Composite likelihood inference for hidden Markov models for dynamic networks

Bartolucci, Francesco and Marino, Maria Francesca and Pandolfi, Silvia

14 October 2015

Composite likelihood inference for hidden Markov models for dynamic networks

Francesco Bartolucci
Department of Economics
University of Perugia (IT)
E-mail: francesco.bartolucci@unipg.it

Maria Francesca Marino
Department of Economics
University of Perugia (IT)
E-mail: mariafrancesca.marino@unipg.it

Silvia Pandolfi
Department of Economics
University of Perugia (IT)
E-mail: silvia.pandolfi@unipg.it

October 14, 2015

Abstract

We introduce a hidden Markov model for dynamic network data where directed relations among a set of units are observed at different time occasions. The model can also be used with minor adjustments to deal with undirected networks. In the directional case, dyads referred to each pair of units are explicitly modelled conditional on the latent states of both units. Given the complexity of the model, we propose a composite likelihood method for making inference on its parameters. This method is studied in detail for the directional case by a simulation study in which different scenarios are considered. The proposed approach is illustrated by an example based on the well-known Enron dataset about email exchange.

Keywords: Dyads; EM algorithm; Enron dataset; Latent Markov models

1 Introduction

A number of social and biological phenomena can be naturally represented in terms of networks. Here, the connection between units, that is, "actors" or "nodes", is the main target of inference. On these grounds, in the last decades, statistical models for the analysis of this type of data have known a flowering interest. Most research has focused on static networks, where data consist of a single snapshot of the network at a given time; see, among others, Goldenberg et al.

(2010) for a review. However, in some cases, the research interest may concern the evolution of networks over time. The Enron dataset (Klimt and Yang, 2004) on email exchange between employees of the company gives an interesting empirical example. Here, one may be interested in understanding how email traffic evolves over time.

In this context, standard tools of analysis need to be extended to deal with observations repeatedly taken over time, that is, with multiple snapshots of the network observed at different time points. Thus, the analysis falls into the context of longitudinal data analysis. As it is well known, although repeated measurements allow us to get deeper information on the phenomena of interest, the dependence between measures taken on the same sample units represents a further challenge that has to be faced (e.g., Diggle et al., 2002). Recently, there has been a growing amount of work on analysing dynamic networks. Key contributions are represented by dynamic exponential random graph models (Robins and Pattison, 2001) and continuous latent space models (Sarkar and Moore, 2005; Sarkar et al., 2007; Hoff, 2011; Lee and Priebe, 2011; Durante and Dunson, 2014). Within this latter context, network edges are projected in a reduced latent space where edge relations are explored.

An alternative class of models focuses on clustering nodes. Stochastic Block Models (SBMs; Holland and Leinhardt, 1976) assume that network nodes belong to one of k distinct blocks. These are defined by a discrete latent variable, with the probability of observing a connection between two nodes only depending on the corresponding block membership. That is, units in the same block connect to all the others in a similar fashion and are said to be stochastically equivalent. These models offer a concise description of the network, as a possible large number of connections is summarised by the connections between the blocks to which the units belong.

Yang et al. (2011) extended standard SBMs by considering time-varying block memberships for each unit that evolve over time according to an unobservable Markov chain. The resulting model can be conceived as a particular kind of hidden Markov model (for general references, see Zucchini and MacDonald, 2009; Bartolucci et al., 2013) for dynamic networks. Xu and Hero (2014) further extended the dynamic SBM of Yang et al. (2011) by considering time-varying edge probabilities, while Xu (2015) proposed an approach in which the presence of an edge at a given occasion directly influences future edge probabilities. An approach that is in between the dynamic latent space and the dynamic SBM is the dynamic mixed-membership SBM by Xing et al. (2010) and Ho et al. (2011). Within this context, each node may have partial memberships in several blocks.

In the framework of static networks, a number of statistical models for dyadic mutual dependences have been defined; in this context, *reciprocal* relations between units are the main target of inference. To the best of our knowledge, these types of relation have not been deeply investigated for dynamic networks. Avoiding restrictive assumptions about the dependence/independence between reciprocal relations turns out to be crucial in order to ensure model flexibility.

Here, starting from the proposal by Yang et al. (2011), we develop a SBM for dynamic networks, observed in discrete time, in which the unit of analysis is the *dyad*. As opposed to the Bayesian approaches suggested in the above mentioned works which are all based on MCMC algorithms, we obtain parameter estimates in a maximum-likelihood perspective. In this respect, a reduced computational effort is required and assumptions on the prior distribution of the model parameters can be avoided. In particular, in order to overcome the intractability of the observed data likelihood, we propose a composite likelihood approach (Lindsay, 1988; Cox and Reid, 2004) that consistently simplifies the estimation procedure and leads to reliable parameter estimates. The implementation of this method is based on an Expectation-Maximisation algorithm (EM; Dempster et al., 1977) implemented using the standard Baum-Welch recursions (Baum et al., 1970). For a related approach we refer to Bartolucci and Lupparelli (2015) who dealt with composite likelihood inference for hidden Markov models but in a different context, which is that of multilevel longitudinal data without a social network perspective. The proposed composite likelihood estimation method is studied via simulation and through the application to the Enron dataset, which represents a benchmark in the dynamic network literature. Upon request, we make available to the reader our `R` implementation of the algorithm.

The paper is organised as follows. Section 2 introduces the dynamic SBM, while Section 3 entails the description of the algorithm for parameter estimation. The results of the simulation study and of the real data application are provided in Sections 4 and 5, respectively. Last section gives some concluding remarks and outlines potential future developments.

2 The dynamic stochastic blockmodel

Let $\boldsymbol{Y}_{ij}^{(t)} = (Y_{ij}^{(t)}, Y_{ji}^{(t)})'$ denote the random vector corresponding to the dyad recorded at time occasion t between units i and j, with $i, j = 1, \dots, n$ and $t = 1, \dots, T$, where n is the number of units and T is the number of time periods of observation. Each element of the dyad, $Y_{ij}^{(t)}$, is equal to 1 if there exist an edge from unit i to unit j at occasion t and to 0 otherwise. We assume that units in the network remain unchanged during time. Also, we denote by $\tilde{\boldsymbol{Y}}_{ij} = (\boldsymbol{Y}_{ij}^{(1)}, \ldots, \boldsymbol{Y}_{ij}^{(T)})$ the matrix of dyadic relations between i and j observed during the analysed time window. As usual, realisations of random variables and related objects will be denoted by lower case letters, so that, for instance, $y_{ij}^{(t)}$ is the observed value of $Y_{ij}^{(t)}$. Finally, we define the set of all network snapshots taken across time as $\mathcal{Y} = \{\boldsymbol{Y}_{ij}^{(t)},\ i = 1, \ldots, n-1, j = i+1, \ldots, n,\ t = 1, \ldots, T\}$.

In this paper, we focus on *directed networks*, where the existence of an edge from unit i to unit j at a given occasion does not imply an edge from j to i. The extension to the undirected case is straightforwardly obtained with minor changes to the estimation algorithm. A typical example of directed networks is that of friendship nominations, where relations are not necessarily mutual, or that of email exchange that is object of the present paper; see Section 5. In this context, the dyadic relation between i and j can be either *null* ("00"), *asymmetric* ("01" or "10"), or *mutual* ("11").

In the spirit of dynamic SBMs, we assume the existence of a hidden (or latent) Markov chain $\boldsymbol{U}_i = (U_i^{(1)}, \ldots, U_i^{(T)})'$ for each sample unit i, which is defined over the finite state space $\{1, \ldots, k\}$. Latent processes $\boldsymbol{U}_i, i = 1, \ldots, n$, are assumed to be mutually independent and identically distributed, with initial probability vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)'$ and transition probability matrix $\boldsymbol{\Lambda}$ of dimension $k \times k$ with elements $\lambda_{u|v}$. The elements of this vector and matrix are defined as

$$\begin{aligned} \lambda_u &= p(U_i^{(1)} = u), \quad u = 1, \ldots, k, \\ \lambda_{u|v} &= p(U_i^{(t)} = u \mid U_i^{(t-1)} = v), \quad u, v = 1, \ldots, k,\ t = 2, \ldots, T. \end{aligned}$$

Note that these parameters are assumed to be constant over time and shared by all units in the network. These assumptions are seldom restrictive, even if generalisations are easily obtained by introducing unit and time-dependent covariates in the model; see Bartolucci et al. (2013) for a thorough discussion on the topic.

