Post print version of the paper published in Computational Statistics and Data Analysis:
https://doi.org/10.1016/j.csda.2018.01.010

Dealing with Reciprocity in Dynamic Stochastic Block Models

Francesco Bartolucci[a], Maria Francesca Marino[b], Silvia Pandolfi[a,*]

[a]*Department of Economics, University of Perugia*
[b]*Department of Statistics, Computer Science, Applications "G. Parenti", University of Florence*

Abstract

A stochastic block model for dynamic network data is introduced, where directed relations among a set of nodes are observed at different time occasions and the blocks are represented by a sequence of latent variables following a Markov chain. Dyads are explicitly modeled conditional on the states occupied by both nodes involved in the relation. With respect to the approaches already available in the literature, the main focus is on reciprocity. In this regard, three different parameterizations are proposed in which: (*i*) reciprocity is allowed to depend on the blocks of the nodes in the dyad; (*ii*) reciprocity is assumed to be constant across blocks; and (*iii*) reciprocity is ruled out. The assumption of conditional independence between dyads (referred to different pairs of nodes and time occasions) given the latent blocks is always retained. Given the complexity of the model, inference on its parameters is based on a variational approach, where a lower bound of the log-likelihood function is maximized instead of the intractable full model log-likelihood. An approximate likelihood ratio test statistic is proposed which compares the value at convergence of this lower bound under different model specifications. This allows us to formally test for both the hypothesis of no reciprocity and that of constant reciprocity with respect to the latent blocks. The proposed approach is illustrated via a simulation study based on different scenarios. The application to two benchmark datasets in the social network literature is also proposed to illustrate the effectiveness of the proposal in studying reciprocity and identifying groups of nodes having a similar social behavior.

Keywords: Dyads, EM algorithm, hidden Markov models, likelihood ratio test, variational inference

1. Introduction

A number of social, behavioral, and biological phenomena can be naturally represented in terms of networks. In this literature, the relation between units, that is, "actors" or "nodes",

*Corresponding author. Silvia Pandolfi, Department of Economics, University of Perugia, Via A. Pascoli 20. 06123 Perugia, ITALY. Phone: +39 075 5855278

Email addresses: francesco.bartolucci@unipg.it (Francesco Bartolucci), mariafrancesca.marino@unifi.it (Maria Francesca Marino), silvia.pandolfi@unipg.it (Silvia Pandolfi)

is the main target of inference and statistical models for the analysis of these relations have known a flowering interest. Most methods available in the literature are tailored to deal with static networks, where data consist of a single snapshot observed at a given occasion; see, among others, Goldenberg et al. (2010) and Amati et al. (2018) for a review.

Within this context, models for *clustering* and *community discovering* based on latent variables play an important role. Among these models, it is worth mentioning latent space models (Sarkar and Moore, 2006; Sarkar et al., 2007; Hoff, 2011; Lee and Priebe, 2011; Durante and Dunson, 2014), which project network nodes on a reduced latent space where relations between them are explored, and Stochastic Block Models (SBMs; Holland et al., 1983; Snijders and Nowicki, 1997; Nowicki and Snijders, 2001; Daudin et al., 2008), which assume that network nodes belong to one of k distinct blocks. In this latter framework, relational variables are assumed to be independent conditional on the blocks of the nodes involved in the relation (*local independence assumption*). Blocks are defined by a discrete latent variable, with the probability of observing a connection between two nodes only depending on the corresponding block memberships. Therefore, nodes in the same block connect to all the others in a similar fashion and are said to be stochastically equivalent. The identification of these blocks provides a concise description of the network.

However, in some cases, the research interest may be on the evolution of the network over time, provided that longitudinal network data are available. 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. Although longitudinal data permit a deeper study of the phenomenon 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).

The literature about models specifically tailored to deal with dynamic networks is rather recent, with most proposals starting from approaches developed for static networks. In this article, we focus on extensions of the SBM for longitudinal data. In particular, Yang et al. (2011) developed a dynamic SBM by considering time-varying block memberships that evolve over time according to an unobservable Markov chain. The resulting model can be conceived as a particular type of hidden (latent) Markov model (for general references, see Bartolucci et al., 2013; Zucchini et al., 2016) for dynamic networks. Xu and Hero (2014) further extended the dynamic SBM of Yang et al. (2011) by considering time-varying edge probabilities. The same model has been recently discussed by Matias and Miele (2017), who proposed an approach to solve the lack of identifiability due to label switching between time steps and a well-principled estimation approach. Finally, Xu (2015) proposed a model in which the presence of a relation at a given occasion directly influences future relation probabilities. An approach that is in between the dynamic latent space and the dynamic SBM is the dynamic mixed-membership SBM of Xing et al. (2010) and Ho et al. (2011). In this context, each node may have partial membership to different blocks.

Alternative proposals to the dynamic SBM are represented by the dynamic exponential random graph model for the analysis of social networks observed in discrete time (Robins and Pattison, 2001; Hanneke et al., 2010; Lusher et al., 2013, Chapter 10). Further references include the stochastic actor-oriented model (Snijders, 1996, 2001, 2005; Snijders et al., 2010)

and the relational event model (Butts, 2008; Quintane et al., 2014), which are based on a continuous time Markov process and on a time-to-event representation, respectively.

Extending the proposal of Yang et al. (2011), we develop an SBM for dynamic networks observed in discrete time in which the principal element of analysis is the *dyad* referred to each pair of nodes, conditional on the hidden states they occupy at each occasion. Therefore, we avoid restrictive assumptions about the dependence/independence between reciprocal relations and, thus, obtain higher flexibility than that of standard dynamic SBMs. The main assumption is that of conditional independence between the dyads, given the corresponding latent variables representing the blocks. Note, however, that marginal dependence between dyads is not ruled out, but is explained in a meaningful way by the latent variables. Therefore, triangulation or similar higher-order effects are accounted for. In agreement with Vu et al. (2013), among others, conditional independence between dyads offers at least three advantages: (*i*) it implies simplification in the estimation process; (*ii*) it facilitates data simulation; and (*iii*) it avoids the degeneracy issue which is frequently encountered when dealing with SBMs.

