}}{\alpha_{i,j}+\beta_{i,j}}$ towards the limit $d_H(\infty) = \sum \frac{2\alpha_{i,j}\beta_{i,j}}{(\alpha_{i,j}+\beta_{i,j})^2}$ which is also the expected Hamming distance of the two independent networks sharing the same marginal distribution of $\mathbf{X}_t$.

Remark 2. The ACF of the process defined (2.1) is always non-negative (see (2.9)). This reflects, for example, the fact that in social networks 'friends' tend to remain as 'friends'. On the other

hand, an AR(1) process with alternating ACFs can be defined as

$$X_{i,j}^t = (1 - X_{i,j}^{t-1})I(\varepsilon_{i,j}^t = 0) + I(\varepsilon_{i,j}^t = 1),$$

where $\varepsilon_{i,j}^t$ are defined as in (2.2). Then under condition (2.5), $\{\mathbf{X}_t = (X_{i,j}^t),\, t = 0, 1, 2, \cdots\}$ is stationary if $P(X_{i,j}^0 = 1) = (1 - \beta_{i,j})/(2 - \alpha_{i,j} - \beta_{i,j})$. Furthermore, the ACF of $\{\mathbf{X}_t\}$ is

$$\mathrm{Corr}(X_{i,j}^t, X_{i,j}^{t+k}) = (-1)^k(1 - \alpha_{i,j} - \beta_{i,j})^k, \quad k = 0, 1, 2, \cdots,$$

and the expected Hamming distance is

$$E\{D_H(\mathbf{X}_t, \mathbf{X}_{t+k})\} = \sum_{(i,j)\in\mathcal{J}} \frac{2(1 - \alpha_{i,j})(1 - \beta_{i,j})}{(2 - \alpha_{i,j} - \beta_{i,j})^2}\{1 - (-1)^k(1 - \alpha_{i,j} - \beta_{i,j})^k\}, \quad k = 0, 1, 2, \cdots.$$

2.3 Estimation

To simplify the notation, we assume the availability of the observations $\mathbf{X}_0, \mathbf{X}_1, \cdots, \mathbf{X}_n$ from a stationary network process which satisfying the condition of Theorem 1. Without imposing any further structure on the model, the parameters $(\alpha_{i,j}, \beta_{i,j})$, for different (i, j), can be estimated separately.

The log-likelihood for $(\alpha_{i,j}, \beta_{i,j})$, conditionally on $\mathbf{X}_0$, is of the form

$$\begin{aligned} l(\alpha_{i,j}, \beta_{i,j}) = \log(\alpha_{i,j}) \sum_{t=1}^n X_{i,j}^t(1 - X_{i,j}^{t-1}) + \log(1 - \alpha_{i,j}) \sum_{t=1}^n (1 - X_{i,j}^t)(1 - X_{i,j}^{t-1}) \\ + \log(\beta_{i,j}) \sum_{t=1}^n (1 - X_{i,j}^t)X_{i,j}^{t-1} + \log(1 - \beta_{i,j}) \sum_{t=1}^n X_{i,j}^t X_{i,j}^{t-1}. \end{aligned}$$

See (2.4). Maximizing this log-likelihood leads to the maximum likelihood estimators

$$\widehat{\alpha}_{i,j} = \frac{\sum_{t=1}^n X_{i,j}^t(1 - X_{i,j}^{t-1})}{\sum_{t=1}^n (1 - X_{i,j}^{t-1})}, \qquad \widehat{\beta}_{i,j} = \frac{\sum_{t=1}^n (1 - X_{i,j}^t)X_{i,j}^{t-1}}{\sum_{t=1}^n X_{i,j}^{t-1}}. \tag{2.13}$$

To state the asymptotic properties, we list some regularity conditions first.

C1. There exists a constant l such that $0 < l \le \alpha_{i,j}, \beta_{i,j}, \alpha_{i,j} + \beta_{i,j} \le 1$ holds for all $(i, j) \in \mathcal{J}$.

C2. $n, p \to \infty$, and $(\log n)(\log\log n)\sqrt{\frac{\log p}{n}} \to 0$.

Condition C1 defines the parameter space, and Condition C2 indicates that the number of nodes is allowed to diverge in a smaller order than $\exp\big\{\frac{n}{(\log n)^2(\log\log n)^2}\big\}$. Relaxing condition C1 to include sparse networks is possible. We leave it to a follow-up study to keep the exploration in this paper as simple as possible.

Theorem 3. *Let conditions (2.5), C1 and C2 hold. Then it holds that*

$$\max_{(i,j)\in\mathcal{J}} |\widehat{\alpha}_{i,j} - \alpha_{i,j}| = O_p\left(\sqrt{\frac{\log p}{n}}\right) \quad \text{and} \quad \max_{(i,j)\in\mathcal{J}} |\widehat{\beta}_{i,j} - \beta_{i,j}| = O_p\left(\sqrt{\frac{\log p}{n}}\right).$$

Theorem 3 provides a uniform convergence rate for the MLEs in (2.13). Theorem 4 below establishes the joint asymptotic normality of any fixed numbers of the estimators. To this end, let $J_1 = \{(i_1, j_1), \ldots, (i_{m_1}, j_{m_1})\}$ and $J_2 = \{(k_1, \ell_1), \ldots, (k_{m_2}, \ell_{m_2})\}$ be two arbitrary subsets of $\mathcal{J}$ with $m_1, m_2 \geq 1$ fixed. Denote $\boldsymbol{\Theta}_{J_1,J_2} = (\alpha_{i_1,j_1}, \ldots, \alpha_{i_{m_1},j_{m_1}}, \beta_{k_1,\ell_1}, \ldots, \beta_{k_{m_2},\ell_{m_2}})^\top$ and correspondingly denote the MLEs as $\widehat{\boldsymbol{\Theta}}_{J_1,J_2} = (\widehat{\alpha}_{i_1,j_1}, \ldots, \widehat{\alpha}_{i_{m_1},j_{m_1}}, \widehat{\beta}_{k_1,\ell_1}, \ldots, \widehat{\beta}_{k_{m_2},\ell_{m_2}})^\top$.

Theorem 4. *Let conditions (2.5), C1 and C2 hold. Then*

$$\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{J_1,J_2} - \boldsymbol{\Theta}_{J_1,J_2}) \to N(\mathbf{0}, \boldsymbol{\Sigma}_{J_1,J_2}),$$

where $\boldsymbol{\Sigma}_{J_1,J_2} = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{m_1+m_2,m_1+m_2})$ *is a diagonal matrix with*

$$\sigma_{rr} = \frac{\alpha_{i_r,j_r}(1-\alpha_{i_r,j_r})(\alpha_{i_r,j_r}+\beta_{i_r,j_r})}{\beta_{i_r,j_r}}, \quad 1 \leq r \leq m_1,$$
$$\sigma_{rr} = \frac{\beta_{k_r,\ell_r}(1-\beta_{k_r,\ell_r})(\alpha_{k_r,\ell_r}+\beta_{k_r,\ell_r})}{\alpha_{k_r,\ell_r}}, \quad m_1+1 \leq r \leq m_1+m_2.$$

To prove Theorems 3 and 4, we need to show that $\{\mathbf{X}_t,\, t = 0, 1, \cdots\}$ is α-mixing with the exponentially decaying coefficients. The result is of independent interest and is presented as Lemma 1 below. It shows that the α-mixing coefficients are dominated by the autocorrelation coefficients (2.9). Note that the conventional mixing results for ARMA processes do not apply here, as they typically require that the innovation distribution is continuous; see, e.g., Section 2.6.1 of Fan and Yao (2003).

Let $\mathcal{F}_a^b$ be the σ-algebra generated by $\{X_{i,j}^k, a \leq k \leq b\}$. The α-mixing coefficient of process $\{X_{i,j}^t, t = 0, 1, \cdots\}$ is defined as

$$\alpha^{i,j}(\tau) = \sup_{k \in \mathbb{N}} \sup_{A \in \mathcal{F}_0^k, B \in \mathcal{F}_{k+\tau}^\infty,} |P(A \cap B) - P(A)P(B)|.$$

Lemma 1. *Let condition (2.5) hold,* $\alpha_{i,j}, \beta_{i,j} > 0$, *and* $\alpha_{i,j} + \beta_{i,j} \leq 1$. *Then* $\alpha^{i,j}(\tau) \leq \rho_{i,j}(\tau) = (1 - \alpha_{i,j} - \beta_{i,j})^\tau$ *for any* $\tau \geq 1$.

Note that under condition C1, $\alpha^{i,j}(\tau) \leq (1-l)^\tau$ for any $\tau \geq 1$ and $(i,j) \in \mathcal{J}$.

2.4 Model diagnostic check

Based on estimators in (2.13), we define 'residual' $\widehat{\varepsilon}_{i,j}^t$, resulted from fitting model (2.1) to the data, as the estimated value of $E(\varepsilon_{i,j}^t | X_{i,j}^t, X_{i,j}^{t-1})$, i.e.

$$\widehat{\varepsilon}_{i,j}^t = \frac{\widehat{\alpha}_{i,j}}{1-\widehat{\beta}_{i,j}} I(X_{i,j}^t = 1, X_{i,j}^{t-1} = 1) - \frac{\widehat{\beta}_{i,j}}{1-\widehat{\alpha}_{i,j}} I(X_{i,j}^t = 0, X_{i,j}^{t-1} = 0)$$
$$+ I(X_{i,j}^t = 1, X_{i,j}^{t-1} = 0) - I(X_{i,j}^t = 0, X_{i,j}^{t-1} = 1), \quad (i,j) \in \mathcal{J},\ t = 1, \cdots, n.$$

One way to check the adequacy of the model is to test for the independence of $\widehat{\boldsymbol{\mathcal{E}}}_t \equiv (\widehat{\varepsilon}_{i,j}^t)$ for $t = 1, \cdots, n$. Since $\widehat{\varepsilon}_{i,j}^t$, $t = 1, \cdots, n$, only take 4 different values for each $(i,j) \in \mathcal{J}$, we adopt the two-way, or three-way contingency table to test the independence of $\widehat{\boldsymbol{\mathcal{E}}}_t$ and $\widehat{\boldsymbol{\mathcal{E}}}_{t-1}$, or $\widehat{\boldsymbol{\mathcal{E}}}_t, \widehat{\boldsymbol{\mathcal{E}}}_{t-1}$ and $\widehat{\boldsymbol{\mathcal{E}}}_{t-2}$. For example the test statistic for the two-way contingency table is

$$T = \frac{1}{n\,|\mathcal{J}|} \sum_{(i,j)\in\mathcal{J}} \sum_{k,\ell=1}^{4} \{n_{i,j}(k,\ell) - n_{i,j}(k,\cdot) n_{i,j}(\cdot,\ell)/(n-1)\}^2 / \{n_{i,j}(k,\cdot) n_{i,j}(\cdot,\ell)/(n-1)\}, \quad (2.14)$$

where $|\mathcal{J}|$ denotes the cardinality of $\mathcal{J}$, and for $1 \le k, \ell \le 4$,

$$n_{i,j}(k,\ell) = \sum_{t=2}^{n} I\{\widehat{\varepsilon}_{i,j}^t = u_{i,j}(k),\ \widehat{\varepsilon}_{i,j}^{t-1} = u_{i,j}(\ell)\},$$
$$n_{i,j}(k,\cdot) = \sum_{t=2}^{n} I\{\widehat{\varepsilon}_{i,j}^t = u_{i,j}(k)\}, \quad n_{i,j}(\cdot,\ell) = \sum_{t=2}^{n} I\{\widehat{\varepsilon}_{i,j}^{t-1} = u_{i,j}(\ell)\}.$$

In the above expressions, $u_{i,j}(1) = -1, u_{i,j}(2) = -\frac{\widehat{\beta}_{i,j}}{1-\widehat{\alpha}_{i,j}}, u_{i,j}(3) = \frac{\widehat{\alpha}_{i,j}}{1-\widehat{\beta}_{i,j}}$ and $u_{i,j}(4) = 1$. We calculate the P-values of the test T based on the following permutation algorithm:

1. Permute $\widehat{\boldsymbol{\mathcal{E}}}_1, \cdots, \widehat{\boldsymbol{\mathcal{E}}}_n$ to obtain a new sequence $\boldsymbol{\mathcal{E}}_1^\star, \cdots, \boldsymbol{\mathcal{E}}_n^\star$. Calculate the test statistic $T^\star$ in the same manner as T with $\{\widehat{\boldsymbol{\mathcal{E}}}_t\}$ replaced by $\{\boldsymbol{\mathcal{E}}_t^\star\}$.