Concerning the relations between the units in the network, we assume the following model specification. For a given $t = 1, \ldots, T$, the dyad $\boldsymbol{Y}_{ij}^{(t)}$ only depends on the latent states occupied by units i and j at occasion t, that is, $U_i^{(t)}$ and $U_j^{(t)}$. Given these latent variables and for $i = 1 \ldots, n-1, j = i+1, \ldots, n$, the random variables $\boldsymbol{Y}_{ij}^{(t)}$ are conditionally independent of any other dyad and the corresponding probabilities are defined as

$$\psi_{y_1y_2|u_1u_2} = p(Y_{ij}^{(t)} = y_1, Y_{ji}^{(t)} = y_2 \mid U_i^{(t)} = u_1, U_j^{(t)} = u_2), \quad u_1, u_2 = 1, \ldots, k,\ y_1, y_2 = 0, 1. \quad (1)$$

That is, conditional on the states occupied by units in the dyad $(U_i^{(t)} = u_1, U_j^{(t)} = u_2)$ at a given time occasion, the following 2×2 matrix, denoted by $\boldsymbol{\Psi}(\boldsymbol{u})$, completely describes the corresponding dyadic relation

Y_{ij} \ Y_{ji}	0	1	
0	$\psi_{00\|u_1u_2}$	$\psi_{01\|u_1u_2}$	$\psi_{0\cdot\|u_1}$
1	$\psi_{10\|u_1u_2}$	$\psi_{11\|u_1u_2}$	$\psi_{1\cdot\|u_1}$
	$\psi_{\cdot 0\|u_2}$	$\psi_{\cdot 1\|u_2}$	1

It is worth noticing that a different way to describe the model introduced so far is by assuming that the dyad $\boldsymbol{Y}_{ij}^{(t)}$, for each $i < j$, follows a bivariate hidden Markov model defined on an augmented state space. In detail, let $\tilde{\boldsymbol{U}}_{ij} = (\boldsymbol{U}_{ij}^{(1)}, \ldots \boldsymbol{U}_{ij}^{(T)})$ be the augmented hidden Markov process with k^2 states denoted by $\boldsymbol{u}$, where $\boldsymbol{U}_{ij}^{(t)} = (U_i^{(t)}, U_j^{(t)})'$ and $\boldsymbol{u} = (u_1, u_2)'$. The corresponding initial and transition probabilities are completely defined by the parameters of the univariate latent process $\boldsymbol{U}_i$ according to the following expressions

$$\pi_{\boldsymbol{u}} = p(\boldsymbol{U}_{ij}^{(1)} = \boldsymbol{u}) = \lambda_{u_1}\lambda_{u_2},$$
$$\pi_{\boldsymbol{u}|\boldsymbol{v}} = p(\boldsymbol{U}_{ij}^{(t)} = \boldsymbol{u} \mid \boldsymbol{U}_{ij}^{(t-1)} = \boldsymbol{v}) = \lambda_{u_1|v_1}\lambda_{u_2|v_2},$$

where $\boldsymbol{v} = (v_1, v_2)'$ stands for the latent state at the previous occasion. In a more compact form, the quantities above can be directly obtained as $\boldsymbol{\pi} = \boldsymbol{\lambda} \otimes \boldsymbol{\lambda}$ and $\boldsymbol{\Pi} = \boldsymbol{\Lambda} \otimes \boldsymbol{\Lambda}$, where $\otimes$ denotes the Kronecker product. As it is standard in the Markov model literature, due to the Markovian property, the marginal distribution of the augmented latent process is given by

$$p(\tilde{\boldsymbol{U}}_{ij} = \tilde{\boldsymbol{u}}) = \pi_{\boldsymbol{u}^{(1)}} \prod_{t=2}^{T} \pi_{\boldsymbol{u}^{(t)}|\boldsymbol{u}^{(t-1)}},$$

where $\tilde{\boldsymbol{u}} = (\boldsymbol{u}^{(1)}, \ldots, \boldsymbol{u}^{(T)})$. Finally, conditional response probabilities are directly those defined in equation (1) and the joint conditional probability of all dyadic relations observed across the analysed time window is given by

$$p(\tilde{\boldsymbol{Y}}_{ij} = \tilde{\boldsymbol{y}} \mid \tilde{\boldsymbol{U}}_{ij} = \tilde{\boldsymbol{u}}) = \prod_{t=1}^{T} \psi_{\boldsymbol{y}^{(t)}|\boldsymbol{u}^{(t)}},$$

where, in general, we define $\psi_{\boldsymbol{y}|\boldsymbol{u}} = p(\boldsymbol{Y}_{ij}^{(t)} = \boldsymbol{y} \mid \boldsymbol{U}_{ij}^{(t)} = \boldsymbol{u})$. Note that, in the above expression, $\tilde{\boldsymbol{y}}$ is a realisation of $\tilde{\boldsymbol{Y}}_{ij}$, whereas $\boldsymbol{y} = (y_1, y_2)'$ is the configuration of the dyad $\boldsymbol{Y}_{ij}^{(t)}$ at time t.

As suggested by Nowicki and Snijders (2001), the parameters $\psi_{y_1 y_2|u_1 u_2}$ must be invariant with respect to *reflection*. Then, we assume the following constraints

$$\psi_{01|uu} = \psi_{10|uu}, \quad u = 1, \ldots, k, \tag{2}$$

$$\psi_{01|u_1 u_2} = \psi_{10|u_2 u_1}, \quad u_1, u_2 = 1, \ldots, k,\ u_1 \neq u_2. \tag{3}$$

This implies that the 2×2 matrix of conditional response probabilities is symmetric, that is, $\boldsymbol{\Psi}(\boldsymbol{u}) = \boldsymbol{\Psi}(\boldsymbol{u})'$, when $u_1 = u_2$. Moreover, $\boldsymbol{\Psi}(\boldsymbol{u}) = \boldsymbol{\Psi}(\boldsymbol{u}^*)'$ when $u_1 \neq u_2$, where $\boldsymbol{u}^* = (u_2, u_1)'$ is obtained by switching the elements of $\boldsymbol{u}$.

The probability of observed network $\mathcal{Y}$ is obtained by marginalising with respect to all latent variables. More precisely, we have

$$p(\mathcal{Y}) = \sum_{\tilde{\boldsymbol{u}}_{12}} \cdots \sum_{\tilde{\boldsymbol{u}}_{n-1,n}} p(\mathcal{Y} \mid \tilde{\boldsymbol{U}}_{12} = \tilde{\boldsymbol{u}}_{12}, \ldots, \tilde{\boldsymbol{U}}_{n-1,n} = \tilde{\boldsymbol{u}}_{n-1,n}) \Pr(\tilde{\boldsymbol{U}}_{12} = \tilde{\boldsymbol{u}}_{12}, \ldots, \tilde{\boldsymbol{U}}_{n-1,n} = \tilde{\boldsymbol{u}}_{n-1,n}),$$

where the sum $\sum_{\tilde{\boldsymbol{u}}_{12}} \cdots \sum_{\tilde{\boldsymbol{u}}_{n-1,n}}$ is extended to all possible configurations of the bivariate latent processes $\tilde{\boldsymbol{U}}_{ij}$ and

$$p(\mathcal{Y} \mid \tilde{\boldsymbol{U}}_{12} = \tilde{\boldsymbol{u}}_{12}, \ldots, \tilde{\boldsymbol{U}}_{n-1,n} = \tilde{\boldsymbol{u}}_{n-1,n}) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} p(\tilde{\boldsymbol{y}}_{ij} \mid \tilde{\boldsymbol{U}}_{ij} = \tilde{\boldsymbol{u}}),$$

$$p(\tilde{\boldsymbol{U}}_{12} = \tilde{\boldsymbol{u}}_{12}, \ldots, \tilde{\boldsymbol{U}}_{n-1,n} = \tilde{\boldsymbol{u}}_{n-1,n}) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} p(\tilde{\boldsymbol{U}}_{ij} = \tilde{\boldsymbol{u}}).$$

As it is clear, computation of the network distribution needs the solution of a summation over $k^{Tn(n+1)/2}$ terms that, therefore, becomes quickly cumbersome as the number of units in the network n increases. To obtain parameter estimates for the simpler model without dyads, Yang et al. (2011) proposed either the use of a variational EM algorithm based on the independence between latent variables or a Bayesian approach. In the next section, we show that a composite likelihood approach is an efficient and valid tool of analysis. It allows us to avoid the specification of assumptions on the prior distribution of model parameters and leads to estimators with properties similar to those that could by derived in a standard maximum likelihood framework.