To permit a deeper insight into reciprocity effects, we propose to parametrically specify every dyadic relation between nodes in the network by means of a suitably formulated log-linear model, given the latent blocks. Therefore, we may distinguish between main and reciprocal effects reflecting the tendency to observe asymmetric and symmetric relations, respectively, and therefore we may obtain information on the network's cohesion. In particular, our approach allows us to formulate three different hypotheses: (*i*) reciprocity may depend on the blocks in which the nodes involved in the relation belong to; (*ii*) reciprocity is constant across blocks; and (*iii*) reciprocity is absent.

Estimation of the proposed model represents a challenging matter as computing the log-likelihood function would require the evaluation of a multiple summation defined over all possible configurations of the latent variables. Clearly, this becomes quickly unfeasible as the size of the network, and then the number of such latent variables, increases. In the literature, two main approaches are available to derive model parameter estimates. Markov Chain Monte Carlo (MCMC) algorithms represent a typical option in the Bayesian framework (e.g., Yang et al., 2011), while variational approximation methods represent a quite classical solution in the frequentist context (e.g., Yang et al., 2011; Matias and Miele, 2017). In this paper, we start from the proposal of Yang et al. (2011) and obtain parameter estimates through a Variational Expectation-Maximization (VEM) algorithm based on the assumption of posterior independence between dyads. Also, we propose an approximate inferential procedure with the aim of testing for the presence of reciprocity effects in the network. Starting from the lower bound of the likelihood function required for variational inference, we show how an approximate Likelihood Ratio (LR) test statistic, which is simply computed, may be used for inferential purposes on the reciprocity parameters.

Properties of the proposed inferential method, and in particular of the approximate LR test, are investigated via simulation and through the application to two benchmark datasets in the dynamic network literature: the Newcomb Fraternity network and the Enron dataset. The results show the potentialities of the proposed approach. Upon request, we make available the `R` implementation of the proposed estimation algorithm.

The paper is organized as follows. Section 2 introduces the dynamic SBM according to the initial proposal of Yang et al. (2011) and then illustrates the proposed extension to deal with different forms of reciprocity. Section 3 entails the description of the VEM algorithm for parameter estimation and introduces the approximate LR test for hypotheses of reciprocity. The results of the simulation study and of the real data applications are provided in Sections 4 and 5, respectively. The last section contains some concluding remarks and outlines potential future developments.

2. Dynamic stochastic block models

For a network of n individuals observed at T time occasions, let $Y_{ij}^{(t)}, i, j = 1, \ldots, n, j \neq i$, denote a binary response variable which is equal to 1 if there exists an edge from unit i to unit j at occasion t and is equal to 0 otherwise; $y_{ij}^{(t)}$ is used to denote a realization of $Y_{ij}^{(t)}$. Moreover, let $\boldsymbol{Y}^{(t)}$ be the binary adjacency matrix recorded at occasion $t = 1, \ldots, T$, which summarizes the relations between nodes. Here, we focus on directed networks without self-loops, so that $\boldsymbol{Y}^{(t)}$ is not constrained to be symmetric and all the corresponding diagonal elements are missing. Finally, we define the set of all network snapshots taken across time as $\mathcal{Y} = \{\boldsymbol{Y}^{(1)}, \ldots, \boldsymbol{Y}^{(T)}\}$.

2.1. Current approaches

In the spirit of standard dynamic SBMs (e.g., Yang et al., 2011; Matias and Miele, 2017), each node in the network is assumed to belong to one of k distinct blocks. These blocks are identified by the individual- and time-specific latent variables $U_i^{(t)}$ that are discrete with support $\{1, \ldots, k\}$. Moreover, dynamic SBMs assume that the vectors $\boldsymbol{U}_i = (U_i^{(1)}, \ldots, U_i^{(T)})'$ are mutually independent and identically distributed; each of them follows a Markov chain with initial probability vector $\boldsymbol{\lambda}$, having elements λ_u, $u = 1, \ldots, k$, and transition probability matrix $\boldsymbol{\Pi}$, having dimension $k \times k$ and elements $\pi_{u|v}$. These parameters are defined as follows:

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

Note that the initial and the transition probabilities are assumed to be the same for all nodes, with transition probabilities that are also time homogeneous. These assumptions are seldom restrictive, even if generalizations may be easily obtained by parametrically specifying λ_u and $\pi_{u|v}$ and by considering unit- and time-dependent covariates; see Bartolucci et al. (2013) for a thorough discussion on the topic.

A further crucial assumption of dynamic SBMs is that of *local independence*: given all latent variables $U_i^{(t)}$ and $U_j^{(t)}$, the relational variables $Y_{ij}^{(t)}$ (response variables) are conditionally independent. Moreover, each of these variables follows a Bernoulli distribution with success probability that only depends on the hidden states (or blocks) occupied by node i and j at occasion t, that is,

$$Y_{ij}^{(t)} \mid U_i^{(t)} = u_1, U_j^{(t)} = u_2 \sim \text{Bern}(\phi_{u_1 u_2}), \quad u_1, u_2 = 1, \ldots, k,\ t = 1, \ldots, T,$$

where $\phi_{u_1u_2} = p(Y_{ij}^{(t)} = 1 | U_i^{(t)} = u_1, U_j^{(t)} = u_2)$.

Different generalizations of the above model are available in the literature. Weighted networks may be easily accommodated by considering a different conditional distribution for the random variables $Y_{ij}^{(t)}$; affiliation structures or sparse data may be both modeled by considering suitable mixtures with two components; time-varying parameters for the response distribution may also be considered to improve model flexibility. See, among others, Ambroise and Matias (2012), Xu and Hero (2014), or Matias and Miele (2017) for a detailed description of these approaches.

In the next section, we will further extend the dynamic SBM. In particular, we partially relax the local independence assumption and account for reciprocal effects by directly modeling dyadic relations. This allows us to avoid restrictive assumptions about reciprocity, while simplifying both data simulation and the estimation of model parameters.

2.2. Proposed dyadic formulation

Let $\boldsymbol{D}_{ij}^{(t)} = (Y_{ij}^{(t)}, Y_{ji}^{(t)})'$ denote the random vector corresponding to the dyad involving nodes i and j at occasion t, with $i, j = 1, \ldots, n$, $i < j$, and $t = 1, \ldots, T$. In this context, the dyadic relation between i and j can be *null* – $(0, 0)$ –, *asymmetric* – $\{(0, 1), (1, 0)\}$ –, or *mutual* – $(1, 1)$.