2. Repeat 1 above M times, obtaining permutation test statistics $T_j^\star$, $j = 1, \cdots, M$, where $M > 0$ is a large integer. The P-value of the test (for rejecting the stationary AR(1) model) is then
$$\frac{1}{M} \sum_{j=1}^{M} I(T < T_j^\star).$$

3 Autoregressive stochastic block models

The general setting in Section 2 may apply to various Erdös-Renyi network processes with some specific underlying structures. In this section we illustrate the idea with a new dynamic stochastic block (DSB) model.

3.1 Models

The DSB networks are undirected (i.e. $X_{i,j}^t \equiv X_{j,i}^t$) with no self-loops (i.e. $X_{i,i}^t \equiv 0$). Most available DSB models assume that the networks observed at different times are independent (Pensky, 2019; Bhattacharjee et al., 2020) or conditional independent (Xu and Hero, 2014; Durante et al., 2016; Matias and Miele, 2017) while connection probabilities and node memberships evolve over time. We take a radically different approach as we impose autoregressive structure (2.1) in the network process. Furthermore instead of assuming that the members in the same communities

share the same (unconditional) connection probabilities — a condition directly taken from static stochastic blocks models, we entertain the idea that the transition probabilities (2.3) for the members in the same communities are the same. This reflects more directly the dynamic behaviour of the process, and implies the former assumption on the unconditional connection probabilities under the stationarity. See (2.7). Furthermore, since the information on both $\alpha_{i,j}$ and $\beta_{i,j}$, instead of that on $\pi_{i,j} = \alpha_{i,j}/(\alpha_{i,j} + \beta_{i,j})$ only, will be used in estimation, we expect that the new approach leads to more efficient inference. This is confirmed by both the theory (Theorem 5 and also Remark 5 below) and the numerical experiments (Section 4.2 below).

Let ν_t be the membership function at time t, i.e. for any $1 \le i \le p$, $\nu_t(i)$ takes an integer value between 1 and q ($\le p$); indicating that node i belongs to the $\nu_t(i)$-th community at time t, where q is a fixed integer. This effectively assumes that the p nodes are divided into the q communities. We assume that q is fixed though some communities may contain no nodes at some times.

Definition 3. *An AR(1) stochastic block network process* $\{\mathbf{X}_t = (X^t_{i,j}),\ t = 0, 1, 2, \cdots\}$ *is defined by (2.1), where for* $1 \le i < j \le p$,

$$
\begin{aligned}
&P(\varepsilon^t_{i,j} = 1) = \alpha^t_{i,j} = \theta^t_{\nu_t(i),\nu_t(j)}, \quad P(\varepsilon^t_{i,j} = -1) = \beta^t_{i,j} = \eta^t_{\nu_t(i),\nu_t(j)}, \\
&P(\varepsilon^t_{i,j} = 0) = 1 - \alpha^t_{i,j} - \beta^t_{i,j} = 1 - \theta^t_{\nu_t(i),\nu_t(j)} - \eta^t_{\nu_t(i),\nu_t(j)}.
\end{aligned} \tag{3.1}
$$

In the above expressions, $\theta^t_{k,\ell}, \eta^t_{k,\ell}$ *are non-negative constants, and* $\theta^t_{k,\ell} + \eta^t_{k,\ell} \le 1$ *for all* $1 \le k \le \ell \le q$.

The evolution of membership process ν_t and/or the connection probabilities was often assumed to be driven by some latent (Markov) processes. The statistical inference for those models is carried out using computational Bayesian methods such as MCMC. See, for example, Yang et al. (2011); Xu and Hero (2014); Durante et al. (2016); Matias and Miele (2017). Bhattacharjee et al. (2020) adopted a change-point approach: assuming both the membership and the connection probabilities remain constants either before or after a change point. See also Ludkin et al. (2018); Wilson et al. (2019). This reflects the fact that many networks (e.g. social networks) hardly change, and a sudden change is typically triggered by some external events.

We adopt a change-point approach in this paper. Section 3.2 considers the estimation for both the community membership and transition probabilities when there are no change points in the process. This will serve as a building block for the inference with a change-point in Section 3.3. Note that detecting change-points in dynamic networks is a surging research area. More recent development include, in addition to the aforementioned references, Wang et al. (2018); Zhu et al. (2020a). Also note that the method of Zhao et al. (2019) can be applied to detect multiple change-points for any dynamic networks.

3.2 Estimation without change-points

We first consider a simple scenario of no change-points in the observed period, i.e.

$$\nu_t(\cdot) \equiv \nu(\cdot) \quad \text{and} \quad (\theta^t_{k,\ell}, \eta^t_{k,\ell}) \equiv (\theta_{k,\ell}, \eta_{k,\ell}), \quad t = 1, \cdots, n, \;\; 1 \le k \le \ell \le q. \tag{3.2}$$

Then fitting the DSB model consists of two steps: (i) estimating $\nu(\cdot)$ to cluster the p nodes into q communities, and (ii) estimating transition probabilities $\theta_{k,\ell}$ and $\eta_{k,\ell}$ for $1 \le k \le \ell \le q$. To simplify the presentation, we assume q is known, which is the assumption taken by most papers on change-point detection for DSB networks. In practice, one can determine q by, for example, the jittering method of Chang et al. (2020b), or a Bayesian information criterion; see an example in Section 5.2 below.

3.2.1 Why it works?

We first provide a theoretical underpinning (Proposition 1 below) on identifying the latent communities based on $\alpha_{i,j}$ and $\beta_{i,j}$. The stochastic block model with p nodes and q communities can be parameterized by a pair of matrices $(\mathbf{Z}, \mathbf{\Omega})$, where $\mathbf{Z} = (z_{i,j}) \in \{0,1\}^{p\times q}$ is the membership matrix such that it has exactly one 1 in each row and at least one 1 in each column, and $\mathbf{\Omega} = (\omega_{k,\ell})_{q\times q} \in [0,1]^{q\times q}$ is a symmetric and full rank connectivity matrix, with $\omega_{k,\ell} = \frac{\theta_{k,\ell}}{\theta_{k,\ell}+\eta_{k,\ell}}$. Then $z_{i,j} = 1$ if and only if the i-th node belongs to the j-th community. On the other hand, $\omega_{k,\ell}$ is the connection probability between the nodes in community k and the nodes in community ℓ, and $s_k \equiv \sum_{i=1}^p z_{i,k}$ is the size of community $k \in \{1, \ldots, q\}$. Clearly matrix $\mathbf{Z}$ and function $\nu(\cdot)$ are the two equivalent representations for the community membership of the network nodes. They are uniquely determined by each other.

Let $\mathbf{W} = \mathbf{Z\Omega Z}^\top$. Under model (3.2), the marginal edge formation probability is given as $E(\mathbf{X}_t) = \mathbf{W} - \text{diag}(\mathbf{W})$. Define

$$\begin{aligned} &\mathbf{\Omega}_1 = (\theta_{k,\ell})_{q\times q}, \qquad \mathbf{\Omega}_2 = (1-\eta_{k,\ell})_{q\times q}, \\ &\mathbf{W}_1 = \mathbf{Z\Omega}_1\mathbf{Z}^\top = (\alpha_{i,j})_{p\times p}, \qquad \mathbf{W}_2 = \mathbf{Z\Omega}_2\mathbf{Z}^\top = (1-\beta_{i,j})_{p\times p}, \end{aligned}$$

where $\alpha_{i,j} = \theta_{\nu(i),\nu(j)}$, $\beta_{i,j} = \eta_{\nu(i),\nu(j)}$. Then $\mathbf{W}_1 - \text{diag}(\mathbf{W}_1)$ can be viewed as the edge formation probability matrix of the latent noise process $\varepsilon^{1,t}_{i,j} \equiv I(\varepsilon^t_{i,j} = 1)$. Furthermore, under model (3.2), the latent network process $\{(\varepsilon^{1,t}_{i,j})_{1\le i,j\le p}, t = 0, 1, 2, \ldots\}$ has the same membership structures as $\{\mathbf{X}_t = (X^t_{i,j}),\, t = 0, 1, 2, \cdots\}$. Since $X^t_{i,j} = 1$ is implied by $\varepsilon^t_{i,j} = 1$ under model (2.1), the elements in $\mathbf{W}_1 - \text{diag}(\mathbf{W}_1)$ are thus positively correlated with the elements in $\mathbf{W} - \text{diag}(\mathbf{W})$. Similarly $\mathbf{W}_2 - \text{diag}(\mathbf{W}_2)$ can be viewed as the edge formation probability matrix of the latent noise process $\varepsilon^{2,t}_{i,j} = I(\varepsilon^t_{i,j} \ne -1)$, and $\{(\varepsilon^{2,t}_{i,j})_{1\le i,j\le p}, t = 0, 1, 2, \ldots\}$ has the same membership structures as $\{\mathbf{X}_t,\, t = 0, 1, 2, \cdots\}$. Since $\varepsilon^t_{i,j} = -1$ implies $X^t_{i,j} = 0$, the elements in $\mathbf{W}_2 - \text{diag}(\mathbf{W}_2)$ are also

positively correlated to those in $\mathbf{W} - \text{diag}(\mathbf{W})$. Let $\mathbf{D}_1$ and $\mathbf{D}_2$ be two $p \times p$ diagonal matrices with, respectively, $d_{i,1}, d_{i,2}$ as their (i,i)-th elements, where

$$d_{i,1} = \sum_{j=1}^{p} \alpha_{i,j}, \qquad d_{i,2} = \sum_{j=1}^{p} (1 - \beta_{i,j}).$$

The normalized Laplacian matrices based on $\mathbf{W}_1$ and $\mathbf{W}_2$ are then defined as:

$$\mathbf{L}_1 = \mathbf{D}_1^{-1/2} \mathbf{W}_1 \mathbf{D}_1^{-1/2}, \qquad \mathbf{L}_2 = \mathbf{D}_2^{-1/2} \mathbf{W}_2 \mathbf{D}_2^{-1/2}, \qquad \mathbf{L} = \mathbf{L}_1 + \mathbf{L}_2. \tag{3.3}$$

Correspondingly, we denote the degree corrected connectivity matrices as

$$\widetilde{\mathbf{\Omega}}_1 = \mathbf{D}_1^{-1/2} \mathbf{\Omega}_1 \mathbf{D}_1^{-1/2}, \qquad \widetilde{\mathbf{\Omega}}_2 = \mathbf{D}_2^{-1/2} \mathbf{\Omega}_2 \mathbf{D}_2^{-1/2}, \qquad \widetilde{\mathbf{\Omega}} = \widetilde{\mathbf{\Omega}}_1 + \widetilde{\mathbf{\Omega}}_2.$$

The following lemma shows that the block structure in the membership matrix $\mathbf{Z}$ can be recovered by the leading eigenvectors of $\mathbf{L}$.