3 Composite likelihood inference

Given the difficulties in computing the network distribution, we rely on a composite likelihood method based on the dyad probabilities for each ordered pair of units. Let $\boldsymbol{\theta}$ denote the vector of all model parameters, that is, λ_u, $\lambda_{u|v}$, $\psi_{y_1y_2|u_1u_2}$, arranged in a suitable order; the composite log-likelihood function is defined as

$$c\ell(\boldsymbol{\theta}) = \prod_{i=1}^{n-1}\prod_{j=i+1}^{n} p(\tilde{\boldsymbol{y}}_{ij}),$$

where

$$p(\tilde{\boldsymbol{y}}_{ij}) = \sum_{\tilde{\boldsymbol{u}}} p(\tilde{\boldsymbol{y}}_{ij} \mid \tilde{\boldsymbol{U}}_{ij} = \tilde{\boldsymbol{u}}) p(\tilde{\boldsymbol{U}}_{ij} = \tilde{\boldsymbol{u}}). \tag{4}$$

In order to maximise the expression above, we rely on the EM algorithm (Dempster et al., 1977) described in the following section.

3.1 Expectation-Maximisation algorithm

Let $a_{ij}^{(t)}(\boldsymbol{u})$ denote the indicator variable which is equal to 1 if, at occasion t, unit i is in state u_1 and unit j is in state u_2. Also, let $a_{ij}^{(t)}(\boldsymbol{u},\boldsymbol{v}) = a_{ij}^{(t)}(\boldsymbol{u})a_{ij}^{(t-1)}(\boldsymbol{v})$. The complete composite log-likelihood corresponding to equation (4) is defined as

$$c\ell^*(\boldsymbol{\theta}) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[\sum_{\boldsymbol{u}} a_{ij}^{(1)}(\boldsymbol{u})\log\pi_{\boldsymbol{u}} + \sum_{t=2}^{T}\sum_{\boldsymbol{u}}\sum_{\boldsymbol{v}} a_{ij}^{(t)}(\boldsymbol{u},\boldsymbol{v})\log\pi_{\boldsymbol{u}|\boldsymbol{v}} + \sum_{t=1}^{T}\sum_{\boldsymbol{u}} a_{ij}^{(t)}(\boldsymbol{u})\log\psi_{\boldsymbol{y}_{ij}^{(t)}|\boldsymbol{u}}\right]. \tag{5}$$

At the E-step, the EM algorithm computes the expected value of expression (5), conditional on the observed data and the current parameter values. This amounts to compute the posterior expectation of each dummy variable $a_{ij}^{(t)}(\boldsymbol{u})$ and $a_{ij}^{(t)}(\boldsymbol{u},\boldsymbol{v})$. To simplify the procedure, we can rely on standard forward and backward variables used in the hidden Markov model framework

(Baum et al., 1970; Welch, 2003)

$$\alpha_{ij}^{(t)}(\boldsymbol{u}) = p(\boldsymbol{y}_{ij}^{(1)}, \dots, \boldsymbol{y}_{ij}^{(t)}, \boldsymbol{U}_{ij}^{(t)} = \boldsymbol{u}),$$
$$\beta_{ij}^{(t)}(\boldsymbol{u}) = p(\boldsymbol{y}_{ij}^{(t+1)}, \dots, \boldsymbol{y}_{ij}^{(T)} \mid \boldsymbol{U}_{ij}^{(t)} = \boldsymbol{u}).$$

These can be recursively obtained by following similar arguments as those detailed by Baum et al. (1970).

Once the quantities above have been derived, the posterior expectations of $a_{ij}^{(t)}(\boldsymbol{u})$ and $a_{ij}^{(t)}(\boldsymbol{u}, \boldsymbol{v})$ can be computed as

$$\hat{a}_{ij}^{(t)}(\boldsymbol{u}) = p(\boldsymbol{U}_{ij}^{(t)} = \boldsymbol{u} \mid \boldsymbol{y}_{ij}^{(1)}, \dots, \boldsymbol{y}_{ij}^{(T)}) = \frac{\alpha_{ij}^{(t)}(\boldsymbol{u})\beta_{ij}^{(t)}(\boldsymbol{u})}{\sum_{\boldsymbol{u}} \alpha_{ij}^{(t)}(\boldsymbol{u})\beta_{ij}^{(t)}(\boldsymbol{u})},$$

$$\hat{a}_{ij}^{(t)}(\boldsymbol{u}, \boldsymbol{v}) = p(\boldsymbol{U}_{ij}^{(t-1)} = \boldsymbol{v}, \boldsymbol{U}_{ij}^{(t)} = \boldsymbol{u}, \mid \boldsymbol{y}_{ij}^{(1)}, \dots, \boldsymbol{y}_{ij}^{(T)}) = \frac{\alpha_{ij}^{(t-1)}(\boldsymbol{v})\, \pi_{\boldsymbol{u}|\boldsymbol{v}}\, \psi_{\boldsymbol{y}_{ij}^{(t)}|\boldsymbol{u}}\, \beta_{ij}^{(t)}(\boldsymbol{u})}{\sum_{\boldsymbol{v}} \sum_{\boldsymbol{u}} \alpha_{ij}^{(t-1)}(\boldsymbol{v})\, \pi_{\boldsymbol{u}|\boldsymbol{v}}\, \psi_{\boldsymbol{y}_{ij}^{(t)}|\boldsymbol{u}}\, \beta_{ij}^{(t)}(\boldsymbol{u})}.$$

In the M-step, model parameters are updated by maximising the expected composite log-likelihood for complete data. It is worth reminding that the parameters for the distribution of $\{\tilde{\boldsymbol{U}}_{ij}\}$, that is $\boldsymbol{\pi}$ and $\boldsymbol{\Pi}$, are fully determined by the parameters of the univariate latent process $\{\boldsymbol{U}_i\}$, that is $\boldsymbol{\lambda}$ and $\boldsymbol{\Lambda}$. Therefore, $\hat{a}_{ij}^{(t)}(\boldsymbol{u})$ and $\hat{a}_{ij}^{(t)}(\boldsymbol{u}, \boldsymbol{v})$ have to be properly marginalised to get the posterior probability of each state/pair of states for the latent process $\{\boldsymbol{U}_i\}$. In this respect, let

$$\hat{s}_{ij}^{(t)}(u) = \sum_{\boldsymbol{u}:u_1=u} \hat{a}_{ij}^{(t)}(\boldsymbol{u}) + \sum_{\boldsymbol{u}:u_2=u} \hat{a}_{ij}^{(t)}(\boldsymbol{u}), \quad u = 1, \dots, k, \tag{6}$$

where the first sum is extended to all latent configurations $\boldsymbol{u} = (u_1, u_2)'$ with $u_1 = u$ and the second is defined accordingly, considering all $\boldsymbol{u}$ such that $u_2 = u$. Similarly, let

$$\begin{aligned} \hat{s}_{ij}^{(t)}(u, v) = & \sum_{\boldsymbol{u}:u_1=u} \sum_{\boldsymbol{v}:v_1=v} \hat{a}_{ij}^{(t)}(\boldsymbol{u}, \boldsymbol{v}) + \sum_{\boldsymbol{u}:u_1=u} \sum_{\boldsymbol{v}:v_2=v} \hat{a}_{ij}^{(t)}(\boldsymbol{u}, \boldsymbol{v}) + \\ & + \sum_{\boldsymbol{u}:u_2=u} \sum_{\boldsymbol{v}:v_1=v} \hat{a}_{ij}^{(t)}(\boldsymbol{u}, \boldsymbol{v}) + \sum_{\boldsymbol{u}:u_2=u} \sum_{\boldsymbol{v}:v_2=v} \hat{a}_{ij}^{(t)}(\boldsymbol{u}, \boldsymbol{v}). \end{aligned} \tag{7}$$