In order to model reciprocity, we assume that the dyads $\boldsymbol{D}_{ij}^{(t)}$ (rather than the single response variables $Y_{ij}^{(t)}$) are conditionally independent given all latent variables $U_1^{(t)}, \ldots, U_n^{(t)}$, relaxing in this way the local independence assumption. As already mentioned, the proposed model does not rule out marginal dependence between the dyads; on the contrary, this dependence is explained in a meaningful way by the latent variables. More precisely, dependence between two dyads of type $\boldsymbol{D}_{ij}^{(t_1)}$ and $\boldsymbol{D}_{ij}^{(t_2)}$, for two different time occasions t_1 and t_2, or of type $\boldsymbol{D}_{hi}^{(t)}$, $\boldsymbol{D}_{hj}^{(t)}$, and $\boldsymbol{D}_{ij}^{(t)}$ for three units h, i, and j at the same time occasion is allowed and, thus, triangulation or higher order effects that are well known in the social network literature (Holland and Leinhardt, 1976) are accounted for. Clearly, the type of local independence that we assume here is less restrictive than the standard assumption described in the previous section. Conditional on $U_i^{(t)} = u_1$ and $U_j^{(t)} = u_2$, the blocks occupied by the nodes in a given dyad at a certain occasion t, we denote the dyad probabilities as

$$\begin{aligned} \psi_{y_1y_2|u_1u_2} &= p(Y_{ij}^{(t)} = y_1, Y_{ji}^{(t)} = y_2 \mid U_i^{(t)} = u_1, U_j^{(t)} = u_2) \\ &= p(\boldsymbol{D}_{ij}^{(t)} = \boldsymbol{d} \mid U_i^{(t)} = u_1, U_j^{(t)} = u_2), \end{aligned} \tag{1}$$

with $u_1, u_2 = 1, \ldots, k$, $y_1, y_2 = 0, 1$, and $\boldsymbol{d} = (y_1, y_2)$, so that $\boldsymbol{d} \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. Therefore, relations between two nodes are completely described by the following 2×2 matrix of conditional dyad probabilities:

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

(2)

In the following, we will denote the above matrix by $\boldsymbol{\Psi}(u_1, u_2)$.

As already clarified, reciprocal relations between units play a major role in this work. In particular, we propose to parametrically specify dyad probabilities in equation (1) according to the following log-linear model:

$$\psi_{y_1 y_2|u_1 u_2} = \frac{\exp\left[\alpha_{u_1 u_2} y_1 + (\alpha_{u_1 u_2} + \beta_{u_1 u_2}) y_2 + \rho_{u_1 u_2}\, y_1 y_2\right]}{M(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\rho})} \tag{3}$$

where $M(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\rho})$ is the normalizing constant obtained by summing the numerator for all possible configurations of (y_1, y_2) and, as before, $u_1, u_2 = 1, \ldots, k$, and $y_1, y_2 = 0, 1$.

To ensure identifiability, we assume that $\beta_{uu} = 0$, for $u = 1, \ldots, k$. Furthermore, $\alpha_{u_1 u_2} = \alpha_{u_2 u_1} + \beta_{u_2 u_1}$, $\beta_{u_1 u_2} = -\beta_{u_2 u_1}$, and $\rho_{u_1 u_2} = \rho_{u_2 u_1}$, for all $u_1 \neq u_2$. In practice, the parameters that must be estimated are $\alpha_{u_1 u_2}$ and $\rho_{u_1 u_2}$ for $u_1 \leqslant u_2$ and $\beta_{u_1 u_2}$ for $u_1 < u_2$, with $u_1, u_2 = 1, \ldots, k$. These free parameters are collected in the vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\rho}$, whose elements are ordered as follows: $\boldsymbol{\alpha} = (\alpha_{11}, \ldots, \alpha_{1k}, \ldots, \alpha_{kk})'$, $\boldsymbol{\beta} = (\beta_{12}, \ldots, \beta_{k-1,k})'$, and $\boldsymbol{\rho} = (\rho_{11}, \ldots, \rho_{1k}, \ldots, \rho_{kk})'$.

Besides permitting a clear analysis of reciprocity, the proposed formulation improves the interpretability of the results. In this regard, it is worth noting that parameters $\alpha_{u_1 u_2}$ and $\beta_{u_1 u_2}$ may be directly expressed in terms of logits. In fact, we have that

$$\alpha_{u_1 u_2} = \log \frac{\psi_{10|u_1 u_2}}{\psi_{00|u_1 u_2}} = \log \frac{p(Y_{ij}^{(t)} = 1 | U_i^{(t)} = u_1, U_j^{(t)} = u_2, Y_{ji}^{(t)} = 0)}{p(Y_{ij}^{(t)} = 0 | U_i^{(t)} = u_1, U_j^{(t)} = u_2, Y_{ji}^{(t)} = 0)},$$

corresponding to the tendency of observing a relation from a node in block u_1 to a node in block u_2 (given that such a relation is not reciprocated). As regards the second set of parameters, we have that

$$\beta_{u_1 u_2} = \log \frac{\psi_{01|u_1 u_2}}{\psi_{10|u_1 u_2}}, \tag{4}$$

so that positive values for $\beta_{u_1 u_2}$ correspond to a higher probability of observing asymmetric relations from units in block u_2 to units in block u_1 than the opposite. Negative values for $\beta_{u_1 u_2}$ correspond to a higher probability of asymmetric relations from units in block u_1 to units in block u_2 than that from u_2 to u_1. Clearly, when $\beta_{u_1 u_2} = 0$, no differences in terms of non-reciprocated relations between the two blocks are present. Last, the parameter $\rho_{u_1 u_2}$ in equation (3) characterizes mutual relations between nodes in the network. It corresponds to the following log-odds ratio:

$$\rho_{u_1 u_2} = \log \frac{\psi_{00|u_1 u_2} \psi_{11|u_1 u_2}}{\psi_{01|u_1 u_2} \psi_{10|u_1 u_2}},$$

which is a well-known measure of association between two binary variables in the statistical literature (e.g., Agresti, 2013, Ch. 2).