Proposition 1. *Suppose $\widetilde{\mathbf{\Omega}}$ is full rank, and $rank(\mathbf{L}) = q$. Let $\mathbf{\Gamma}_q \mathbf{\Lambda} \mathbf{\Gamma}_q^\top$ be the eigen-decomposition of $\mathbf{L}$, where $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_q\}$ is the diagonal matrix consisting of the nonzero eigenvalues of $\mathbf{L}$ arranged in the order $|\lambda_1| \geq \cdots \geq |\lambda_q| > 0$. There exists a matrix $\mathbf{U} \in \mathcal{R}^{q \times q}$ such that $\mathbf{\Gamma}_q = \mathbf{Z}\mathbf{U}$. Further, for any $1 \leq i, j \leq p$, $\mathbf{z}_i \mathbf{U} = \mathbf{z}_j \mathbf{U}$ if and only if $\mathbf{z}_i = \mathbf{z}_j$, where $\mathbf{z}_i$ denotes the i-th row of $\mathbf{Z}$.*

Remark 3. The q columns of $\mathbf{\Gamma}_q$ are the orthonormal eigenvectors of $\mathbf{L}$ corresponding to the q non-zero eigenvalues. Proposition 1 implies that there are only q distinct rows in the $p \times q$ matrix $\mathbf{\Gamma}_q$, and two nodes belong to a same community if and only if the corresponding rows in $\mathbf{\Gamma}_q$ are the same. Intuitively the discriminant power of $\mathbf{\Gamma}_q$ can be understood as follows. For any unit vector $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_p)^\top$,

$$\boldsymbol{\gamma}^\top \mathbf{L} \boldsymbol{\gamma} = 2 - \sum_{1 \leq i < j \leq p} \alpha_{i,j} \Big(\frac{\gamma_i}{\sqrt{d_{i,1}}} - \frac{\gamma_j}{\sqrt{d_{j,1}}} \Big)^2 - \sum_{1 \leq i < j \leq p} (1 - \beta_{i,j}) \Big(\frac{\gamma_i}{\sqrt{d_{i,2}}} - \frac{\gamma_j}{\sqrt{d_{j,2}}} \Big)^2. \tag{3.4}$$

For $\boldsymbol{\gamma}$ being an eigenvector corresponding to the positive eigenvalue of $\mathbf{L}$, the sum of the 2nd and the 3rd terms on the RHS (3.4) is minimized. Thus $|\gamma_i - \gamma_j|$ is small when $\alpha_{i,j}$ and/or $(1 - \beta_{i,j})$ are large; noting that $d_{i,k} = d_{j,k}$ for $k = 1, 2$ when nodes i and j belong to the same community. The communities in a network are oftentimes formed in the way that the members within the same community are more likely to be connected with each other, and the members belong to different communities are unlikely or less likely to be connected. Hence when nodes i and j belong to the same community, $\alpha_{i,j}$ tends to be large and $\beta_{i,j}$ tends to be small (see (2.3)). The converse is true when the two nodes belong to two different communities. The eigenvectors corresponding to negative eigenvalues are capable to identify the so-called heterophilic communities, see pp.1892-3 of Rohe et al. (2011).

3.2.2 Estimating membership $\nu(\cdot)$

It follows from Theorem 1, (3.1) and (3.2) that

$$P(X_{ij}^t = 1) = \theta_{\nu(i),\nu(j)}/(\theta_{\nu(i),\nu(j)} + \eta_{\nu(i),\nu(j)}) \equiv \omega_{\nu(i),\nu(j)}, \quad 1 \le i < j \le p,$$

provided that X_{ij}^0 is initiated with the same marginal distribution. The conventional approach adopted in literature is to apply a community detection method for static stochastic block models using the averaged data $\bar{\mathbf{X}} = \sum_{1\le t\le n} \mathbf{X}_t/n$ to detect the latent communities characterized by the connection probabilities $\{\omega_{k,\ell}, 1 \le k \le \ell \le q\}$. We take a different approach based on estimators $\{(\widehat{\alpha}_{i,j}, \widehat{\beta}_{i,j}), 1 \le i < j \le p\}$ defined in (2.13) to identify the clusters determined by the transition probabilities $\{(\theta_{k,\ell}, \eta_{k,\ell}), 1 \le k \le \ell \le q\}$ instead. More precisely, we propose a new spectral clustering algorithm to estimate $\mathbf{\Gamma}_q$ specified in Proposition 1 above.

Let $\widehat{\mathbf{W}}_1, \widehat{\mathbf{W}}_2$ be two $p\times p$ matrices with, respectively, $\widehat{\alpha}_{i,j}, (1-\widehat{\beta}_{i,j})$ as their (i,j)-th elements for $i \ne j$, and 0 on the main diagonals. Let $\widehat{\mathbf{D}}_1, \widehat{\mathbf{D}}_2$ be two $p \times p$ diagonal matrices with, respectively, $\widehat{d}_{i,1}, \widehat{d}_{i,2}$ as their (i,i)-th elements, where

$$\widehat{d}_{i,1} = \sum_{j=1}^{p} \widehat{\alpha}_{i,j}, \qquad \widehat{d}_{i,2} = \sum_{j=1}^{p} (1 - \widehat{\beta}_{i,j}).$$

Define two (normalized) Laplacian matrices

$$\widehat{\mathbf{L}}_1 = \widehat{\mathbf{D}}_1^{-1/2} \widehat{\mathbf{W}}_1 \widehat{\mathbf{D}}_1^{-1/2}, \qquad \widehat{\mathbf{L}}_2 = \widehat{\mathbf{D}}_2^{-1/2} \widehat{\mathbf{W}}_2 \widehat{\mathbf{D}}_2^{-1/2}. \tag{3.5}$$

Perform the eigen-decomposition for the sum of $\mathbf{L}_1$ and $\mathbf{L}_2$:

$$\widehat{\mathbf{L}} \equiv \widehat{\mathbf{L}}_1 + \widehat{\mathbf{L}}_2 = \widehat{\mathbf{\Gamma}}\, \mathrm{diag}(\widehat{\lambda}_1, \cdots, \widehat{\lambda}_p) \widehat{\mathbf{\Gamma}}^\top, \tag{3.6}$$

where the eigenvalues are arranged in the order $\widehat{\lambda}_1^2 \ge \ldots \ge \widehat{\lambda}_p^2$, and the columns of the $p \times p$ orthogonal matrix $\widehat{\mathbf{\Gamma}}$ are the corresponding eigenvectors. We call $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_q$ the q leading eigenvalues of $\widehat{\mathbf{L}}$. Denote by $\widehat{\mathbf{\Gamma}}_q$ the $p \times q$ matrix consisting of the first q columns of $\widehat{\mathbf{\Gamma}}$, which are called the leading eigenvectors of $\widehat{\mathbf{L}}$. The spectral clustering applies the k-means clustering algorithm to the p rows of $\widehat{\mathbf{\Gamma}}_q$ to obtain the community assignments for the p nodes $\widehat{\nu}(i) \in \{1, \cdots, q\}$ for $i = 1, \cdots, p$.

Remark 4. Proposition 1 implies that the true memberships can be recovered by the q distinct rows of $\mathbf{\Gamma}_q$. Note that

$$\widehat{\mathbf{L}} = \widehat{\mathbf{L}}_1 + \widehat{\mathbf{L}}_2 \approx \mathbf{L}_1 - \mathrm{diag}(\mathbf{L}_1) + \mathbf{L}_2 - \mathrm{diag}(\mathbf{L}_2) = \mathbf{L} - \mathrm{diag}(\mathbf{L}).$$

We shall see that the effect of the term $\mathrm{diag}(\mathbf{L})$ on the eigenvectors $\mathbf{\Gamma}_q$ is negligible when p is large (see for example (A.13) in the proof of Lemma 6 in the Appendix), and hence the rows of $\widehat{\mathbf{\Gamma}}_q$ should be slightly perturbed versions of the q distinct rows in $\mathbf{\Gamma}_q$.

The following theorem justified the validity of using $\widehat{\mathbf{L}}$ for spectral clustering. Note that $\|\cdot\|_2$ and $\|\cdot\|_F$ denote, respectively, the L_2 and the Frobenius norm of matrices.

Theorem 5. *Let conditions (2.5), C1 and C2 hold, and* $\lambda_q^{-2}\left(\sqrt{\frac{\log(pn)}{np}}+\frac{1}{n}+\frac{1}{p}\right)\to 0$, *as* $n,p\to\infty$. *Then it holds that*

$$\max_{i=1,\ldots,p}|\lambda_i^2-\widehat{\lambda}_i^2|\le\|\widehat{\mathbf{L}}\widehat{\mathbf{L}}-\mathbf{L}\mathbf{L}\|_2\le\|\widehat{\mathbf{L}}\widehat{\mathbf{L}}-\mathbf{L}\mathbf{L}\|_F=O_p\left(\sqrt{\frac{\log(pn)}{np}}+\frac{1}{n}+\frac{1}{p}\right). \tag{3.7}$$

Moreover, for any constant $B>0$, *there exists a constant* $C>0$ *such that the inequality*

$$\|\widehat{\mathbf{\Gamma}}_q-\mathbf{\Gamma}_q\mathbf{O}_q\|_F\le 4\lambda_q^{-2}C\left(\sqrt{\frac{\log(pn)}{np}}+\frac{1}{n}+\frac{1}{p}\right) \tag{3.8}$$

holds with probability greater than $1-16p\left[(pn)^{-(1+B)}+\exp\{-B\sqrt{p}\}\right]$, *where* $\mathbf{O}_q$ *is a* $q\times q$ *orthogonal matrix.*

It follows from (3.7) that the leading eigenvalues of $\mathbf{L}$ can be consistently recovered by the leading eigenvalues of $\widehat{\mathbf{L}}$. By (3.8), the leading eigenvectors of $\mathbf{L}$ can also be consistently estimated, subject to a rotation (due to the possible multiplicity of some leading eigenvalues $\mathbf{L}$). Proposition 1 indicates that there are only q distinct rows in $\mathbf{\Gamma}_q$, and, therefore, also q distinct rows in $\mathbf{\Gamma}_q\mathbf{O}_q$, corresponding to the q latent communities for the p nodes. This paves the way for the k-means algorithm stated below. Put

$$\mathcal{M}_{p,q}=\{\mathbf{M}\in\mathcal{R}^{p\times q}:\mathbf{M}\text{ has }q\text{ distinct rows}\}.$$

The k-means clustering algorithm: Let

$$(\widehat{\mathbf{c}}_1,\cdots,\widehat{\mathbf{c}}_p)^\top=\arg\min_{\mathbf{M}\in\mathcal{M}_{p,q}}\|\widehat{\mathbf{\Gamma}}_q-\mathbf{M}\|_F^2.$$

There are only q distinct vectors among $\widehat{\mathbf{c}}_1,\cdots,\widehat{\mathbf{c}}_p$, forming the q communities. Theorem 6 below shows that they are identical to the latent communities of the p nodes under (3.8) and (3.9). The latter holds if $\sqrt{s_{\max}}\lambda_q^{-2}C\left(\sqrt{\frac{\log(pn)}{np}}+\frac{1}{n}+\frac{1}{p}\right)\to 0$, where $s_{\max}=\max\{s_1,\ldots,s_q\}$ is the size of the largest community.

Theorem 6. *Let* (3.8) *hold and*

$$\sqrt{\frac{1}{s_{\max}}}>2\sqrt{6}\lambda_q^{-2}C\left(\sqrt{\frac{\log(pn)}{np}}+\frac{1}{n}+\frac{1}{p}\right). \tag{3.9}$$

Then $\widehat{\mathbf{c}}_i=\widehat{\mathbf{c}}_j$ *if and only if* $\nu(i)=\nu(j)$, $1\le i,j\le p$.

Remark 5. By Lemma A.1 of Rohe et al. (2011), the error bound for the standard spectral clustering algorithm (with $n=1$) is $O_p\left(\frac{\log p}{\sqrt{p}}+\frac{1}{p}\right)$, where the term $\frac{1}{p}$ reflects the bias caused by

the inconsistent estimation of diagonal terms (see equation (A.5) and subsequent derivations in Rohe et al. (2011)). This bias comes directly from the removal of the diagonal elements of $\mathbf{L}$, as pointed out in Remark 4 above. Although the algorithm was designed for static networks, it has often be applied to dynamic networks using $\frac{1}{n}\sum_t \mathbf{X}_t$ in the place of the single observed network; see, e.g. Bhattacharjee et al. (2020). With some simple modification to the proof of Lemma A.1 of Rohe et al. (2011), it can be shown that the error bound is then reduced to

$$O_p\left(\frac{\log(pn)}{\sqrt{np}} + \frac{1}{p}\right) \tag{3.10}$$

provided that the observed networks are i.i.d. The error would only increase when the observations are not independent. On the other hand, our proposed spectral clustering algorithm for (dependent) dynamic networks entails the error rate specified in (3.7) and (3.8) which is smaller than (3.10) as long as n is sufficiently large (i.e. $(p/n)^{\frac{1}{2}}/\log(np) \to 0$). Note that we need n to be large enough in relation to p in order to capture the dynamic dependence of the networks.