Based on expressions (6)-(7), initial and transition probabilities $\boldsymbol{\lambda}$ and $\boldsymbol{\Lambda}$ are updated as follows:

$$\begin{aligned}\hat{\lambda}_u &= \frac{1}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\hat{s}_{ij}^{(1)}(u), \quad u=1,\ldots,k,\\ \hat{\lambda}_{v|u} &= \left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{t=1}^{T-1}\hat{s}_{ij}^{(t)}(u)\right]^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{t=2}^{T}\hat{s}_{ij}^{(t)}(u,v), \quad u,v=1,\ldots,k.\end{aligned}$$

For the conditional response probabilities $\psi_{y_1y_2|u_1u_2}$, we have to distinguish the case $u_1 = u_2$ from the case $u_1 \neq u_2$. When units in the dyad are in the same latent state at a given occasion, due to the constraints defined by expression (2), the following result holds

$$\begin{aligned}\hat{\psi}_{\boldsymbol{y}|\boldsymbol{u}} = &\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{t=1}^{T}\hat{a}_{ij}^{(t)}(\boldsymbol{u})\right]^{-1}\left\{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{t=1}^{T}\mathbb{I}(\boldsymbol{y}_{ij}^{(t)}=\boldsymbol{y})\left[\hat{a}_{ij}^{(t)}(\boldsymbol{u})\mathbb{I}(y_1=y_2)+\right.\right.\\ &\left.\left.+\frac{1}{2}\left(\hat{a}_{ij}^{(t)}(\boldsymbol{u})I(\boldsymbol{y}_{ij}^{(t)}=\boldsymbol{y})+\hat{a}_{ij}^{(t)}(\boldsymbol{u})I(\boldsymbol{y}_{ij}^{(t)}=\boldsymbol{y}^*)\right)\mathbb{I}(y_1\neq y_2)\right]\right\},\end{aligned}$$

where $\mathbb{I}(\cdot)$ denotes the indicator function and $\boldsymbol{y}^* = (y_2, y_1)$ is obtained by switching the elements of $\boldsymbol{y}$. For units being in different latent states ($u_1 \neq u_2$), based on the constraints defined by expression (3), we only need to consider the conditional response probabilities for $u_1 < u_2$; these are updated as

$$\hat{\psi}_{\boldsymbol{y}|\boldsymbol{u}} = \left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{t=1}^{T}[\hat{a}_{ij}^{(t)}(\boldsymbol{u})+\hat{a}_{ij}^{(t)}(\boldsymbol{u}^*)]\right]^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{t=1}^{T}[\hat{a}_{ij}^{(t)}(\boldsymbol{u})\mathbb{I}(\boldsymbol{y}_{ij}^{(t)}=\boldsymbol{y})+\hat{a}_{ij}^{(t)}(\boldsymbol{u}^*)\mathbb{I}(\boldsymbol{y}_{ij}^{(t)}=\boldsymbol{y}^*)].$$

The E- and the M-step of the algorithm are iterated until convergence, that is until the (relative) difference between subsequent likelihood values is lower than an arbitrary small quantity $\epsilon > 0$. In this regard, special attention must be payed on the initialisation of the EM algorithm. In fact, as typically happens when dealing with latent variables, the (composite) likelihood surface may be multimodal. Therefore, we adopt a multi-start strategy based both on a deterministic and a random starting rule. For instance, according to the first rule, we set $\lambda_u = 1/k$, $u = 1,\ldots,k$, whereas, according to the second, we first draw each λ_u from a uniform distribution between 0 and 1 and then normalise the obtained values. The random starting rule is repeatedly applied to generate a number of initial parameter values increasing with k. Overall, the solution that at convergence of the EM algorithm corresponds to the highest composite log-likelihood value is taken as the maximum composite likelihood estimate, denoted by $\hat{\boldsymbol{\theta}}$.

3.2 Standard errors and model selection

Once the model is estimated, the variance-covariance matrix of the composite likelihood estimator $\hat{\boldsymbol{\theta}}$ and the corresponding standard errors may be obtained through the standard sandwich formula and the delta method. In this regard, we first re-parametrise $\boldsymbol{\theta}$ obtaining a vector of free parameters, which is denoted by $\boldsymbol{\theta}^*$; it contains the following transformations of the initial, transition, and conditional response probabilities:

- the initial probabilities are re-parametrised according to a multinomial logit transformation using as the reference category the first hidden state:

$$\hat{\lambda}_u^* = \log \frac{\hat{\lambda}_u}{\hat{\lambda}_1}, \quad u = 2, \dots, k;$$

- the transition probabilities are re-parametrised according to a multinomial logit in which, for each row of the transition matrix, the reference state is the central one:

$$\hat{\lambda}_{v|u}^* = \log \frac{\hat{\lambda}_{v|u}}{\hat{\lambda}_{u|u}}, \quad u, v = 1, \dots, k, \ u \neq v;$$

- regarding the conditional distribution of the dyads given the hidden states u_1 and u_2, and considering constraints (2) and (3), we consider the following logits:

$$\hat{\psi}_{01|uu}^*, \ \hat{\psi}_{11|uu}^*, \quad \text{for } u = 1, \dots, k,$$

$$\hat{\psi}_{01|u_1u_2}^*, \ \hat{\psi}_{10|u_1u_2}^*, \ \hat{\psi}_{11|u_1u_2}^*, \quad \text{for } u_1 = 1, \dots, k-1, \ u_2 = u_1 + 1, \dots, k,$$

  where

$$\hat{\psi}_{y_1y_2|u_1u_2}^* = \log \frac{\hat{\psi}_{y_1,y_2|u_1u_2}}{\hat{\psi}_{00|u_1u_2}}.$$

The sandwich formula used to estimate the variance-covariance matrix for $\hat{\boldsymbol{\theta}}^*$ may be expressed as (see, among others Godambe, 1960; Varin et al., 2011),

$$\hat{\boldsymbol{\Sigma}}^*(\hat{\boldsymbol{\theta}}^*) = \hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}}^*)^{-1}\hat{\boldsymbol{K}}(\hat{\boldsymbol{\theta}}^*)\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}}^*)^{-1}, \tag{8}$$

where $\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}}^*)$ is an estimate of

$$\boldsymbol{J}_0^* = E\left[-\frac{\partial^2 c\ell(\boldsymbol{\theta}_0^*)}{\partial\boldsymbol{\theta}^*\partial(\boldsymbol{\theta}^*)'}\right],$$

and $\boldsymbol{\theta}_0^*$ is the true value of $\boldsymbol{\theta}^*$. This latter quantity is computed as minus the numerical derivative of the score vector $\boldsymbol{s}(\hat{\boldsymbol{\theta}}^*) = \partial c\ell(\hat{\boldsymbol{\theta}})/\partial\hat{\boldsymbol{\theta}}^*$ at convergence that, in turn, is equal to the first derivative with respect to $\boldsymbol{\theta}^*$ of the corresponding conditional expected value of $c\ell^*(\hat{\boldsymbol{\theta}})$. Moreover, $\hat{\boldsymbol{K}}(\hat{\boldsymbol{\theta}}^*)$ is an estimate of

$$\boldsymbol{K}_0 = V\left[-\frac{\partial c\ell(\boldsymbol{\theta}_0^*)}{\partial\boldsymbol{\theta}^*}\right],$$

which is obtained by a Monte Carlo method, as suggested by Varin et al. (2011) and recently by Bartolucci and Lupparelli (2015). For this aim, we draw a suitable number of independent samples from the fitted model and, then, we compute the score $\boldsymbol{s}(\hat{\boldsymbol{\theta}}^*)$ for each simulated sample. Finally, $\hat{\boldsymbol{K}}(\hat{\boldsymbol{\theta}}^*)$ is obtained as the variance-covariance matrix of the simulated score vectors.

Standard errors for $\hat{\boldsymbol{\theta}}^*$ are given by the square root of the diagonal elements in the variance-covariance matrix defined by equation (8). Moreover, we can obtain the standard errors for $\hat{\boldsymbol{\theta}}$, that is for the parameters expressed in the original scale, by the delta method, as described in Bartolucci and Farcomeni (2015). This amounts to derive the variance-covariance matrix as

$$\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\theta}}) = \boldsymbol{M}(\hat{\boldsymbol{\theta}}^*)\hat{\boldsymbol{\Sigma}}^*(\hat{\boldsymbol{\theta}}^*)\boldsymbol{M}(\hat{\boldsymbol{\theta}}^*)',$$

where $\boldsymbol{M}(\hat{\boldsymbol{\theta}}^*)$ is the derivative $\partial\boldsymbol{\theta}/\partial(\boldsymbol{\theta}^*)'$. Then, we again compute the square root of each diagonal element of the obtained matrix.