Note that the log-linear parameterization we propose is in agreement with the constraints suggested by Nowicki and Snijders (2001) to avoid redundancy in the parameters $\psi_{y_1 y_2|u_1 u_2}$

and requiring these probabilities to be invariant with respect to *reflection*:

$$\psi_{01|uu} = \psi_{10|uu}, \quad u = 1, \ldots, k,$$
$$\psi_{01|u_1u_2} = \psi_{10|u_2u_1}, \quad u_1, u_2 = 1, \ldots, k,\ u_1 \neq u_2.$$

This implies that the matrix of conditional dyad probabilities $\boldsymbol{\Psi}(u_1, u_2)$ defined in (2) is symmetric when $u_1 = u_2$, that is, $\boldsymbol{\Psi}(u, u) = \boldsymbol{\Psi}(u, u)'$. Under the proposed model, this holds as a consequence of the constraints $\beta_{uu} = 0$ and is in agreement with the interpretation of these parameters provided in equation (4). Moreover, we have $\boldsymbol{\Psi}(u_1, u_2) = \boldsymbol{\Psi}(u_2, u_1)'$ when $u_1 \neq u_2$ as a consequence of the constraints we assume to ensure identifiability.

Different versions of the proposed model specification are obtained by imposing constraints on the $\rho_{u_1u_2}$ parameters. In particular, when we consider the hypothesis

$$H_I : \rho_{u_1u_2} = 0, \quad u_1, u_2 = 1, \ldots, k,\ u_1 \leqslant u_2, \tag{5}$$

the model directly reduces to the standard dynamic SBM of Yang et al. (2011), based on the local independence between responses $Y_{ij}^{(t)}$. In denoting this hypothesis, the suffix I stands for independence in the sense of absence of reciprocity. In the following, we will refer to this model as the null (or independence) model, denoted by M_I. Furthermore, reciprocity effects which do not vary with blocks may be obtained by assuming the following hypothesis:

$$H_C : \rho_{u_1u_2} = \rho, \quad u_1, u_2 = 1, \ldots, k,\ u_1 \leqslant u_2, \tag{6}$$

where the suffix C stands for constant reciprocity. In the following, we will refer to this model as the constant reciprocity model, denoted by M_C. The unconstrained model, based on non-homogeneous $\rho_{u_1u_2}$ parameters will be denoted by M_U. In the next section, we will present an approximate procedure to test for the above hypotheses.

3. Model inference

Let $\mathcal{U} = \{\boldsymbol{U}_i, i = 1, \ldots, n\}$ denote the overall set of latent variables in the model; based on the assumptions introduced so far, the observed network distribution is obtained by marginalizing out all these latent variables from the joint distribution of $\mathcal{Y}$ and $\mathcal{U}$. In particular, we have

$$p(\mathcal{Y}) = \sum_{\mathcal{U}} p(\mathcal{Y}, \mathcal{U}) = \sum_{\mathcal{U}} p(\mathcal{Y} \mid \mathcal{U}) p(\mathcal{U}), \tag{7}$$

where $\sum_{\mathcal{U}}$ denotes the sum over the support of $\mathcal{U}$ and

$$p(\mathcal{Y} \mid \mathcal{U}) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \prod_{t=1}^{T} p(y_{ij}^{(t)}, y_{ji}^{(t)} \mid U_i^{(t)} = u_i^{(t)}, U_j^{(t)} = u_j^{(t)}),$$
$$p(\mathcal{U}) = \prod_{i=1}^{n} \lambda_{u_i^{(t)}} \prod_{t=2}^{T} \pi_{u_i^{(t)} | u_i^{(t-1)}}.$$

Computation of the network distribution through (7) requires the evaluation of a sum over $k^{Tn(n-1)/2}$ terms that becomes quickly cumbersome as the number of nodes in the network (n) increases. Consequently, full maximum likelihood estimates are not achievable, apart from networks of a very limited size.

3.1. Variational inference

Let $\boldsymbol{\theta}$ denote the vector of all free model parameters collected in $\boldsymbol{\lambda}$, $\boldsymbol{\Pi}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\rho}$. When dealing with discrete latent variables, obtaining parameter estimates via an Expectation Maximization algorithm (EM - Dempster et al., 1977) is a quite popular approach. However, due to the computational difficulties in obtaining the marginal network distribution, posterior expectations of the complete data log-likelihood which are required for the EM algorithm are themselves intractable. In this article, following the approach suggested by Daudin et al. (2008) for static SBMs and by Yang et al. (2011) and Matias and Miele (2017) for the corresponding dynamic versions, we propose to derive parameter estimates by exploiting a VEM algorithm.

Let $p(\mathcal{U} \mid \mathcal{Y})$ denote the posterior distribution of $\mathcal{U}$ given the observed data $\mathcal{Y}$ and let $Q(\mathcal{U})$ denote its approximation. The VEM algorithm maximizes the following lower bound of the log-likelihood function:

$$\begin{aligned}\mathcal{J}(\boldsymbol{\theta}) &= \log p(\mathcal{Y}) - KL\left[Q(\mathcal{U}) \,||\, p(\mathcal{U} \mid \mathcal{Y})\right] \\ &= \log p(\mathcal{Y}) - \sum_{\mathcal{U}} Q(\mathcal{U}) \log \frac{Q(\mathcal{U})}{p(\mathcal{U} \mid \mathcal{Y})} \\ &= \log p(\mathcal{Y}) - \sum_{\mathcal{U}} Q(\mathcal{U}) \log Q(\mathcal{U}) + \sum_{\mathcal{U}} Q(\mathcal{U}) \log p(\mathcal{Y},\mathcal{U}) - \sum_{\mathcal{U}} Q(\mathcal{U}) \log p(\mathcal{Y}). \end{aligned} \tag{8}$$