3.2.3 Estimation for $\theta_{k,\ell}$ and $\eta_{k,\ell}$

For any $1 \le k \le \ell \le q$, we define

$$S_{k,l} = \begin{cases} \{(i,j) : 1 \le i \ne j \le p, \nu(i) = k, \nu(j) = \ell\} & \text{if } k \ne l, \\ \{(i,j) : 1 \le i < j \le p, \nu(i) = k = \nu(j) = \ell\} & \text{if } k = l, \end{cases} \tag{3.11}$$

Clearly the cardinality of $S_{k,\ell}$ is $n_{k,\ell} = s_k s_\ell$ when $k \ne \ell$ and $n_{k,\ell} = s_k(s_k - 1)/2$ when $k = \ell$.

Based on the procedure presented in Section 3.2.2, we obtain an estimated membership function $\widehat{\nu}(\cdot)$. Consequently, the MLEs for $(\theta_{k,\ell}, \eta_{k,\ell})$, $1 \le k \le \ell \le q$, admit the form

$$\widehat{\theta}_{k,\ell} = \sum_{(i,j)\in\widehat{S}_{k,\ell}} \sum_{t=1}^{n} X_{i,j}^t (1 - X_{i,j}^{t-1}) \Big/ \sum_{(i,j)\in\widehat{S}_{k,\ell}} \sum_{t=1}^{n} (1 - X_{i,j}^{t-1}), \tag{3.12}$$

$$\widehat{\eta}_{k,\ell} = \sum_{(i,j)\in\widehat{S}_{k,\ell}} \sum_{t=1}^{n} (1 - X_{i,j}^t) X_{i,j}^{t-1} \Big/ \sum_{(i,j)\in\widehat{S}_{k,\ell}} \sum_{t=1}^{n} X_{i,j}^{t-1}, \tag{3.13}$$

where

$$\widehat{S}_{k,\ell} = \begin{cases} \{(i,j) : 1 \le i \ne j \le p,\ \widehat{\nu}(i) = k,\ \widehat{\nu}(j) = \ell\} & \text{if } k \ne \ell, \\ \{(i,j) : 1 \le i < j \le p,\ \widehat{\nu}(i) = \widehat{\nu}(j) = k\} & \text{if } k = \ell. \end{cases}$$

See (2.13) and also (3.1).

Theorem 6 implies that the memberships of the nodes can be consistently recovered. Consequently, the consistency and the asymptotic normality of the MLEs $\widehat{\theta}_{k,\ell}$ and $\widehat{\eta}_{k,\ell}$ can be established in the same manner as for Theorems 3 and 4. We state the results below.

Let $\mathcal{K}_1 = \{(i_1, j_1), \ldots, (i_{m_1}, j_{m_1})\}$ and $\mathcal{K}_2 = \{(k_1, \ell_1), \ldots, (k_{m_2}, \ell_{m_2})\}$ be two arbitrary subsets of $\{(k, \ell) : 1 \le k \le \ell \le q\}$ with $m_1, m_2 \ge 1$ fixed. Let

$$\mathbf{\Psi}_{\mathcal{K}_1,\mathcal{K}_2} = (\theta_{i_1,j_1}, \ldots, \theta_{i_{m_1},j_{m_1}}, \eta_{k_1,\ell_1}, \ldots, \eta_{k_{m_2},\ell_{m_2}})',$$

and let $\widehat{\mathbf{\Psi}}_{\mathcal{K}_1,\mathcal{K}_2}$ denote its MLE. Put $\mathbf{N}_{\mathcal{K}_1,\mathcal{K}_2} = \mathrm{diag}(n_{i_1,j_1}, \ldots, n_{i_{m_1},j_{m_1}}, n_{k_1,\ell_1}, \ldots, n_{k_{m_2},\ell_{m_2}})$ where $n_{k,\ell}$ is the cardinality of $S_{k,\ell}$ defined as in (3.11).

Theorem 7. *Let conditions (2.5), C1 and C2 hold, and* $\sqrt{s_{\max}}\lambda_q^{-2} C\left(\sqrt{\frac{\log(pn)}{np}} + \frac{1}{n} + \frac{1}{p}\right) \to 0$. *Then it holds that*

$$\max_{1\le k,\ell\le q} |\widehat{\theta}_{k,\ell} - \theta_{k,\ell}| = O_p\left(\sqrt{\frac{\log q}{n s_{\min}^2}}\right) \quad \text{and} \quad \max_{1\le k,\ell\le q} |\widehat{\eta}_{k,\ell} - \eta_{k,\ell}| = O_p\left(\sqrt{\frac{\log q}{n s_{\min}^2}}\right),$$

where $s_{\min} = \min\{s_1, \ldots, s_q\}$.

Theorem 8. *Let the condition of Theorem 7 holds. Then*

$$\sqrt{n}\mathbf{N}_{\mathcal{K}_1,\mathcal{K}_2}^{\frac{1}{2}}(\widehat{\mathbf{\Psi}}_{\mathcal{K}_1,\mathcal{K}_2} - \mathbf{\Psi}_{\mathcal{K}_1,\mathcal{K}_2}) \to N(\mathbf{0}, \widetilde{\mathbf{\Sigma}}_{\mathcal{K}_1,\mathcal{K}_2}),$$

where $\widetilde{\mathbf{\Sigma}}_{\mathcal{K}_1,\mathcal{K}_2} = \mathrm{diag}(\widetilde{\sigma}_{11}, \ldots, \widetilde{\sigma}_{m_1+m_2,m_1+m_2})$ *with*

$$\widetilde{\sigma}_{rr} = \frac{\theta_{i_r,j_r}(1-\theta_{i_r,j_r})(\theta_{i_r,j_r} + \eta_{i_r,j_r})}{\eta_{i_r,j_r}}, \quad 1 \le r \le m_1,$$
$$\widetilde{\sigma}_{rr} = \frac{\eta_{k_r,\ell_r}(1-\eta_{k_r,\ell_r})(\theta_{k_r,\ell_r} + \eta_{k_r,\ell_r})}{\theta_{k_r,\ell_r}}, \quad m_1 + 1 \le r \le m_1 + m_2.$$

Finally to prepare for the inference in Section 3.3 below, we introduce some notation. First we denote $\widehat{\nu}$ by $\widehat{\nu}^{1,n}$, to reflect the fact that the community clustering was carried out using the data $\mathbf{X}_1, \cdots, \mathbf{X}_n$ (conditionally on $\mathbf{X}_0$). See Section 3.2.2 above. Further we denote the maximum log likelihood by

$$\widehat{l}(1, n;\ \widehat{\nu}^{1,n}) = l(\{\widehat{\theta}_{k,\ell}, \widehat{\eta}_{k,\ell}\};\ \widehat{\nu}^{1,n}) \tag{3.14}$$

to highlight the fact that both the node clustering and the estimation for transition probabilities are based on the data $\mathbf{X}_1, \cdots, \mathbf{X}_n$.

3.3 Inference with a change-point

Now we assume that there is a change-point τ_0 at which both the membership of nodes and the transition probabilities $\{\theta_{k,\ell}, \eta_{k,\ell}\}$ change. It is necessary to assume $n_0 \le \tau_0 \le n - n_0$, where n_0 is an integer and $n_0/n \equiv c_0 > 0$ is a small constant, as we need enough information before and after the change in order to detect τ_0. We assume that within the time period $[0, \tau_0]$, the network

follows a stationary model (3.2) with parameters $\{(\theta_{1,k,\ell}, \eta_{1,k,\ell}) : 1 \le k, l \le q\}$ and a membership map $\nu^{1,\tau_0}(\cdot)$. Within the time period $[\tau_0 + 1, n]$ the network follows a stationary model (3.2) with parameters $\{(\theta_{2,k,\ell}, \eta_{2,k,\ell}) : 1 \le k, l \le q\}$ and a membership map $\nu^{\tau_0+1,n}(\cdot)$. Though we assume that the number of communities is unchanged after the change point, our results can be easily extended to the case that the number of communities also changes.

We estimate the change point τ_0 by the maximum likelihood method:

$$\widehat{\tau} = \arg \max_{n_0 \le \tau \le n - n_0} \{ \widehat{l}(1, \tau;\ \widehat{\nu}^{1,\tau}) + \widehat{l}(\tau + 1, n;\ \widehat{\nu}^{\tau+1,n}) \}, \tag{3.15}$$

where $\widehat{l}(\cdot)$ is given in (3.14).

To measure the difference between the two sets of transition probabilities before and after the change, we put

$$\Delta_F^2 = \frac{1}{p^2} \big(\|\mathbf{W}_{1,1} - \mathbf{W}_{2,1}\|_F^2 + \|\mathbf{W}_{1,2} - \mathbf{W}_{2,2}\|_F^2 \big),$$

where the four $p \times p$ matrices are defined as

$$\mathbf{W}_{1,1} = (\theta_{1,\nu^{1,\tau_0}(i),\nu^{1,\tau_0}(j)}), \quad \mathbf{W}_{1,2} = (1 - \eta_{1,\nu^{1,\tau_0}(i),\nu^{1,\tau_0}(j)}),$$

$$\mathbf{W}_{2,1} = (\theta_{2,\nu^{\tau_0+1,n}(i),\nu^{\tau_0}(j)+1,n}), \quad \mathbf{W}_{2,2} = (1 - \eta_{1,\nu^{\tau_0+1,n}(i),\nu^{\tau_0+1,n}(j)}).$$

Note that Δ_F can be viewed as the signal strength for detecting the change-point τ_0. Let $s_{\max}$, $s_{\min}$ denote, respectively, the largest, the smallest community size among all the communities before and after the change. Similar to (3.3), we denote the normalized Laplacian matrices corresponding to on $\mathbf{W}_{i,j}$ as $\mathbf{L}_{i,j}$ for $i, j = 1, 2$. Let $|\lambda_{i,1}| \ge |\lambda_{i,2}| \ge \ldots \ge |\lambda_{i,q}|$ the absolute nonzero eigenvalues of $\mathbf{L}_{i,1} + \mathbf{L}_{i,2}$ for $i = 1, 2$, and we denote $\lambda_{\min} = \min\{|\lambda_{1,q}|, |\lambda_{2,q}|\}$. Now some regularity conditions are in order.

C3. For some constant $l > 0$, $\theta_{i,k,\ell}, \eta_{i,k,\ell} > l$, and $\theta_{i,k,\ell} + \eta_{i,k,\ell} \le 1$ for all $i = 1, 2$ and $1 \le k \le \ell \le q$.

C4. $\log(np)/\sqrt{p} \to 0$, and $\sqrt{s_{\max}} \lambda_{\min}^{-2} \big(\sqrt{\log(pn)/np} + \frac{1}{n} + \frac{1}{p} + \frac{\log(np)/n + \sqrt{\log(np)/(np^2)}}{\Delta_F^2} \big) \to 0$.

C5. $\frac{\Delta_F^2}{\log(np)/n + \sqrt{\log(np)/(np^2)}} \to \infty$.

Condition C3 is similar to C1. The condition $\log(np)/\sqrt{p} \to 0$ in C4 controls the misclassification rate of the k-means algorithm. Recall that there is a bias term $O(p^{-1})$ in spectral clustering caused by the removal of the diagonal of the Laplacian matrix (see Remark 4 above). Intuitively, as p increases, the effect of this bias term on the misclassification rate of the k-means algorithm becomes negligible. On the other hand, note that the length of the time interval for searching for the change point in (3.15) is of order $O(n)$, the $\log(n)$ term here in some sense reflects the

effect of the difficulty in detecting the true change point when the searching interval is extended as n increases. The second condition in C4 is similar to (3.9), which ensures that the true communities can be recovered. Condition C5 requires that the average signal strength $\Delta_F^2 = p^{-2}\big[\|\mathbf{W}_{1,1} - \mathbf{W}_{2,1}\|_F^2 + \|\mathbf{W}_{1,2} - \mathbf{W}_{2,2}\|_F^2\big]$ is of higher order than $\frac{\log(np)}{n} + \sqrt{\frac{\log(np)}{np^2}}$ for change point detection.