In order to select the number of latent states k, we can use a version of the Akaike Information Criterion (Akaike, 1973) for composite likelihood, as suggested in Varin and Vidoni (2005), denoted as CL-AIC. This criterion is based on the index

$$\text{CL-AIC} = -2\,c\ell(\hat{\boldsymbol{\theta}}) + 2\ \text{tr}\left(\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})^{-1}\hat{\boldsymbol{K}}(\hat{\boldsymbol{\theta}})\right). \tag{9}$$

Alternatively, we rely on the composite Bayesian Information Criterion (CL-BIC; Gao and Song, 2010) which is based on the index

$$\text{CL-BIC} = -2\,c\ell(\hat{\boldsymbol{\theta}}) + \text{tr}\left(\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})^{-1}\hat{\boldsymbol{K}}(\hat{\boldsymbol{\theta}})\right)\log n, \tag{10}$$

where $\text{tr}\left(\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})^{-1}\hat{\boldsymbol{K}}(\hat{\boldsymbol{\theta}})\right)$ represents the penalty accounting for the model complexity. According to both criteria, the model to be selected is the one corresponding to the minimum value of the indexes in (10) and (9).

4 Simulation study

In the following, we illustrate the results of a large scale Monte Carlo simulation study aimed at assessing the performance of the proposed approach. Different experimental scenarios have been considered to evaluate the empirical behaviour of our proposal when both the sample size and the number of measurement occasions vary.

4.1 Design

In the simulation, data are generated from a two state ($k = 2$) and a three state ($k = 3$) dynamic SBM considering three different experimental scenarios. Scenario 1 (our benchmark) entails a dynamic network referred to $n = 100$ units observed at $T = 10$ measurement occasions. To understand how our proposal performs when n or T increase, we considered two additional scenarios: the former (Scenario 2) involves $T = 20$ snapshots of a network with $n = 100$ units, the latter (Scenario 3) involves $T = 10$ snapshots of a network with $n = 200$ units.

For the dynamic SBM with $k = 2$ hidden states, we fixed the following values for the initial probability vector and the transition probability matrix:

$$\boldsymbol{\lambda} = (0.4, 0.6)', \qquad \boldsymbol{\Lambda} = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{pmatrix},$$

while the conditional response probabilities we assumed

(y_1, y_2) \ (u_1, u_2)	(1,1)	(1,2)	(2,2)
00	0.60	0.20	0.20
01	0.10	0.50	0.10
10	0.10	0.10	0.10
11	0.20	0.20	0.60

For the dynamic SBM with $k = 3$ latent states, the parameters of the hidden Markov process are fixed to

$$\boldsymbol{\lambda} = (0.25, 0.5, 0.25)', \qquad \boldsymbol{\Lambda} = \begin{pmatrix} 0.80 & 0.15 & 0.05 \\ 0.10 & 0.80 & 0.10 \\ 0.05 & 0.15 & 0.80 \end{pmatrix}.$$

The snapshot of the network at a given measurement occasion is obtained on the basis of the following conditional probabilities

(y_1, y_2) \ (u_1, u_2)	(1,1)	(1,2)	(1,3)	(2,2)	(2,3)	(3,3)
00	0.91	0.69	0.29	0.29	0.05	0.00
01	0.04	0.13	0.21	0.21	0.13	0.04
10	0.04	0.13	0.21	0.21	0.13	0.04
11	0.00	0.05	0.29	0.29	0.69	0.91

4.2 Results

Tables 1-3 report the simulation results for the dynamic SBM with $k = 2$ hidden states under the different experimental scenarios. The results are based on $B = 200$ simulated datasets. The performance of our approach is evaluated in terms of bias, standard deviation (sd), and root mean square error (rmse) of the estimators.

By looking at the estimated rmses, we observe that, in general, the quality of the results improves when the number of measurement occasions increases (Scenario 2) and, more substantially, when a higher number of units is available (Scenario 3). Focusing on the parameters of the latent process, it may be noticed that the initial probability vector is estimated with slightly lower accuracy than the transition matrix, regardless the value of n and T.

As for the parameters of the latent Markov process, parameters defining the conditional response probability of the dyads are estimated with high accuracy, in terms of bias, and precision, in terms of variability; see Table 3. When both the dimension of the network and the number of observed snapshots increase, the variability of the parameter estimates seems to reduce, ensuring the consistency of the proposed estimation approach in recovering the true data structure.

Tables 4-8 report the estimation results for the dynamic SBM with $k = 3$ latent states under the three experimental scenarios. As expected, the quality of the results obtained under this model specification turns out to be lower with respect to that observed for the model with $k = 2$ states due to the higher uncertainty on the latent structure of the model. A higher bias and a higher variability for the initial probability vector may be observed (see Table 4). This result is

Table 1: *Bias, standard deviation (sd), and root mean square error (rmse) for the estimator of the initial probabilities* (λ_u) *under different scenarios, with* $k = 2$ *latent states (Sc.1:* $n = 100, T = 10$*; Sc.2:* $n = 100, T = 20$*; Sc.3:* $n = 200, T = 10$*).*

	Sc.1			*Sc.2*			*Sc.3*		
	bias($\hat{\boldsymbol{\lambda}}$)	sd($\hat{\boldsymbol{\lambda}}$)	rmse($\hat{\boldsymbol{\lambda}}$)	bias($\hat{\boldsymbol{\lambda}}$)	sd($\hat{\boldsymbol{\lambda}}$)	rmse($\hat{\boldsymbol{\lambda}}$)	bias($\hat{\boldsymbol{\lambda}}$)	sd($\hat{\boldsymbol{\lambda}}$)	rmse($\hat{\boldsymbol{\lambda}}$)
$u = 1$	-0.007	0.080	0.080	-0.002	0.061	0.061	0.001	0.046	0.046
$u = 2$	0.007	0.080	0.080	0.002	0.061	0.061	-0.001	0.046	0.046

Table 2: *Bias, standard deviation (sd), and root mean square error (rmse) for the estimator of the transition probabilities* ($\lambda_{u|v}$) *under different scenarios, with* $k = 2$ *latent states (Sc.1:* $n = 100, T = 10$*; Sc.2:* $n = 100, T = 20$*; Sc.3:* $n = 200, T = 10$*).*

		bias($\hat{\boldsymbol{\Lambda}}$)		sd($\hat{\boldsymbol{\Lambda}}$)		rmse($\hat{\boldsymbol{\Lambda}}$)	
		$u = 1$	$u = 2$	$u = 1$	$u = 2$	$u = 1$	$u = 2$
Sc.1	$u = 1$	0.000	-0.000	0.061	0.061	0.061	0.061
	$u = 2$	0.000	-0.000	0.047	0.047	0.047	0.047
Sc.2	$u = 1$	-0.003	0.003	0.028	0.028	0.028	0.028
	$u = 2$	0.002	-0.002	0.025	0.025	0.025	0.025
Sc.3	$u = 1$	-0.001	0.001	0.025	0.025	0.025	0.025
	$u = 2$	0.000	-0.000	0.023	0.023	0.023	0.023

in agreement with the reduced amount of information which is available for each parameter. A slight reduction in the rmse values is only observed under Scenario 3 due to a higher number of units in the network. When focusing on the estimation of the transition probabilities reported in Table 5, a higher accuracy and a reduced variability of parameter estimates may be observed, with very few exceptions, which are due to certain samples corresponding to estimates that considerably differ from those obtained for the other samples.

Tables 6-8 report the estimated parameters for the conditional response probabilities of the dyads under the three experimental scenarios. On the basis of these results, we notice that parameter estimates present quite a similar behaviour with respect to those observed for the initial and the transition probabilities. The quality of results improves when a higher number of units or a higher number of repeated measurements is available. Thus, based on these empirical findings, the proposed approach may be considered as an interesting tool of analysis to understand temporal dynamics characterising network data.