Clearly, the first and the last term of equation (8) cancel out, so that

$$\mathcal{J}(\boldsymbol{\theta}) = \sum_{\mathcal{U}} Q(\mathcal{U}) \log p(\mathcal{Y},\mathcal{U}) - \sum_{\mathcal{U}} Q(\mathcal{U}) \log Q(\mathcal{U}), \tag{9}$$

and the evaluation of the intractable likelihood $p(\mathcal{Y})$ is not required any longer. As regards $Q(\mathcal{U})$, we focus on the class of approximate distributions assuming conditional independence between the latent variables in the network given the observed data, namely

$$Q(\mathcal{U}) = \prod_{i=1}^{n} \prod_{t=1}^{T} h\left(u_i^{(t)}; \boldsymbol{\tau}_i^{(t)}\right), \tag{10}$$

where $h(\cdot; \boldsymbol{\tau}_i^{(t)})$ denotes a multinomial probability distribution with parameters 1 and $\boldsymbol{\tau}_i^{(t)} = (\tau_{i1}^{(t)}, \ldots, \tau_{ik}^{(t)})'$. In this respect, equation (9) can be rewritten as

$$\mathcal{J}(\boldsymbol{\theta}) \;\; = \;\; \mathcal{J}_1(\boldsymbol{\theta}) + \mathcal{J}_2(\boldsymbol{\theta}) - \mathcal{J}_3(\boldsymbol{\theta}),$$

where

$$\begin{aligned}
\mathcal{J}_1(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\sum_{u=1}^{k}\tau_{iu}^{(1)}\log\lambda_u + \sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{u=1}^{k}\sum_{v=1}^{k}\tau_{iu}^{(t-1)}\tau_{iv}^{(t)}\log\pi_{v|u},\\
\mathcal{J}_2(\boldsymbol{\theta}) &= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{t=1}^{T}\sum_{u=1}^{k}\sum_{v=1}^{k}\tau_{iu}^{(t)}\tau_{jv}^{(t)}\log p(y_{ij}^{(t)}, y_{ji}^{(t)} \mid U_i^{(t)}=u, U_j^{(t)}=v),\\
\mathcal{J}_3(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{u=1}^{k}\tau_{iu}^{(t)}\log\tau_{iu}^{(t)}.
\end{aligned}$$

To obtain parameter estimates, the VEM algorithm alternates two separate steps until convergence: the E-step and the M-step.

The E-step consists in maximizing $\mathcal{J}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\tau}_i^{(t)}, i = 1,\ldots,n,\ t = 1,\ldots,T$, under the constraint that these quantities are non-negative and $\sum_u \tau_{iu}^{(t)} = 1$. This leads to the following updating rule:

$$\tau_{iu}^{(t)} \propto \exp\left[\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{v=1}^{k}\tau_{jv}^{(t)}\log p(y_{ij}^{(t)}, y_{ji}^{(t)} \mid U_i^{(t)}=u, U_j^{(t)}=v) + r_{iu}^{(t)}\right], \tag{11}$$

with

$$r_{iu}^{(t)} = \begin{cases} \log\lambda_u + \sum_{v=1}^{k}\tau_{iv}^{(2)}\log\pi_{v|u}, & t=1,\\ \sum_{s=1}^{k}\tau_{is}^{(t-1)}\log\pi_{u|s} + \sum_{v=1}^{k}\tau_{iv}^{(t+1)}\log\pi_{v|u}, & 1<t<T,\\ \sum_{s=1}^{k}\tau_{is}^{(T-1)}\log\pi_{u|s}, & t=T. \end{cases}$$

In equation (11), the symbol $\propto$ means that the quantities $\tau_{iu}^{(t)}$ are obtained by normalizing the expression at the right hand side.

At the M-step of the VEM algorithm, we maximize $\mathcal{J}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. In particular, we first maximize $\mathcal{J}_1(\boldsymbol{\theta})$ with respect to the initial and the transition probabilities, λ_u and $\pi_{v|u}$, of the hidden Markov process, under the constraints that these are all non-negative with $\sum_{u=1}^{k}\lambda_u = 1$ and $\sum_{v=1}^{k}\pi_{v|u} = 1$, $u = 1,\ldots,k$; the corresponding parameter estimates are updated as

$$\begin{aligned}
\lambda_u &= \frac{\sum_{i=1}^{n}\tau_{iu}^{(1)}}{n}, \quad u=1,\ldots,k,\\
\pi_{v|u} &= \frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\tau_{iu}^{(t-1)}\tau_{iv}^{(t)}}{\sum_{i=1}^{n}\sum_{t=2}^{T}\tau_{iu}^{(t-1)}}, \quad u,v=1,\ldots,k.
\end{aligned}$$

Moreover, to update $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\rho}$, we maximize $\mathcal{J}_2(\boldsymbol{\theta})$ with respect to these parameters. This requires the use of a Newton-Raphson algorithm having the same structure of a standard iterative algorithm used to estimate log-linear models for contingency tables. Obviously,

this algorithm needs to take into account if we are assuming the unconstrained model, M_U, in which the $\rho_{u_1u_2}$ parameters are free, or its constrained versions, M_I or M_C, under the hypotheses defined in (5) and (6), respectively.

The E- and the M-step of the VEM algorithm are iterated until convergence, that is, until the (relative) difference between subsequent approximate log-likelihood values $\mathcal{J}(\boldsymbol{\theta})$ is lower than an arbitrary small quantity $\epsilon > 0$, such as 10^{-10}.

3.1.1. Algorithm initialization

The initialization of the VEM algorithm plays a central role (see e.g., Vu et al., 2013; Matias and Miele, 2017). In fact, the target function is typical multimodal; this is a common problem in estimation of discrete latent variable models, even when a full log-likelihood function is used for estimation. Therefore, a multi-start strategy, based both on a deterministic and a random starting rule, is necessary.

A deterministic starting rule is based on firstly clustering nodes in the network via a k-means procedure and then obtaining parameter estimates accordingly. In this regard, Matias and Miele (2017) proposed to run a k-means algorithm on the rows of an extended data matrix obtained concatenating by columns all adjacency matrices $\boldsymbol{Y}^{(t)}$. In practice, this correspond to a time-constant clustering of the n nodes to initialize the algorithm. However, we found that a more efficient choice is considering a sort of time-varying clustering. To this purpose, we suggest to run a k-means algorithm on the rows of the extended data matrix obtained via row-concatenation of the adjacency matrices $\boldsymbol{Y}^{(t)}$, and derive model parameter estimates accordingly. As a result, we obtain the initial clustering of $n \times T$ observations, which takes into account the dynamic structure of the data.