Theorem 9. *Let conditions C2-C5 hold. Then the following assertions hold.*

(i) When $\nu^{1,\tau_0} \equiv \nu^{\tau_0+1,n}$,

$$\frac{|\tau_0 - \widehat{\tau}|}{n} = O_p\left(\frac{\frac{\log(np)}{n} + \sqrt{\frac{\log(np)}{np^2}}}{\Delta_F^2} \times \min\left\{1, \frac{\min\left\{1, (n^{-1}p^2\log(np))^{\frac{1}{4}}\right\}}{\Delta_F s_{\min}}\right\}\right).$$

(ii) When $\nu^{1,\tau_0} \neq \nu^{\tau_0+1,n}$,

$$\frac{|\tau_0 - \widehat{\tau}|}{n} = O_p\left(\frac{\frac{\log(np)}{n} + \sqrt{\frac{\log(np)}{np^2}}}{\Delta_F^2} \times \min\left\{1, \frac{\min\left\{1, (n^{-1}p^2\log(np))^{\frac{1}{4}}\right\}}{\Delta_F s_{\min}} + \frac{1}{\Delta_F^2}\right\}\right).$$

Notice that for $\tau < \tau_0$, the observations in the time interval $[\tau+1, n]$ are a mixture of the two different network processes if $\nu^{1,\tau_0} \neq \nu^{\tau_0+1,n}$. In the worst case scenario then, all q communities can be changed after the change point τ_0. This causes the extra estimation error term $\frac{1}{\Delta_F^2}$ in Theorem 9(ii).

4 Simulations

4.1 Parameter estimation

We generate data according to model (2.1) in which the parameters α_{ij} and β_{ij} are drawn independently from $U[0.1, 0.5]$, $1 \leq i, j \leq p$. The initial value $\mathbf{X}_0$ was simulated according to (2.6) with $\pi_{ij} = 0.5$. We calculate the estimates according to (2.13). For each setting (with different p and n), we replicate the experiment 500 times. Furthermore we also calculate the 95% confidence intervals for α_{ij} and β_{ij} based on the asymptotically normal distributions specified in Theorem 4, and report the relative frequencies of the intervals covering the true values of the parameters. The results are summarized in Table 1.

The MSE decreases as n increases, showing steadily improvement in performance. The coverage rates of the asymptotic confidence intervals are very close to the nominal level when $n \geq 50$. The results hardly change between $p = 100$ and 200.

Table 1: The mean squared errors (MSE) of the estimated parameters in AR(1) network model (2.1) and the relative frequencies (coverage rates) of the event that the asymptotic 95% confidence intervals cover the true values in a simulation with 500 replications.

		$\widehat{\alpha}_{i,j}$		$\widehat{\beta}_{i,j}$	
n	p	MSE	Coverage (%)	MSE	Coverage (%)
5	100	.130	39.2	.131	39.3
5	200	.131	39.3	.131	39.4
20	100	.038	86.1	.037	86.0
20	200	.037	86.1	.037	86.0
50	100	.012	92.3	.012	92.2
50	200	.011	92..2	.012	92.2
100	100	.005	93.7	.005	93.8
100	200	.005	93.8	.005	93.9
200	100	.002	94.5	.002	94.5
200	200	.002	94.6	.002	94.5

4.2 Community Detection

We now consider model (3.1) with $q = 2$ or 3 clusters, in which $\theta_{i,i} = \eta_{i,i} = 0.4$ for $i = 1, \cdots, q$, and $\theta_{i,j}$ and $\eta_{i,j}$, for $1 \leq i, j \leq q$, are drawn independently from $U[0.05, 0.25]$. For each setting, we replicate the experiment 500 times.

We identify the q latent communities using the newly proposed spectral clustering algorithm based on matrix $\widehat{\mathbf{L}} = \widehat{\mathbf{L}}_1 + \widehat{\mathbf{L}}_2$ defined in (3.6). For the comparison purpose, we also implement the standard spectral clustering method for static networks (cf. Rohe et al. (2011)) but using the average

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{t=1}^{n} \mathbf{X}_t \tag{4.1}$$

in place of the single observed adjacency matrix. This idea has been frequently used in spectral clustering for dynamic networks; see, for example, Wilson et al. (2019); Zhao et al. (2019); Bhattacharjee et al. (2020). We report the normalized mutual information (NMI) and the adjusted Rand index (ARI): Both metrics take values between 0 and 1, and both measure the closeness between the true communities and the estimated communities in the sense that the larger the values of NMI and ARI are, the closer the two sets of communities are; see Vinh et al. (2010). The results are summarized in Table 2. The newly proposed algorithm based on $\widehat{\mathbf{L}}$ always outperforms the algorithm based on $\bar{\mathbf{X}}$, even when n is as small as 5. The differences between the two methods are substantial in terms of the scores of both NMI and ARI. For example when $q = 2, p = 100$ and $n = 5$, NMI and ARI are, respectively, 0.621 and 0.666 for the new method, and they are merely 0.148 and 0.158 for the standard method based on $\bar{\mathbf{X}}$. This is due to the fact that the new method identifies the latent communities using the information on both $\alpha_{i,j}$ and $\beta_{i,j}$ while the standard method uses the information on $\pi_{i,j} = \frac{\alpha_{i,j}}{\alpha_{i,j}+\beta_{i,j}}$ only.

After the communities were identified, we estimate $\theta_{i,j}$ and $\eta_{i,j}$ by (3.12) and (3.13), respectively. The mean squared errors (MSE) are evaluated for all the parameters. The results are summarized in Table 3. For the comparison purpose, we also report the estimates based on the identified communities by the $\bar{\mathbf{X}}$-based clustering. The MSE values of the estimates based on the communities identified by the new clustering method are always smaller than those of based on $\bar{\mathbf{X}}$. Noticeably now the estimates with small n such as $n = 5$ are already reasonably accurate, as the information from all the nodes within the same community is pulled together.

Table 2: Normalized mutual information (NMI) and adjusted Rand index (ARI) of the true communities and the estimated communities in the simulation with 500 replications. The communities are estimated by the spectral clustering algorithm (SCA) based on either matrix $\widehat{\mathbf{L}}$ in (3.6) or matrix $\bar{\mathbf{X}}$ in (4.1).

			SCA based on $\widehat{\mathbf{L}}$		SCA based on $\bar{\mathbf{X}}$	
q	p	n	NMI	ARI	NMI	ARI
2	100	5	.621	.666	.148	.158
		20	.733	.755	.395	.402
		50	.932	.938	.572	.584
		100	.994	.995	.692	.696
2	200	5	.808	.839	.375	.406
		20	.850	.857	.569	.589
		50	.949	.953	.712	.722
		100	.994	.995	.790	.796
3	100	5	.542	.536	.078	.057
		20	.686	.678	.351	.325
		50	.931	.929	.581	.562
		100	.988	.987	.696	.670
3	200	5	.729	.731	.195	.175
		20	.779	.763	.550	.542
		50	.954	.952	.726	.711
		100	.994	.994	.822	.802

5 Illustration with real data

5.1 RFID sensors data

Contacts between patients, patients and health care workers (HCW) and among HCW represent one of the important routes of transmission of hospital-acquired infections. Vanhems et al. (2013) collected records of contacts among patients and various types of HCW in the geriatric unit of a hospital in Lyon, France, between 1pm on Monday 6 December and 2pm on Friday 10 December 2010. Each of the $p = 75$ individuals in this study consented to wear Radio-Frequency IDentification (RFID) sensors on small identification badges during this period, which made it possible to record when any two of them were in face-to-face contact with each other (i.e. within

Table 3: The mean squared errors (MSE) of the estimated parameters in AR(1) network models with q communities. The communities are estimated by the spectral clustering algorithm (SCA) based on either matrix $\widehat{\mathbf{L}}$ in (3.6) or matrix $\bar{\mathbf{X}}$ in (4.1).

			SCA based on $\widehat{\mathbf{L}}$		SCA based on $\bar{\mathbf{X}}$	
q	p	n	$\widehat{\theta}_{i,j}$	$\widehat{\eta}_{i,j}$	$\widehat{\theta}_{i,j}$	$\widehat{\eta}_{i,j}$
2	100	5	.0149	.0170	.0298	.0312
		20	.0120	.0141	.0229	.0233
		50	.0075	.0083	.0178	.0177
		100	.0058	.0061	.0147	.0148
2	200	5	.0099	.0116	.0223	.0248
		20	.0093	.0111	.0219	.0248
		50	.0068	.0073	.0140	.0145
		100	.0061	.0062	.0117	.0118
3	100	5	.0194	.0211	.0318	.0325
		20	.0156	.0181	.0251	.0255
		50	.0093	.0104	.0193	.0193
		100	.0081	.0085	.0163	.0162
3	200	5	.0143	.0162	.0287	.0301
		20	.0134	.0156	.0200	.0205
		50	.0090	.0093	.0156	.0153
		100	.0079	.0083	.0130	.0131

1-1.5 meters) in every 20-second interval during the period. This data set is now available in R packages `igraphdata` and `sand`.

Following Vanhems et al. (2013), we combine together the recorded information in each 24 hours to form 5 daily networks ($n = 5$), i.e. an edge between two individuals is equal to 1 if they made at least one contact during the 24 hours, and 0 otherwise. Those 5 networks are plotted in Figure 1. We fit the data with stationary AR(1) model (2.1) and (2.5). Some summary statistics of the estimated parameters, according to the 4 different roles of the individuals, are presented in Table 4, together with the direct relatively frequency estimates $\widetilde{\pi}_{i,j} = \bar{X}_{i,j} = \sum_{t=1}^{5} X_{i,j}^{t}/5$. We apply the permutation test (2.14) (with 500 permutations) to the residuals resulted from the fitted AR(1) model. The P-value is 0.45, indicating no significant evidence against the stationarity assumption.

Since the original data were recorded for each 20 seconds, they can also be combined into half-day series with $n = 10$. Figure 2 presents the 10 half-day networks. We repeat the above exercise for this new sequence. Now the P-value of the permutation test is 0.008, indicating the stationary AR(1) model should be rejected for this sequence of 10 networks. This is intuitively understandable, as people behave differently at the different times during a day (such as daytime or night). Those within-day nonstationary behaviour shows up in the data accumulation over every 12 hours, and it disappears in the accumulation over 24 hour periods. Also overall the

adjacent two networks in Figure 2 look more different from each other than the adjacent pairs in Figure 1.

There have been no evidences of the existence of any communities among the 75 individuals in this data set. Our analysis confirms this too. For example the results of the spectral clustering algorithm based on, respectively, $\widehat{\mathbf{L}}$ and $\bar{\mathbf{X}}$ do not corroborate with each other at all as the NMI is smaller than 0.1.

Table 4: Mean estimated coefficients (standard errors) for the four types of individuals in RFID data. Status codes: administrative staff (ADM), medical doctor (MED), paramedical staff, such as nurses or nurses' aides (NUR), and patients (PAT).

	$\widehat{\alpha}_{ij}$			
Status	ADM	NUR	MED	PAT
ADM	.1249 (.2212)	.1739 (.2521)	.1666 (.2641)	.1113 (.2021)
NUR		.2347 (.2927)	.2398 (.3022)	.1922 (.2513)
MED			.3594 (.3883)	.1264 (.2175)
PAT				.0089 (.0552)
	$\widehat{\beta}_{ij}$			
Status	ADM	NUR	MED	PAT
ADM	.1666 (.3660)	.2326 (.3883)	.2925 (.4235)	.2061 (.3798)
NUR		.3714 (.4470)	.3001 (.4167)	.3656 (.4498)
MED			.4187 (.3973)	.2311 (.4066)
PAT				.0198 (.1331)
	$\widehat{\pi}_{ij} = \widehat{\alpha}_{ij}/(\widehat{\alpha}_{ij} + \widehat{\beta}_{ij})$			
Status	ADM	NUR	MED	PAT
ADM	.2265 (.3900)	.2478 (.3672)	.1893 (.3119)	.1239 (.2490)
NUR		.2488 (.3244)	.2729 (.3491)	.2088 (.3016)
MED			.3310 (.3674)	.1398 (.2660)
PAT				.0124 (.0928)
	$\widetilde{\pi}_{i,j} = \bar{X}_{ij}$			
Status	ADM	NUR	MED	PAT
ADM	.1250 (.3312)	.1583 (.3652)	.1704 (.3764)	.0887 (.2845)
NUR		.1854 (.3887)	.1730 (.3784)	.1542 (.3612)
MED			.3901 (.4881)	.0927 (.2902)
PAT				.0090 (.0946)

Figure 1: The RFID sensors data: the 5 networks obtained by combining together the information within each of the five 24-hour periods. The four different identities of the individuals are marked in four different colours.