Table 3: *Bias, standard deviation (sd), and root mean square error (rmse) for the estimator of the conditional response probabilities* ($\psi_{y_1 y_2|u_1 u_2}$) *under different scenarios, with* $k = 2$ *latent states (Sc.1:* $n = 100, T = 10$*; Sc.2:* $n = 100, T = 20$*; Sc.3:* $n = 200, T = 10$*).*

	bias($\hat{\mathbf{\Psi}}$)			sd($\hat{\mathbf{\Psi}}$)			rmse($\hat{\mathbf{\Psi}}$)		
$(y_1\, y_2)$ \ (u_1, u_2)	$(1,1)$	$(1,2)$	$(2,2)$	$(1,1)$	$(1,2)$	$(2,2)$	$(1,1)$	$(1,2)$	$(2,2)$
Sc.1									
00	0.011	-0.000	-0.002	0.089	0.029	0.022	0.089	0.029	0.022
01	-0.007	0.008	-0.002	0.038	0.037	0.024	0.038	0.037	0.024
10	-0.007	-0.002	-0.002	0.038	0.008	0.024	0.038	0.008	0.024
11	0.004	-0.006	0.005	0.052	0.038	0.053	0.052	0.038	0.053
Sc.2									
00	0.000	-0.000	-0.000	0.029	0.013	0.011	0.029	0.013	0.011
01	-0.001	0.002	-0.000	0.015	0.016	0.012	0.014	0.016	0.012
10	-0.001	-0.000	-0.000	0.015	0.003	0.012	0.014	0.003	0.012
11	0.001	-0.001	0.001	0.013	0.015	0.026	0.013	0.015	0.026
Sc.3									
00	-0.003	0.000	-0.001	0.026	0.011	0.008	0.026	0.011	0.008
01	0.001	0.000	-0.000	0.012	0.014	0.010	0.012	0.014	0.010
10	0.001	-0.001	-0.000	0.012	0.003	0.010	0.012	0.003	0.010
11	0.002	0.000	0.002	0.010	0.012	0.023	0.010	0.012	0.023

5 Application: Enron email network

A large set of email messages was made public during the legal investigation concerning the Enron corporation. The row Enron corpus (Klimt and Yang, 2004) consists of 619,446 messages that were sent or received by 158 users between 1998 and 2002; the processed version contains information on 200,399 messages with an average of 757 emails per user. Following Tang et al. (2008), we first considered only communications recorded between April, 2001 and March, 2002 involving users who sent and received at least 5 emails during that period, thus obtaining a total number of users equal to 2,359. To further reduce the dimensionality of the network and preserve the real data structure, we randomly chose, among these 2,359 users, $n = 151$ Enron employees to build up the data matrix. In this application, $y_{ij}^{(t)} = 1$ if user i sent at least one email message to user j during the t-th month of the analysed time window, with $i = 1, \ldots, 150, j = i + 1, \ldots, 151$ and $t = 1, \ldots, 12$.

Here, the interest is in understanding the evolution of dyadic relations between users (email exchange) over time, defining groups characterised by similar communication profiles. To this extent, we estimated a dynamic SBM with a varying number of latent states ($k = 2, \ldots, 6$).

Table 4: *Bias, standard deviation (sd), and root mean square error (rmse) for the estimator of the initial probabilities (λ_u) under different scenarios, with $k = 3$ latent states (Sc.1: $n = 100, T = 10$; Sc.2: $n = 100, T = 20$; Sc.3: $n = 200, T = 10$).*

	Sc.1			*Sc.2*			*Sc.3*		
	bias($\hat{\boldsymbol{\lambda}}$)	sd($\hat{\boldsymbol{\lambda}}$)	rmse($\hat{\boldsymbol{\lambda}}$)	bias($\hat{\boldsymbol{\lambda}}$)	sd($\hat{\boldsymbol{\lambda}}$)	rmse($\hat{\boldsymbol{\lambda}}$)	bias($\hat{\boldsymbol{\lambda}}$)	sd($\hat{\boldsymbol{\lambda}}$)	rmse($\hat{\boldsymbol{\lambda}}$)
$u = 1$	0.064	0.147	0.160	0.050	0.152	0.159	0.059	0.128	0.140
$u = 2$	-0.087	0.202	0.220	-0.101	0.196	0.220	-0.107	0.161	0.193
$u = 3$	0.023	0.158	0.159	0.051	0.142	0.150	0.048	0.136	0.144

Table 5: *Bias, standard deviation (sd), and root mean square error (rmse) for the estimator of the transition probabilities ($\lambda_{u|v}$) under different scenarios, with $k = 3$ latent states (Sc.1: $n = 100, T = 10$; Sc.2: $n = 100, T = 20$; Sc.3: $n = 200, T = 10$).*

		bias($\hat{\boldsymbol{\Lambda}}$)			sd($\hat{\boldsymbol{\Lambda}}$)			rmse($\hat{\boldsymbol{\Lambda}}$)		
		$u = 1$	$u = 2$	$u = 3$	$u = 1$	$u = 2$	$u = 3$	$u = 1$	$u = 2$	$u = 3$
Sc.1	$u = 1$	0.031	-0.044	0.013	0.079	0.098	0.059	0.085	0.108	0.061
	$u = 2$	0.039	-0.071	0.033	0.127	0.173	0.154	0.132	0.187	0.157
	$u = 3$	0.018	-0.048	0.029	0.067	0.092	0.089	0.070	0.104	0.093
Sc.2	$u = 1$	0.039	-0.063	0.024	0.059	0.077	0.050	0.071	0.100	0.055
	$u = 2$	0.029	-0.031	0.002	0.106	0.127	0.100	0.110	0.130	0.100
	$u = 3$	0.017	-0.055	0.038	0.048	0.070	0.055	0.051	0.089	0.066
Sc.3	$u = 1$	0.027	-0.048	0.021	0.075	0.083	0.051	0.079	0.095	0.055
	$u = 2$	0.052	-0.079	0.027	0.108	0.147	0.114	0.119	0.166	0.117
	$u = 3$	0.016	-0.059	0.043	0.050	0.075	0.068	0.052	0.095	0.081

Also, in order to reduce the change of being trapped in local maxima, we adopted the multi-start strategy described in Section 3.1; for each value of $k = 2, \dots, 6$, we retained the best solution according to the CL-BIC and CL-AIC indexes. Results are reported in Table 9.

As it frequently happens, BIC-type indexes are more conservative than the corresponding AIC ones, suggesting to select a model with a lower number of parameters. In the present framework, CL-BIC leads to selecting a model with $k = 3$ latent states, while CL-AIC prefers the solution with $k = 4$ states. In the following, we discuss results for both choices in order to assess the sensitivity of the parameter estimates to the value k.

5.1 Dynamic SBM with $k = 3$ latent states

Table 10 shows the parameter estimates of the probability of the dyad $\boldsymbol{Y}_{ij}^{(t)}$, conditional on the latent states occupied by unit i and j at occasion t and the corresponding standard errors obtained by the sandwich formula illustrated in Section 3.2. We report only non-redundant

Table 6: *Bias, standard deviation (sd), and root mean square error (rmse) for the estimator of the conditional response probabilities ($\psi_{y_1 y_2|u_1 u_2}$) under Scenario 1 ($n = 100, T = 10$), with $k = 3$ latent states.*

$(y_1\, y_2)$ \ (u_1, u_2)	$(1,1)$	$(1,2)$	$(1,3)$	$(2,2)$	$(2,3)$	$(3,3)$
	bias($\hat{\mathbf{\Psi}}$)					
00	-0.061	-0.087	-0.000	0.019	0.023	0.017
01	0.023	0.019	-0.002	-0.057	0.006	0.024
10	0.023	0.022	-0.009	-0.057	0.015	0.024
11	0.014	0.046	0.012	0.094	-0.045	-0.065
	sd($\hat{\mathbf{\Psi}}$)					
00	0.077	0.162	0.119	0.273	0.081	0.036
01	0.030	0.055	0.037	0.063	0.056	0.036
10	0.030	0.052	0.041	0.063	0.083	0.036
11	0.031	0.092	0.123	0.277	0.160	0.090
	rmse($\hat{\mathbf{\Psi}}$)					
00	0.098	0.183	0.119	0.273	0.084	0.039
01	0.038	0.058	0.037	0.085	0.056	0.043
10	0.038	0.056	0.042	0.085	0.084	0.043
11	0.034	0.103	0.124	0.292	0.166	0.111

combination of latent states keeping in mind the constraints defined in equation (2) and (3).