In our applications, we experimented both deterministic initialization strategies described above and we found a clear improvement of the results, in terms of precision of estimates and clustering performance, with our proposal. Accordingly, for both the simulation study and the real data applications, we adopted the strategy based on row-concatenation of the adjacency matrices $\boldsymbol{Y}^{(t)}$ in order to obtain the initial estimates of the null model M_I. For the other specifications, M_U and M_C, we exploited the fact that models are nested and, therefore, the initial values of the corresponding parameters can be set equal to the final estimates of the M_I version.

Random starting solutions are obtained by considering a time-varying random assignment of the nodes in the network to one of the k distinct blocks of the dynamic SBM. As before, based on such a clustering, initial values for model parameters may be derived and the algorithm may be run until convergence.

Overall, for a given k, the solution that at convergence of the algorithm corresponds to the highest value of the approximate log-likelihood function $\mathcal{J}(\boldsymbol{\theta})$ is taken as the optimal estimate. Hereafter, the estimates obtained in this way are denoted by $\hat{\boldsymbol{\theta}}$.

3.1.2. Model selection and clustering

To select a suitable number of blocks k for the proposed dynamic SBM for a certain dataset, we rely on the Integrated Classification Likelihood (ICL) approach (Biernacki et al., 2000); see also Daudin et al. (2008) and Matias and Miele (2017). In our case, this criterion

relies on an index that, for a given k and corresponding parameter estimates $\hat{\boldsymbol{\theta}}$, is defined as

$$ICL = \log p(\mathcal{Y}, \hat{\mathcal{U}}) - \frac{k-1}{2}\log n - \frac{k(k-1)}{2}\log[n(T-1)] - \frac{k^2+g}{2}\log[n(n-1)T], \quad (12)$$

where $\hat{\mathcal{U}}$ denotes the optimal clustering of units (obtained as clarified below), computed at $\hat{\boldsymbol{\theta}}$. As regards the penalization terms in equation (12), the first and the second terms refer to the initial and the transition probabilities, respectively, whereas the last term corresponds to the conditional dyad probabilities, with k^2 being the number of free parameters in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and g being the number of free parameters in $\boldsymbol{\rho}$. More precisely, g is equal to $k(k+1)/2$, 1, and 0 under M_U, M_C, and M_I, respectively. According to the criterion at issue, the optimal number of latent blocks is the one corresponding to the maximum value of ICL.

An additional relevant issue when dealing with SBMs concerns the clustering of the observed units into blocks, which may be obtained by the variational inference scheme we propose. In particular, based on the estimated parameters of the multinomial distribution introduced in equation (10), that is, based on $\hat{\boldsymbol{\tau}}_i^{(t)}$, nodes may be assigned to one of the k blocks of the dynamic SBM according to a standard *maximum a posteriori* rule. Clustering performance of the proposed approach will be assessed in the simulation study discussed in Section 4.

3.1.3. Standard errors

Following an approach similar to that described by Vu et al. (2013), we obtain standard errors for the parameter estimates via parametric bootstrap (Efron, 1979). Once parameter estimates $\hat{\boldsymbol{\theta}}$ have been computed and the optimal number of hidden states (k) has been selected, we simulate B random replicates of the network $\mathcal{Y}$ from a dynamic SBM with k blocks and parameters $\hat{\boldsymbol{\theta}}$. For each dataset, a VEM algorithm is used to derive approximate maximum likelihood estimates. Let $\hat{\boldsymbol{\theta}}^{(b)}$ denote the vector of model parameter estimates for the b-th bootstrap sample. Standard error estimates for $\hat{\boldsymbol{\theta}}$ correspond to the square root of the diagonal elements of the matrix

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\boldsymbol{\theta}}^{(b)} - \hat{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}}^{(b)} - \hat{\boldsymbol{\theta}}\right)'.$$

3.2. Testing for reciprocity

It is already clear that reciprocity plays a central role when dealing with directed networks. To test for the presence of this characteristic, we propose an approximate LR test based on the lower bound of the likelihood function, $\mathcal{J}(\boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}}_I$ and $\hat{\boldsymbol{\theta}}_U$ denote the vectors of parameters estimated under models M_I and M_U, respectively. Recall that these two models correspond to the hypothesis H_I and H_C defined in equations (5) and (6), respectively. We propose a general test for the independence assumption H_I based on the following (approximate) LR test statistic

$$R_I = -2\big[\mathcal{J}(\hat{\boldsymbol{\theta}}_I) - \mathcal{J}(\hat{\boldsymbol{\theta}}_U)\big].$$

We compare the observed value of this test statistic against a χ^2 distribution with a number of degrees of freedom equal to the number of free parameters in $\boldsymbol{\rho}$, that is $k(k+1)/2$. In fact, we may consider R_I as an approximation of the LR statistic $-2\big[\ell(\hat{\boldsymbol{\theta}}_I) - \ell(\hat{\boldsymbol{\theta}}_U)\big]$ that, under suitable regularity conditions, has null asymptotic distribution of this type (Cox and Hinkley, 1979). Note that the difference between the exact LR statistic and R_I corresponds to the difference between two Kulback-Leibler distances of the type defined in (8) and we expect this difference to become negligible for n large.

For a more detailed analysis, we can also consider the decomposition of R_I as follows:

$$R_I = R_C + R_{CI},$$

where

$$R_C = -2\big[\mathcal{J}(\hat{\boldsymbol{\theta}}_C) - \mathcal{J}(\hat{\boldsymbol{\theta}}_U)\big]$$

is the approximate LR test statistic for testing the constant reciprocity assumption H_C against the hypothesis of unconstrained parameters, with $\hat{\boldsymbol{\theta}}_C$ denoting the vectors parameter estimates under this hypothesis. The statistic R_C is compared against a χ^2 distribution with $k(k+1)/2-1$ degrees of freedom. On the other hand,

$$R_{CI} = -2\big[\mathcal{J}(\hat{\boldsymbol{\theta}}_I) - \mathcal{J}(\hat{\boldsymbol{\theta}}_C)\big]$$

is the approximate LR test statistic for comparing the independence model M_I against the constant reciprocity model M_C, that is H_I against H_C. This test statistic is compared against a χ^2 distribution with one degree of freedom.