5.2 French high school contact data

Now we consider a contact network data collected in a high school in Marseilles, France (Mastrandrea et al., 2015). The data are the recorded face-to-face contacts among the students from 9 classes during $n = 5$ days in December 2013, measured by the SocioPatterns infrastructure. Those are students in the so-called *classes preparatoires* – a part of the French post-secondary education system. We label the 3 classes majored in mathematics and physics as MP1, MP2 and MP3, the 3 classes majored in biology as BIO1, BIO2 and BIO3, the 2 classes majored in physics and chemistry as PC1 and PC2, and the class majored in engineering as EGI. The data are available at `www.sociopatterns.org/datasets/high-school-contact-and-friendship-networks/`. We have removed the individuals with missing values, and include the remaining $p = 327$ students in our clustering analysis based on the AR(1) stochastic block network model (see Definition 3).

We start the analysis with $q = 2$. The detected 2 clusters by the spectral clustering algorithm (SCA) based on either $\widehat{\mathbf{L}}$ in (3.6) or $\bar{\mathbf{X}}$ are reported in Table 5. The two methods lead to almost identical results: 3 classes majored in biology are in one cluster and the other 6 classes are in the other cluster. The number of 'misplaced' students is 2 and 1, respectively, by the SCA based on $\widehat{\mathbf{L}}$ and $\bar{\mathbf{X}}$. Figure 3 shows that the identified two clusters are clearly separated from each other across all the 5 days. The permutation test (2.14) on the residuals indicates that the stationary AR(1) stochastic block network model seems to be appropriate for this data set, as the P-value is 0.676. We repeat the analysis for $q = 3$, leading to equally plausible results: 3 biology classes are in one cluster, 3 mathematics and physics classes are in another cluster, and the 3 remaining classes form the 3rd cluster. See also Figure 4 for the graphical illustration with the 3 clusters.

Table 5: French high school contact network data: the detected clusters by spectral clustering algorithm (SCA) based on either $\widehat{\mathbf{L}}$ in (3.4) or $\bar{\mathbf{X}}$. The number of clusters is set at $q = 2$.

	SCA based on $\widehat{\mathbf{L}}$		SCA based on $\bar{X}$	
Class	Cluster 1	Cluster 2	Cluster 1	Cluster 2
BIO1	0	37	1	36
BIO2	1	32	0	33
BIO3	1	39	0	40
MP1	33	0	33	0
MP2	29	0	29	0
MP3	38	0	38	0
PC1	44	0	44	0
PC2	39	0	39	0
EGI	34	0	34	0

To choose the number of clusters q objectively, we define the Bayesian information criteria

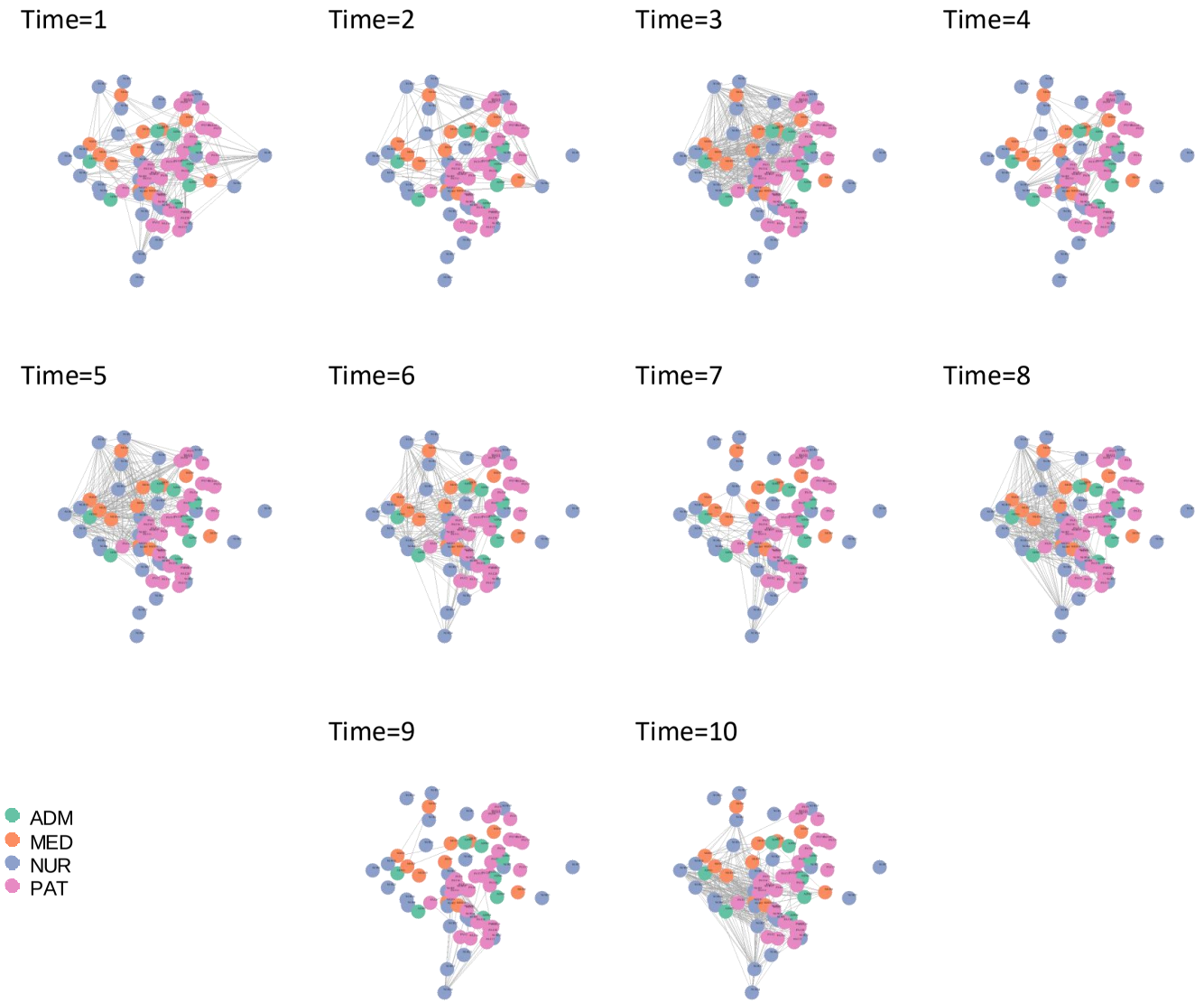

Figure 2: The RFID sensors data: the 10 networks obtained by combining together the information within each of the ten 12-hour periods. The four different identities of the individuals are marked in four different colours.

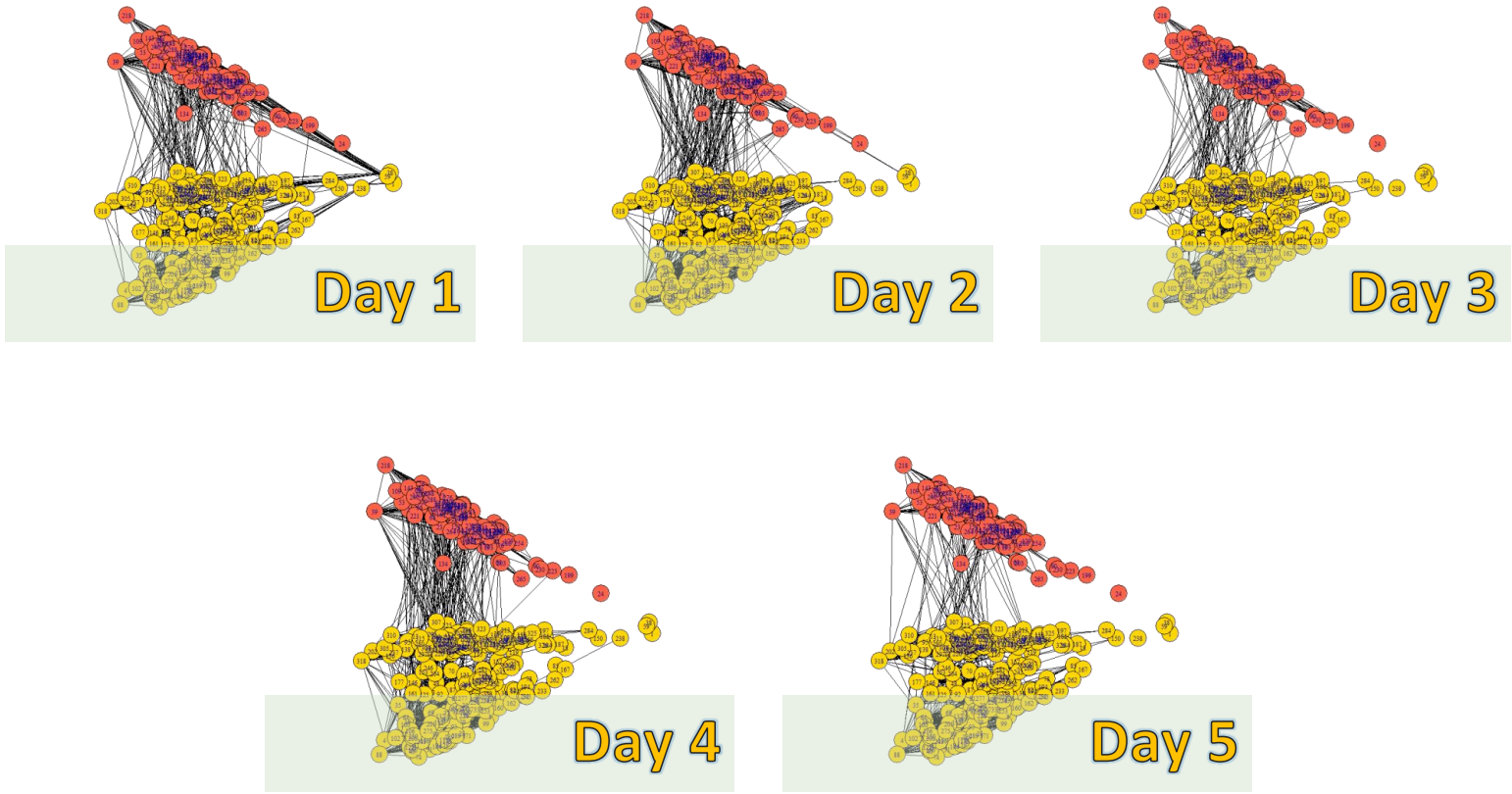

Figure 3: French high school contact networks over 5 days: the nodes marked in two colours represent the $q = 2$ clusters determined by SCA based on $\widehat{\mathbf{L}}$ in (3.4).

(BIC) as follows:

$$\mathrm{BIC}(q) = -2 \max \log(\text{likelihood}) + \log\{n(p/q)^2\} q(q+1).$$

For each fixed q, we effectively build $q(q+1)/2$ models independently and each model has 2 parameters $\theta_{k,\ell}$ and $\eta_{k,\ell}$, $1 \le k \le \ell \le q$. The number of the available observations for each model is approximately $n(p/q)^2$, assuming that the numbers of nodes in all the q clusters are about the same, which is then p/q. Thus the penalty term in the BIC above is $\sum_{1\le k\le \ell\le q} 2\log\{n(p/q)^2\} = \log\{n(p/q)^2\} q(q+1)$.