Based on these results, we are able to identify three groups having quite a different profile. The first hidden state corresponds to *inactive* users, that is, employees that do not interact with any peers. As suggested by Yang et al. (2011), this represents a necessary state to account for the sparsity of the data matrix. Hidden states 2 and 3 identify instead *active* users. More in detail, we may distinguish a group of units (those in state 2) that do not interact with any peers in the same group ($\psi_{00|22} = 1$), but that have a quite high chance of receiving emails from units in the third state ($\psi_{01|23} = 0.51$). Also, mutual communications between units in state 3 at a given occasion are highly likely ($\psi_{11|33} = 0.85$).

The obtained results suggest the presence of three different communication profiles in the Enron company: *inactive* (state 1), *email receivers* (state 2) and *email senders* (state 3). Clearly, the estimates discussed so far allow us to characterise email exchange between Enron employees at a given month of the analysed observation window. To understand how the email traffic of the company evolves over time, we may analyse the estimates of parameters defining the hidden Markov process and the corresponding standard errors; see Table 11. As it is clear, Enron employees present a higher probability of being in the first latent state at the beginning

Table 7: *Bias, standard deviation (sd), and root mean square error (rmse) for the estimator of the conditional response probabilities ($\psi_{y_1y_2|u_1u_2}$) under Scenario 2 ($n = 100, T = 20$), with $k = 3$ latent states.*

(y_1, y_2) \ (u_1, u_2)	$(1,1)$	$(1,2)$	$(1,3)$	$(2,2)$	$(2,3)$	$(3,3)$
	bias($\hat{\mathbf{\Psi}}$)					
(0,0)	-0.066	-0.060	-0.004	0.051	0.029	0.014
(0,1)	0.026	0.011	-0.004	-0.050	0.012	0.024
(1,0)	0.026	0.015	-0.007	-0.050	0.012	0.024
(1,1)	0.015	0.033	0.014	0.049	-0.053	-0.063
	sd($\hat{\mathbf{\Psi}}$)					
(0,0)	0.082	0.135	0.103	0.260	0.074	0.029
(0,1)	0.030	0.045	0.030	0.043	0.047	0.031
(1,0)	0.030	0.045	0.032	0.043	0.043	0.031
(1,1)	0.035	0.066	0.110	0.256	0.141	0.080
	rmse($\hat{\mathbf{\Psi}}$)					
(0,0)	0.105	0.148	0.103	0.264	0.079	0.032
(0,1)	0.039	0.046	0.030	0.066	0.049	0.040
(1,0)	0.039	0.047	0.032	0.066	0.045	0.040
(1,1)	0.038	0.074	0.111	0.260	0.151	0.102

of the observation period ($\hat{\lambda}_1 = 0.66$). Furthermore, state 1 almost represents an absorbing state with a persistence probability equal to $\hat{\lambda}_{1|1} = 0.98$. This result can be related to the sparsity of the data matrix observed over all the analysed time window. Concerning the other latent states, persistence over time is quite evident ($\hat{\lambda}_{2|2} = \hat{\lambda}_{3|3} = 0.84$). Some transitions may still be observed, with units in the *email receiver* group that move with probability $\hat{\lambda}_{1|2} = 0.10$ towards the *inactive* state and with probability $\hat{\lambda}_{3|2} = 0.13$ towards the *email sender* group in two subsequent measurement occasions.

5.2 Dynamic SBM with $k = 4$ latent states

We show in Table 12 the estimated parameters and the corresponding standard errors of the conditional response probabilities under the dynamic SBM with $k = 4$ latent states. As described in Section 5.1, also in this case, results reported in Table 12 allow us to distinguish between *inactive* and *active* users. The former do not send/receive emails neither from/to employees in the same group, nor from/to employees being in other latent states. On the other hand, *active* users are classified in three different latent states corresponding to different communication profiles.

The fourth latent state is characterised by the existence of within group mutual relations

Table 8: *Bias, standard deviation (sd), and root mean square error (rmse) for the estimator of the conditional response probabilities* ($\psi_{y_1 y_2|u_1 u_2}$) *under Scenario 3* ($n = 200, T = 10$)*, with* $k = 3$ *latent states.*

(y_1, y_2) \ (u_1, u_2)	$(1,1)$	$(1,2)$	$(1,3)$	$(2,2)$	$(2,3)$	$(3,3)$
	bias($\hat{\mathbf{\Psi}}$)					
(0,0)	-0.057	-0.064	0.006	0.026	0.029	0.013
(0,1)	0.022	0.016	-0.002	-0.054	0.016	0.022
(1,0)	0.022	0.016	-0.006	-0.054	0.016	0.022
(1,1)	0.012	0.031	0.002	0.081	-0.061	-0.056
	sd($\hat{\mathbf{\Psi}}$)					
(0,0)	0.066	0.142	0.091	0.256	0.070	0.031
(0,1)	0.027	0.045	0.026	0.046	0.039	0.024
(1,0)	0.027	0.045	0.024	0.046	0.043	0.024
(1,1)	0.023	0.072	0.097	0.261	0.133	0.063
	rmse($\hat{\mathbf{\Psi}}$)					
(0,0)	0.087	0.155	0.091	0.257	0.076	0.034
(0,1)	0.035	0.048	0.026	0.071	0.042	0.033
(1,0)	0.035	0.048	0.025	0.071	0.045	0.033
(1,1)	0.026	0.079	0.097	0.273	0.145	0.085

($\hat{\psi}_{11|44} = 1$) and by a high chance of sending email both to units in the second and the third latent state ($\hat{\psi}_{01|24} = 0.65, \hat{\psi}_{01|34} = 0.49$). Such a state can be labelled as the *global sender* group. Regarding states 2 and 3, the distinction between them is mainly associated with the observed relations with units in the fourth latent state. State 2 corresponds to *receiver only* users ($\hat{\psi}_{01|24} = 0.65$), while state 3 identifies employees that are both senders and receivers with respect to the fourth latent group ($\hat{\psi}_{01|34} = 0.49, \hat{\psi}_{11|34} = 0.40$). This state can be labelled as the *sender/receiver* group.

When analysing the estimated initial and transition probabilities, we get similar results as those derived for the dynamic SBM with $k = 3$ latent states; see Table 13. The *inactive* latent states is the most likely one at the beginning of the observation window. Also, the probability of

Table 9: *Enron data. CL-BIC and CL-AIC for different choices of* k*.*

	latent states k				
	2	3	4	5	6
CL-BIC	45857.94	**44417.88**	44428.89	44580.84	44625.63
CL-AIC	45429.84	43804.95	**43710.53**	43764.99	43807.94

Table 10: *Enron data. Estimates and estimated standard errors (se) for the conditional response probabilities of the dynamic SBM with $k = 3$ latent states.*

	estimates						se					
	(u_1, u_2)											
(y_1, y_2)	$(1,1)$	$(1,2)$	$(1,3)$	$(2,2)$	$(2,3)$	$(3,3)$	$(1,1)$	$(1,2)$	$(1,3)$	$(2,2)$	$(2,3)$	$(3,3)$
(0,0)	1.00	1.00	0.95	1.00	0.33	0.01	0.00	0.00	0.01	0.00	0.03	0.02
(0,1)	0.00	0.00	0.04	0.00	0.51	0.07	0.00	0.00	0.01	0.00	0.02	0.01
(1,0)	0.00	0.00	0.01	0.00	0.04	0.07	0.00	0.00	0.00	0.00	0.00	0.01
(1,1)	0.00	0.00	0.01	0.00	0.11	0.85	0.00	0.00	0.00	0.00	0.02	0.04

Table 11: *Enron data. Estimates and estimated standard errors (se) for the latent Markov model parameters of the dynamic SBM with $k = 3$ latent states.*

	estimates				se			
v	$\hat{\lambda}_v$	$\hat{\lambda}_{1\|v}$	$\hat{\lambda}_{2\|v}$	$\hat{\lambda}_{3\|v}$	$\hat{\lambda}_v$	$\hat{\lambda}_{1\|v}$	$\hat{\lambda}_{2\|v}$	$\hat{\lambda}_{3\|v}$
1	0.66	0.98	0.02	0.00	0.05	0.01	0.01	0.00
2	0.21	0.10	0.84	0.06	0.04	0.03	0.03	0.02
3	0.13	0.03	0.13	0.84	0.03	0.01	0.05	0.05

observing no transitions from this latent state is close to 1 ($\hat{\lambda}_{1|1} = 0.98$), thus highlighting the sparsity of the network that remains persistent over all the analysed time window. Similarly, for the other hidden states, transitions are quite unlikely. In particular, units in the *receiver only* group tend to move towards the *inactive* group ($\hat{\lambda}_{1|2} = 0.14$), while units in the *global sender* group move towards the *sender/receiver* one ($\hat{\lambda}_{3|4} = 0.14$). Finally, transitions between state 3 and 4 and between state 3 and 2 within two subsequent measurement occasions seems to be almost equally likely ($\hat{\lambda}_{2|3} = 0.10, \hat{\lambda}_{4|3} = 0.12$).