The simulation study and the real data applications we discuss in the following will illustrate the properties of the proposed approximate inferential procedures. The link between the LR test statistics and the model selection criteria will be also clarified. In fact, it might happen that the outcome of the test depends on the selected number of blocks. To avoid this ambiguity, we can first identify the optimal k under different model specifications and then, if different values of k are selected, compare the values of the corresponding test statistics.

4. Simulation study

In in this section, we illustrate the results of a large scale Monte Carlo simulation study. In this study, we focused both on the performance of the approximate LR statistics and on the clustering performance of the proposed approach. We considered several distinct experimental scenarios, based on different network sizes and different values of the reciprocity parameter.

4.1. Design

We randomly drew 1,000 samples from a two state ($k = 2$) dynamic SBM for $n = 20, 50, 100$ nodes observed at $T = 10$ different time occasions. The following values for the

initial probability vector and the transition probability matrix are assumed:

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

For the parameterization of the dyad probabilities, we set $\boldsymbol{\alpha} = (-2, -3, -1)'$ and $\beta_{12} = 0$. As regards reciprocity parameters, we considered both the independence model M_I, by setting $\rho_{u_1 u_2} = 0$, and the constant reciprocity model M_C, by setting $\rho_{u_1 u_2} = \rho$, $u_1, u_2 = 1, \ldots, k$. In this latter case, we considered a grid of values for ρ, ranging between -2.5 and 2.5. To give an idea of how the matrix $\boldsymbol{\Psi}(u_1, u_2)$ looks like by fixing the above values for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and ρ, we report in Table 1 this matrix for some choices of ρ.

Table 1: *Matrix of conditional dyad probabilities* $\boldsymbol{\Psi}(u_1, u_2)$ *for varying choices of* ρ.

		$u_1 = 1, u_2 = 1$		$u_1 = 1, u_2 = 2$		$u_1 = 2, u_2 = 2$	
ρ		0	1	0	1	0	1
-2.5	0	0.786	0.106	0.909	0.045	0.572	0.211
	1	0.106	0.001	0.045	0.000	0.211	0.006
0.0	0	0.776	0.105	0.907	0.045	0.534	0.197
	1	0.105	0.014	0.045	0.002	0.197	0.072
2.5	0	0.669	0.091	0.885	0.044	0.295	0.109
	1	0.091	0.149	0.044	0.027	0.109	0.487

As is clear from Table 1, when ρ increases the chance of observing mutual relations increases, with the result that the adjacency matrices $\boldsymbol{Y}^{(t)}$ become progressively less sparse.

4.2. Results: LR test statistic performance

To evaluate the performance of the proposed approach, for each simulated scenario we considered the distribution of the approximate LR test statistics R_I and R_{CI}, which allow us to compare the independence model (M_I) against the unconstrained model (M_U) and the constant reciprocity model (M_C), respectively.

Table 2 reports the simulation results for the dynamic SBM under the different experimental scenarios in terms of mean and variance of the test statistic R_I. Furthermore, the table shows the value of $p = p(R_I > \chi^2_{0.05,3})$ that corresponds to the probability of the type I error when we simulate the data under the model M_I ($\rho = 0$), while it corresponds to the power of the test when data are simulated under the model M_C ($\rho \neq 0$). Similar results are reported in Table 3 for the test statistic R_{CI}. In this case, the probability of the type I error and the power of the test are obtained by simulation considering an asymptotic χ^2 distribution with 1 degree of freedom only, that is, as $p = p(R_{CI} > \chi^2_{0.05,1})$.

The results confirm our conjecture that, when simulating data from the model M_I, both approximate test statistics have a distribution reasonably close to a χ^2 distribution, leading to the rejection of H_I in about 5% of the simulated samples. On the other hand, under the homogeneity assumption for the reciprocity effects, we observe that the power of the test

Table 2: *Mean* ($\bar{R}_I$)*, variance* ($Var(R_I)$)*, and estimated probability of type I error/power of the test statistic* R_I (p) *under different scenarios.*

	$n = 20$			$n = 50$			$n = 100$		
ρ	$\bar{R}_I$	Var(R_I)	p	$\bar{R}_I$	Var(R_I)	p	$\bar{R}_I$	Var(R_I)	p
-2.50	61.15	443.79	1.000	390.59	1949.63	1.000	1581.82	12797.59	1.000
-1.50	35.76	200.22	0.993	229.01	1167.33	1.000	922.82	5912.34	1.000
-1.00	21.15	109.27	0.975	129.17	557.34	1.000	523.86	2605.36	1.000
-0.75	13.45	56.39	0.923	82.73	369.29	1.000	333.29	1592.40	1.000
-0.50	7.53	28.92	0.737	42.13	178.39	1.000	167.06	728.66	1.000
-0.25	2.89	9.25	0.272	12.71	49.36	0.922	47.97	192.24	1.000
-0.10	1.40	3.10	0.109	3.06	10.42	0.308	9.29	36.08	0.826
0.00	1.00	2.02	0.052	0.93	1.83	0.045	1.02	2.07	0.052
0.10	1.35	4.56	0.085	3.14	10.40	0.300	9.43	35.60	0.835
0.25	3.15	13.70	0.297	14.68	57.46	0.957	57.44	260.67	1.000
0.50	10.65	46.61	0.861	62.66	275.15	1.000	251.68	1130.22	1.000
0.75	24.68	114.21	1.000	152.31	752.48	1.000	613.80	3461.35	1.000
1.00	45.71	220.82	1.000	293.87	1574.97	1.000	1180.30	7741.51	1.000
1.50	114.84	598.56	1.000	736.43	4668.70	1.000	2977.89	29305.76	1.000
2.50	342.95	1806.82	1.000	2191.43	13394.32	1.000	8893.55	81455.09	1.000

Table 3: *Mean* ($\bar{R}_{CI}$)*, variance* ($Var(R_{CI})$)*, and estimated probability of type I error/power of the test statistic* R_{CI} (p) *under different scenarios.*