Table 6 lists the values of BIC(q) for different q. The minimum is obtained at $q = 9$, exactly the number of original classes in the school. Performing the SCA based on $\widehat{\mathbf{L}}$ with $q = 9$, we obtain almost perfect classification: all the 9 original classes are identified as the 9 clusters with only in total 4 students being placed outside their own classes. Figure 5 plots the networks with the identified 9 clusters in 9 different colours. The estimated $\theta_{i,j}$ and $\eta_{i,j}$, together with their standard errors calculated based on the asymptotic normality presented in Theorem 8, are reported in Table 7. As $\widehat{\theta}_{i,j}$ for $i \neq j$ are very small (i.e. ≤ 0.027), the students from different classes who have not contacted with each other are unlikely to contact next day. See (3.1) and (2.3). On the other hand, as $\widehat{\eta}_{i,j}$ for $i \neq j$ are large (i.e. ≥ 0.761), the students from different classes who have contacted with each other are likely to lose the contacts next day. Note that $\widehat{\theta}_{i,i}$ are greater than $\widehat{\theta}_{i,j}$ for $i \neq j$ substantially, and $\widehat{\eta}_{i,i}$ are smaller than $\widehat{\eta}_{i,j}$ for $i \neq j$ substantially. This implies that the students in the same class are more likely to contact with each other than those across the different classes.

Table 6: Fitting AR(1) stochastic block models to the French high school data: BIC values for different cluster numbers q.

q	2	3	5	7	8	9	10	11
BIC(q)	43624	40586	37726	36112	35224	34943	35002	35120

5.3 Global trade data

Our last example concerns the annual international trades among $p = 197$ countries between 1950 and 2014 (i.e. $n = 65$). We define an edge between two countries to be 1 as long as there exist trades between the two countries in that year (regardless the direction), and 0 otherwise. We take this simplistic approach to illustrate our AR(1) stochastic block model with a change-point. The data used are a subset of the openly available trade data for 205 countries in 1870 – 2014 (Barbieri et al., 2009; Barbieri and Keshk, 2016). We leave out several countries, e.g. Russian and Yugoslavia, which did not exist for the whole period concerned.

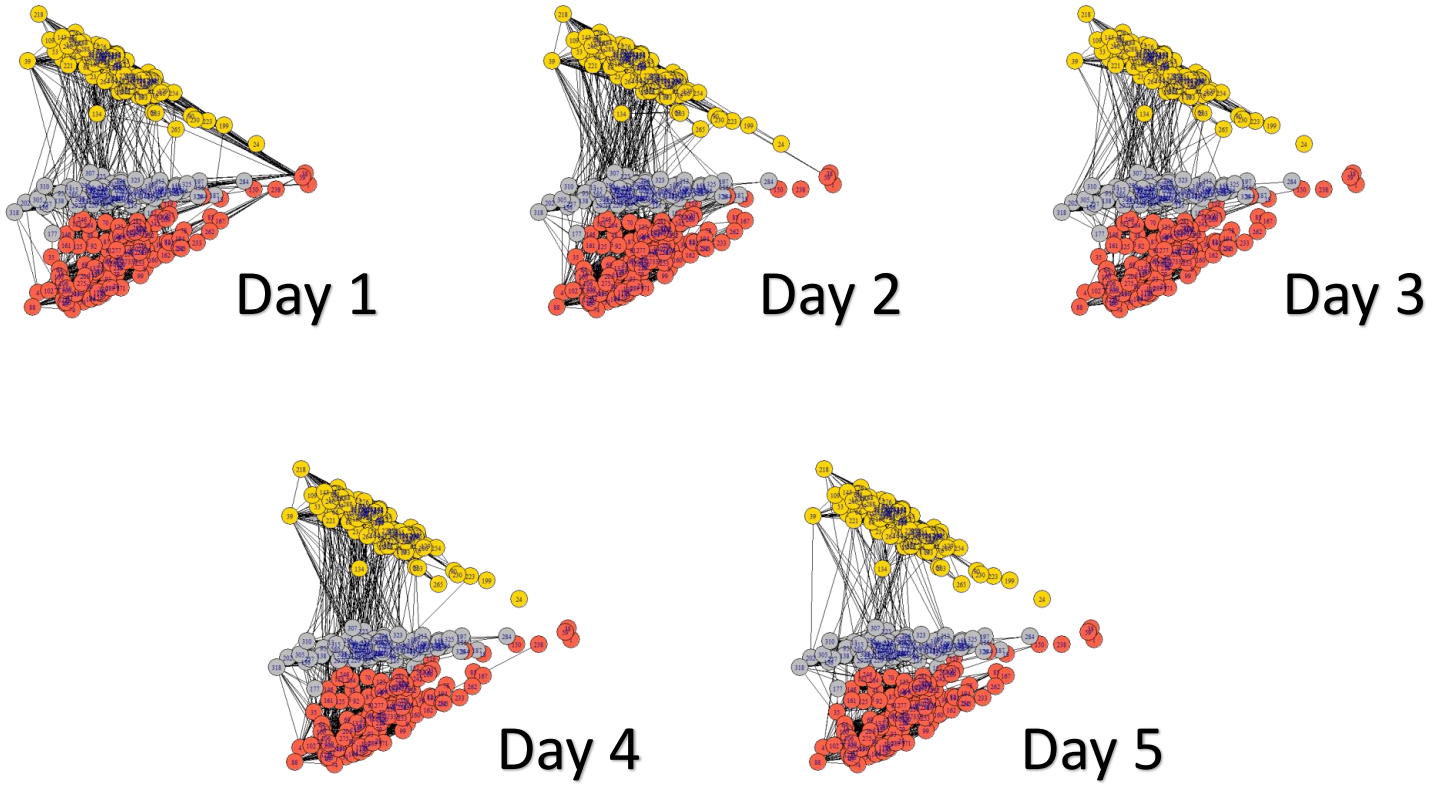

Figure 4: French high school contact networks over 5 days: the nodes marked in three colours represent the $q = 3$ clusters determined by SCA based on $\widehat{\mathbf{L}}$ in (3.4).

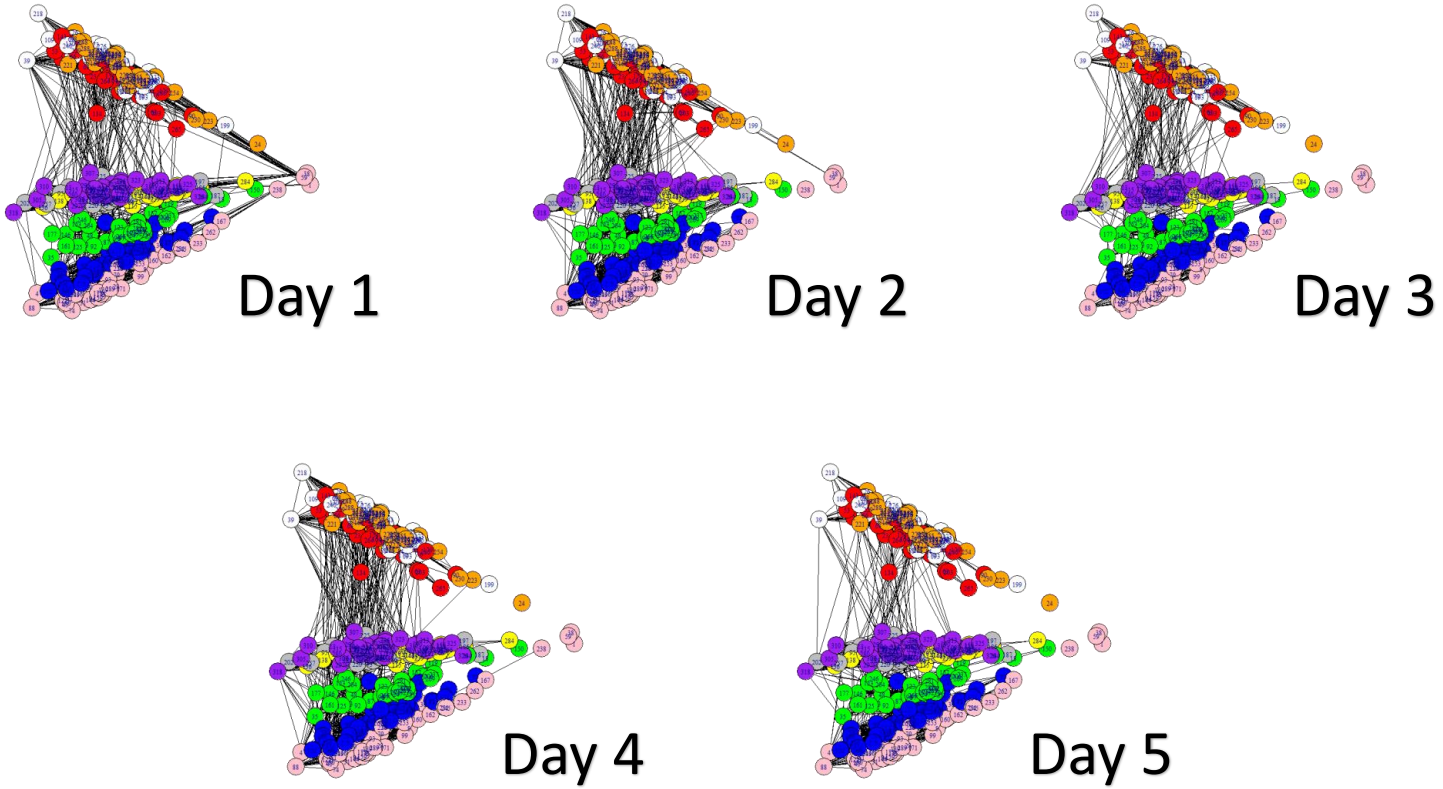

Figure 5: French high school contact networks over 5 days: the nodes marked in seven colours represent the $q = 9$ clusters determined by SCA based on $\widehat{\mathbf{L}}$ in (3.4).

Table 7: Fitting AR(1) stochastic block models with $q = 9$ clusters to the French high school data: the estimation parameters and their standard errors (in parentheses).

	Cluster	1	2	3	4	5	6	7	8	9
$\widehat{\theta}_{i,j}$	1	.246	.001	.004	.006	.001	.009	.003	.024	.003
		(.008)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.002)	(.001)
	2		.136	.024	.0018	.001	.007	.001	.001	.027
			(.009)	(.002)	(.001)	(.001)	(.001)	(.000)	(.001)	(.002)
	3			.252	.001	.002	.007	.001	.001	.022
				(.011)	(.001)	(.001)	(.001)	(.001)	(.001)	(.002)
	4				.234	.020	.001	.024	.002	.001
					(.0099	(.002)	(.001)	(.002)	(.001)	(.001)
	5					.196	.001	.020	.002	.004
						(.008)	(.001)	(.002)	(.000)	(.001)
	6						.181	.001	.010	.007
							(.008)	(.001)	(.001)	(.001)
	7							.252	.003	.006
								(.009)	(.001)	(.001)
	8								.202	.001
									(.006)	(.001)
	9									.219
										(.008)
$\widehat{\eta}_{i,j}$	1	.563	.999	.959	.976	.999	.867	.870	.792	.909
		(.015)	(.001)	(.036)	(.098)	(.001)	(.054)	(.001)	(.000)	(.051)
	2		.472	.761	.888	.999	.866	.999	.999	.866
			(.024)	(.036)	(.097)	(.001)	(.054)	(.001)	(.000)	(.026)
	3			.453	.999	.928	.864	.999	.999	.772
				(.016)	(.000)	(.066)	(.048)	(.000)	(.000)	(.031)
	4				.509	.868	.999	.784	.956	.999
					(.017)	(.028)	(.000)	(.029)	(.041)	(.000)
	5					.544	.999	.929	.842	.935
						(.017)	(.001)	(.021)	(.078)	(.041)
	6						.589	.999	.793	.923
							(.019)	(.001)	(.040)	(.036)
	7							.480	.999	.814
								(.014)	(.000)	(.051)
	8								.504	.999
									(.127)	(.000)
	9									.471
										(.014)

Setting $q = 2$, we fit the data with an AR(1) stochastic block model with two clusters. The P-value of the permutation test for the residuals resulted from the fitted model is 0, indicating overwhelmingly that the stationarity does not hold for the whole period. Applying the maximum likelihood estimator (3.15), the estimated change-point is at year 1991. Before this change-point, the identified Cluster I contains 26 countries, including the most developed industrial countries such as USA, Canada, UK and most European countries. Cluster II contains 171 countries, including all African and Latin American countries, and most Asian countries. After

1991, 41 countries switched from Cluster II and Cluster I, including Argentina, Brazil, Bulgaria, China, Chile, Columbia, Costa Rica, Cyprus, Hungary, Israel, Japan, New Zealand, Poland, Saudi Arabia, Singapore, South Korea, Taiwan, and United Arab Emirates. There was no single switch from Cluster I to II. Note that 1990 may be viewed as the beginning of the globalization. With the collapse of the Soviet Union in 1989, the fall of Berlin Wall and the end of the Cold War in 1991, the world became more interconnected. The communist bloc countries in East Europe, which had been isolated from the capitalist West, began to integrate into the global market economy. Trade and investment increased, while barriers to migration and to cultural exchange were lowered.