6 Concluding remarks

In this paper we discuss dynamic stochastic blockmodels (SBMs) for dynamic networks in a hidden Markov model framework. In this perspective, we are able to identify groups of units characterised by similar profiles, whose composition may change over time. In order to relax the local independence assumption which is typically used when dealing with dynamic SBMs, we analyse the dyads referred to ordered pairs of units.

Reciprocal relations between units in the network are described by means of a bivariate latent Markov process, with augmented latent space which is defined starting from the corresponding univariate latent profiles. For this class of models, computation of the full likelihood to obtain

Table 12: *Enron data. Estimates and estimated standard errors (se) for the conditional response probabilities of the dynamic SBM with $k = 4$ latent states.*

	estimates									
	(u_1, u_2)									
(y_1, y_2)	$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(2,2)$	$(2,3)$	$(2,4)$	$(3,3)$	$(3,4)$	$(4,4)$
(0,0)	1.00	1.00	1.00	0.95	0.54	0.97	0.29	0.67	0.07	0.00
(0,1)	0.00	0.00	0.00	0.04	0.14	0.03	0.65	0.17	0.49	0.00
(1,0)	0.00	0.00	0.00	0.01	0.14	0.00	0.00	0.17	0.04	0.00
(1,1)	0.00	0.00	0.00	0.00	0.17	0.00	0.06	0.00	0.40	1.00

	se									
	(u_1, u_2)									
(y_1, y_2)	$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(2,2)$	$(2,3)$	$(2,3)$	$(3,3)$	$(3,4)$	$(4,4)$
(0,0)	0.00	0.00	0.01	0.03	0.10	0.07	0.05	0.12	0.03	0.00
(0,1)	0.00	0.00	0.00	0.02	0.03	0.07	0.04	0.06	0.05	0.00
(1,0)	0.00	0.00	0.01	0.01	0.03	0.00	0.00	0.06	0.02	0.00
(1,1)	0.00	0.00	0.00	0.01	0.05	0.00	0.02	0.00	0.08	0.00

Table 13: *Enron data. Estimates and estimated standard errors (se) for the latent Markov model parameters of the dynamic SBM with $k = 4$ latent states.*

	estimates					se				
v	$\hat{\lambda}_v$	$\hat{\lambda}_{1\|v}$	$\hat{\lambda}_{2\|v}$	$\hat{\lambda}_{3\|v}$	$\hat{\lambda}_{4\|v}$	$\hat{\lambda}_v$	$\hat{\lambda}_{1\|v}$	$\hat{\lambda}_{2\|v}$	$\hat{\lambda}_{3\|v}$	$\hat{\lambda}_{4\|v}$
1	0.68	0.98	0.01	0.00	0.00	0.04	0.01	0.01	0.00	0.00
2	0.08	0.14	0.85	0.00	0.00	0.03	0.06	0.06	0.01	0.00
3	0.14	0.00	0.10	0.78	0.12	0.05	0.00	0.04	0.05	0.04
4	0.09	0.06	0.00	0.17	0.77	0.03	0.03	0.00	0.05	0.04

parameter estimates becomes progressively infeasible as the dimension of the network increases. For this reason, we propose a composite likelihood approach, defined on all possible pairs of observations. When compared to the Bayesian approaches which are typically used with dynamic SBMs, the composite likelihood method requires a lower computational effort and, also, allows us to avoid the specification of the prior distribution of model parameters that, in some cases, may severely affect inferential conclusions.

The behaviour of the proposed approach is evaluated by means of a large scale simulation study and a real data application. Simulation results suggest that the composite likelihood approach allows us to recover the true data structure with high precision, both for the observed and the latent part of the model. The analysis of the Enron dataset highlights the capability of the dynamic SBMs for dyads in offering a complete and deep description of the relations between units in the network, while avoiding unverifiable model assumptions.

An interesting evolution of the proposed approach may be based on adopting marginal parametrisation for the conditional distribution of each dyad given the underlying Markov chains. This parametrisation is based on two logits for each response variable (marginal with respect to the other response variable) and the log-odds ratio that measures the conditional association between reciprocal relations given the latent states. In this way, it would be also possible to allow for individual covariates in the analysis and to formulate more parsimonious models in which, for instance, the level of conditional dependence is constant across latent states. Also, constraints of interest may be formulated on the Markov chain parameters assuming, for instance, that the initial distribution corresponds to the stationary distribution. In all cases, the composite likelihood inferential approach developed in this paper can be used in these extended versions of the model.

Acknowledgments

We acknowledge the financial support from award RBFR12SHVV of the Italian Government (FIRB "Mixture and latent variable models for causal inference and analysis of socio-economic data", 2012).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, pages 267–281. Akademinai Kiado.

Bartolucci, F. and Farcomeni, A. (2015). Information matrix for hidden markov models with covariates. *Statistics and Computing*, 25:515–526.

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2013). *Latent Markov Models for Longitudinal Data*. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences. Taylor & Francis.

Bartolucci, F. and Lupparelli, M. (2015). Pairwise likelihood inference for nested hidden markov chain models for multilevel longitudinal data. *Journal of the American Statistical Association*, pages 00–00.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41:164–171.

Cox, D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91:729–737.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B. Methodological*, 39:1–38.

Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. (2002). *Analysis of longitudinal data*. Oxford University Press.

Durante, D. and Dunson, D. B. (2014). Nonparametric bayes dynamic modelling of relational data. *Biometrika*.

Gao, X. and Song, P. X.-K. (2010). Composite likelihood Bayesian information criteria for model selection in high-dimensional data. *Journal of the American Statistical Association*, 105:1531–1540.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31:1208–1211.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends® in Machine Learning*, 2:129–233.

Ho, Q., Song, L., and Xing, E. P. (2011). Evolving cluster mixed-membership blockmodel for time-evolving networks. In *International Conference on Artificial Intelligence and Statistics*, pages 342–350.

Hoff, P. D. (2011). Hierarchical multilinear models for multiway data. *Computational Statistics and Data Analysis*, 55:530–543.

Holland, P. and Leinhardt, S. (1976). Local structure in social networks. *Sociological Methodology*, 7:1–45.

Klimt, B. and Yang, Y. (2004). The Enron corpus: a new dataset for email classification research. In *Machine Learning: ECML 2004*, volume 3201, pages 217–226. Springer Berlin Heidelberg.

Lee, N. and Priebe, C. (2011). A latent process model for time series of attributed random graphs. *Statistical inference for stochastic processes*, 14:231–253.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics*, 80:221–39.

Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96:1077–1087.

Robins, G. and Pattison, P. (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, 25:5–41.

Sarkar, P. and Moore, A. W. (2005). Dynamic social network analysis using latent space models. *ACM SIGKDD Explorations Newsletter*, 7:31–40.

Sarkar, P., Siddiqi, S. M., and Gordon, G. J. (2007). A latent space approach to dynamic embedding of co-occurrence data. In *International Conference on Artificial Intelligence and Statistics*, pages 420–427.

Tang, L., Liu, H., Zhang, J., and Nazeri, Z. (2008). Community evolution in dynamic multi-mode networks. In *14th ACM SIGKDD International Conference on Knowledge Discovery and Data ining*, pages 677–685.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21:5–42.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92:519–528.

Welch, L. R. (2003). Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter*, 53:10–13.

Xing, E. P., Fu, W., Song, L., et al. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, 4:535–566.

Xu, K. (2015). Stochastic block transition models for dynamic networks. In *18th International Conference on Artificial Intelligence and Statistics*, pages 1079–1087.

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

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*, 82:157–189.

Zucchini, W. and MacDonald, I. (2009). *Hidden Markov models for time series*. CRC Press.