Figure 6 presents the average adjacency matrix of the 197 countries before and after the change-point, where the cold blue color indicates small value and the warm red color indicates large value. Before 1991, there are only 26 countries in Cluster 1. The intensive red in the small lower left corner indicates the intensive trades among those 26 countries. After 1991, the densely connected lower left corner is enlarged as now there are 67 countries in Cluster 1. Note some members of Cluster 2 also trade with the members of Cluster 1, though not all intensively.

The estimated parameters for the fitted AR(1) stochastic block model with $q = 2$ clusters are reported in Table 8. Since estimated values for $\widehat{\theta}_{1,2}, \widehat{\eta}_{1,2}$ before and after the change-point are always small, the trading status between the countries across the two clusters are unlikely to change. Nevertheless $\widehat{\theta}_{1,2}$ is 0.154 after 1991, and 0.053 before 1991; indicating greater possibility for new trades to happen after 1991.

Table 8: Fitting AR(1) stochastic block model with a change-point and $q = 2$ to the Global trade data: the estimated AR coefficients before and after 1991.

	$t \leq 1991$		$t > 1991$	
Coefficients	Estimates	SE	Estimates	SE
$\theta_{1,1}$	.062	.0092	.046	.0005
$\theta_{1,2}$	.053	.0008	.154	.0013
$\theta_{2,2}$	.023	.0002	.230	.0109
$\eta_{1,1}$	.003	.0005	.144	.0016
$\eta_{1,2}$	.037	.0008	.047	.0007
$\eta_{2,2}$	.148	.0012	.006	.0003

A final remark. We proposed in this paper a simple AR(1) setting to represent the dynamic dependence in network data explicitly. It also facilitates easy inference such as the maximum likelihood estimation and model diagnostic checking. A new class of dynamic stochastic block models illustrates the usefulness of the setting in handling more complex underlying structures including structure breaks due to change-points.

It is conceivable to construct AR(p) or even ARMA network models following the similar lines.

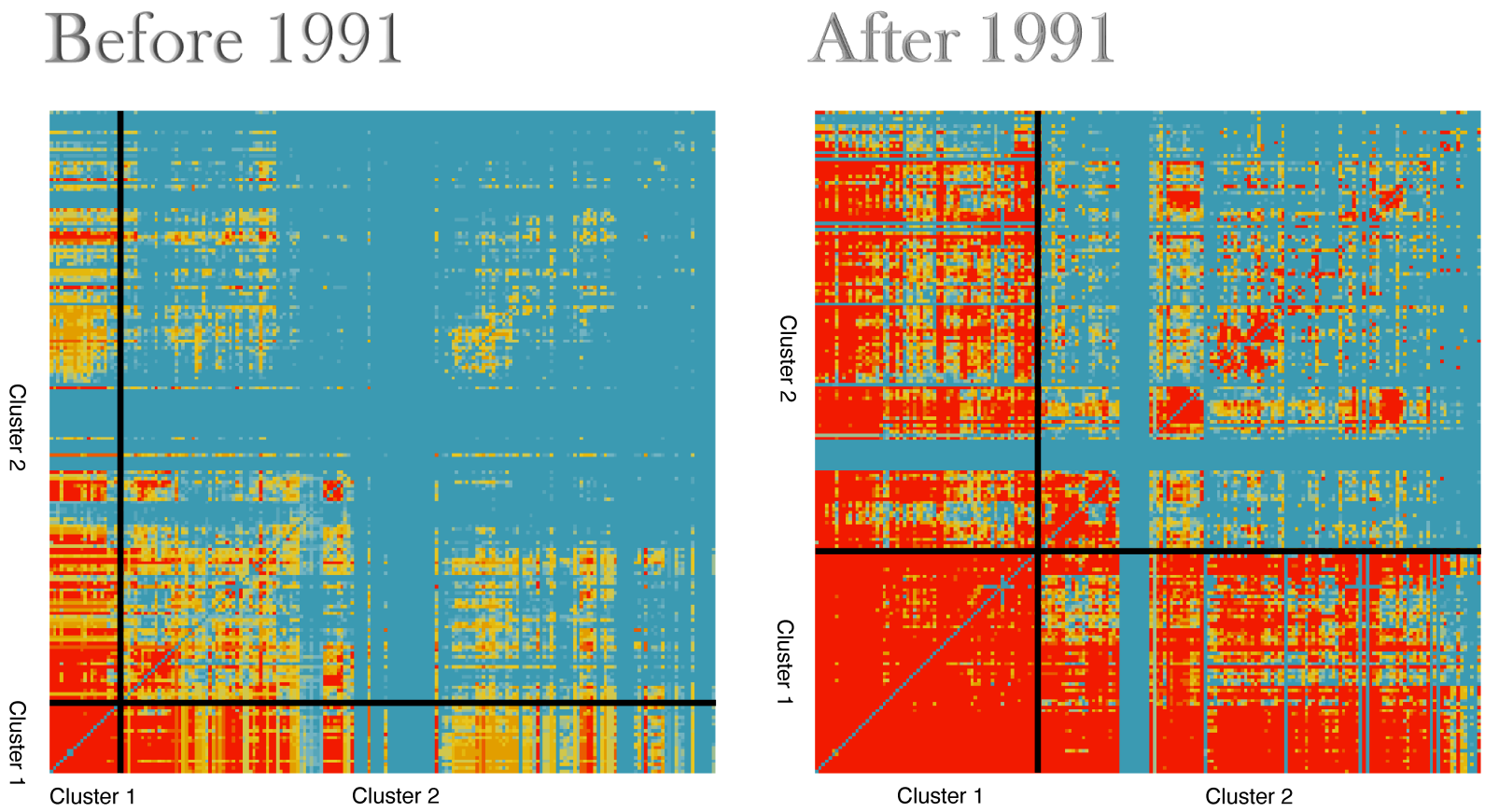

Figure 6: Average adjacency matrix for the trades among the 197 countries before and after 1991. The values from 0 to 1 are colour-coded from blue, light blue, light red to red.

However a more fertile exploration is perhaps to extend the setting for non-Erdös-Renyi networks, incorporating in the model some stylized features of network data such as *transitivity, homophily*. The development in this direction will be reported in a follow-up paper. On the other hand, dynamic networks with weighted edges may be treated as matrix time series for which effective modelling procedures have been developed based on various tensor decompositions (Wang et al., 2019; Chang et al., 2020a).

References

Aggarwal, C. and Subbian, K. (2014). Evolutionary network analysis: A survey. *ACM Computing Surveys (CSUR)*, 47(1):1–36.

Barbieri, K. and Keshk, O. M. G. (2016). *Correlates of War Project Trade Data Set Codebook, Version 4.0.* Online: http://correlatesofwar.org.

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

Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45.

Bhattacharjee, M., Banerjee, M., and Michailidis, G. (2020). Change point estimation in a dynamic stochastic block model. *Journal of Machine Learning Research*, 21(107):1–59.

Bradley, R. C. (2007). *Introduction to strong mixing conditions.* Kendrick press.

Chang, J., He, J., and Yao, Q. (2020a). Modelling matrix time series via a tensor cp-decomposition. *Under preparation.*

Chang, J., Kolaczyk, E. D., and Yao, Q. (2020b). Discussion of "network cross-validation by edge sampling". *Biometrika*, 107(2):277–280.

Chang, J., Kolaczyk, E. D., and Yao, Q. (2020c). Estimation of subgraph densities in noisy networks. *Journal of the American Statistical Association*, (In press):1–40.

Chen, E. Y., Fan, J., and Zhu, X. (2020). Community network auto-regression for high-dimensional time series. *arXiv:2007.05521.*

Crane, H. et al. (2016). Dynamic random networks and their graph limits. *The Annals of Applied Probability*, 26(2):691–721.

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

Durante, D., Dunson, D. B., et al. (2016). Locally adaptive dynamic networks. *The Annals of Applied Statistics*, 10(4):2203–2232.

Durrett, R. (2019). *Probability: theory and examples*, volume 49. Cambridge university press.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer, New York.

Fu, W., Song, L., and Xing, E. P. (2009). Dynamic mixed membership blockmodel for evolving networks. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 329–336.

Hanneke, S., Fu, W., and Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605.

Kang, X., Ganguly, A., and Kolaczyk, E. D. (2017). Dynamic networks with multi-scale temporal structure. *arXiv preprint arXiv:1712.08586*.

Kolaczyk, E. D. (2017). *Topics at the Frontier of Statistics and Network Analysis*. Cambridge University Press.

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

Lin, Z. and Bai, Z. (2011). *Probability inequalities*. Springer Science & Business Media.

Ludkin, M., Eckley, I., and Neal, P. (2018). Dynamic stochastic block models: parameter estimation and detection of changes in community structure. *Statistics and Computing*, 28(6):1201–1213.

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

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

Merlevède, F., Peligrad, M., Rio, E., et al. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In *High dimensional probability V: the Luminy volume*, pages 273–292. Institute of Mathematical Statistics.

Pensky, M. (2019). Dynamic network models and graphon estimation. *Annals of Statistics*, 47(4):2378–2403.

Rohe, K., Chatterjee, S., Yu, B., et al. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915.

Vanhems, P., Barrat, A., Cattuto, C., Pinton, J.-F., Khanafer, N., Regis, C., a. Kim, B., and B. Comte, N. V. (2013). Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PloS ONE*, 8:e73970.

Vinh, N. X., Epps, J., and Bailey, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11:2837–2854.

Wang, D., Liu, X., and Chen, R. (2019). Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics*, 208(1):231–248.

Wang, D., Yu, Y., and Rinaldo, A. (2018). Optimal change point detection and localization in sparse dynamic networks. *arXiv preprint arXiv:1809.09602*.

Wilson, J. D., Stevens, N. T., and Woodall, W. H. (2019). Modeling and detecting change in temporal networks via the degree corrected stochastic block model. *Quality and Reliability Engineering International*, 35(5):1363–1378.

Xu, K. S. and Hero, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):552–562.

Yang, T., Chi, Y., Zhu, S., Gong, Y., and Jin, R. (2011). Detecting communities and their evolutions in dynamic social networks?a bayesian approach. *Machine learning*, 82(2):157–189.

Yu, Y., Wang, T., and Samworth, R. J. (2015). A useful variant of the davis–kahan theorem for statisticians. *Biometrika*, 102(2):315–323.

Zhao, Z., Chen, L., and Lin, L. (2019). Change-point detection in dynamic networks via graphon estimation. *arXiv preprint arXiv:1908.01823*.

Zhu, T., Li, P., Yu, L., Chen, K., and Chen, Y. (2020a). Change point detection in dynamic networks based on community identification. *IEEE Transactions on Network Science and Engineering*.

Zhu, X., Huang, D., Pan, R., and Wang, H. (2020b). Multivariate spatial autoregressive model for large scale social networks. *Journal of Econometrics*, 215(2):591–606.

Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H. (2017). Network vector autoregression. *The Annals of Statistics*, 45(3):1096–1123.

Zhu, X., Wang, W., Wang, H., and Härdle, W. K. (2019). Network quantile autoregression. *Journal of Econometrics*, 212(1):345–358.