	$n = 20$			$n = 50$			$n = 100$		
ρ	$\bar{R}_{CI}$	Var(R_{CI})	p	$\bar{R}_{CI}$	Var(R_{CI})	p	$\bar{R}_{CI}$	Var(R_{CI})	p
-2.50	62.88	441.33	1.000	392.93	1958.23	1.000	1583.89	12813.85	1.000
-1.50	38.11	200.09	0.988	231.12	1177.18	1.000	924.85	5930.55	1.000
-1.00	23.67	117.15	0.959	131.21	563.74	1.000	525.90	2617.52	1.000
-0.75	15.88	61.07	0.868	84.92	375.93	1.000	335.29	1588.78	1.000
-0.50	10.21	37.30	0.601	44.10	182.99	1.000	169.01	733.51	1.000
-0.25	5.42	14.42	0.227	14.77	54.87	0.825	49.89	194.19	1.000
-0.10	3.89	9.24	0.095	5.11	14.46	0.203	11.27	40.50	0.679
0.00	3.51	7.52	0.078	3.00	6.04	0.055	2.94	6.73	0.051
0.10	3.74	9.15	0.086	5.11	15.13	0.210	11.53	40.09	0.703
0.25	5.76	19.62	0.244	16.55	60.25	0.883	59.50	268.90	1.000
0.50	13.06	53.14	0.743	64.70	281.35	1.000	253.65	1128.36	1.000
0.75	26.99	119.26	0.991	154.22	754.42	1.000	615.73	3476.57	1.000
1.00	47.91	224.45	1.000	296.14	1623.78	1.000	1182.33	7749.08	1.000
1.50	116.90	603.89	1.000	738.37	4665.33	1.000	2979.98	29281.81	1.000
2.50	345.05	1816.17	1.000	2193.41	13414.52	1.000	8895.61	81449.12	1.000

increases as much as ρ deviates from 0. Moreover, the power of the test increases as the sample size n increases.

4.3. Results: clustering performance

We also explored the performance of the proposed method for clustering units across time. For this aim, we evaluated the agreement between the estimated and the true latent

structure in terms of Adjusted Rand Index (Hubert and Arabie, 1985), which is denoted by ARI hereafter. This index is obtained from the standard Rand Index (Rand, 1971), which varies between 0 and 1. The main difference is that ARI may also attain negative values, while retaining the same upper bound equal to 1 when there is a perfect agreement between the true and estimated clustering structure. Moreover, it is equal to 0 when the number of nodes correctly classified is equal to that expected by chance.

Table 4 reports the results we obtained in terms of average ARI over the simulated samples under the different model specifications reflecting varying degrees of reciprocity.

Table 4: *Average ARI under different scenarios and the different estimated models.*

	$n = 20$			$n = 50$			$n = 100$		
ρ	M_I	M_C	M_U	M_I	M_C	M_U	M_I	M_C	M_U
-2.50	0.5395	0.5840	0.5838	0.9694	0.9694	0.9695	0.9984	0.9984	0.9984
-1.50	0.6107	0.6512	0.6508	0.9711	0.9723	0.9721	0.9988	0.9989	0.9988
-1.00	0.6845	0.7055	0.7034	0.9786	0.9789	0.9786	0.9990	0.9990	0.9991
-0.75	0.6678	0.6685	0.6690	0.9801	0.9803	0.9803	0.9991	0.9992	0.9991
-0.50	0.7055	0.7063	0.7025	0.9787	0.9788	0.9789	0.9994	0.9995	0.9995
-0.25	0.7417	0.7423	0.7423	0.9841	0.9842	0.9842	0.9998	0.9998	0.9998
-0.10	0.7465	0.7460	0.7468	0.9844	0.9844	0.9846	0.9998	0.9997	0.9998
0.00	0.7776	0.7766	0.7749	0.9872	0.9871	0.9873	0.9998	0.9998	0.9998
0.10	0.7775	0.7767	0.7722	0.9887	0.9889	0.9891	0.9998	0.9998	0.9998
0.25	0.8059	0.8083	0.8054	0.9878	0.9883	0.9881	0.9998	0.9998	0.9998
0.50	0.8295	0.8331	0.8353	0.9934	0.9934	0.9934	0.9999	0.9999	0.9999
0.75	0.8246	0.8291	0.8312	0.9951	0.9951	0.9950	1.0000	1.0000	1.0000
1.00	0.8742	0.8817	0.8788	0.9962	0.9962	0.9961	1.0000	1.0000	1.0000
1.50	0.8999	0.9074	0.9062	0.9988	0.9988	0.9987	1.0000	1.0000	1.0000
2.50	0.9722	0.9731	0.9727	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

On the basis of these results, we conclude that the estimation algorithm performs properly in clustering nodes, with a better behavior (corresponding to an increasing value of the average ARI) for higher values of the reciprocity parameter. This is an expected result, since a larger value of ρ corresponds to a larger probability of mutual relations and, accordingly, to a less sparse network. The lowest values of the index are observed for $n = 20$ and $\rho = -2.5$ although they are always greater than 0.5. Obviously, the best results are obtained when $n = 100$, with an average ARI which is close to 1 under all model specifications.

Results reported in Table 4 do not show substantial differences between models M_I, M_C, and M_U in terms of clustering, besides the true model specification. Accordingly, to better investigate how reciprocity affects the clustering accuracy, we considered a further simulation set up. This is characterized by a less neat clustering structure in the data and by a reduced amount of information. In particular, we considered $n = 20$ nodes and closer values for the tendency parameters $\boldsymbol{\alpha}$, that is, $\boldsymbol{\alpha} = (-2.5, -3, -1.5)'$. We also focused on large (absolute) values for the reciprocity effect: $\rho = -2.5, -1.5, 1.5, 2.5$. The average ARI for the different simulation scenarios is reported in Table 5. Comparing these results with those discussed so far, we clearly observe lower performance of algorithm in recovering the true clustering when the blocks are not well separated. Furthermore, under this latter scenario, differences

between M_I, M_C, and M_U become more evident, especially in the presence of more sparse data matrices (smaller ρ). As expected, we obtain better results in terms of classification under M_C and M_U with respect to M_I. This finding clearly show that the true reciprocity structure in the data may affect the clustering performance, especially when the separation between blocks in the simulated model is less neat.

Table 5: *Average ARI under the additional scenarios with* $n = 20$ *and* $\boldsymbol{\alpha} = (-2.5, -3, -1.5)'$ *and the different estimated models.*

ρ	M_I	M_C	M_U
-2.5	0.1199	0.1838	0.1887
-1.5	0.1368	0.2038	0.1948
1.5	0.6216	0.6470	0.6423
2.5	0.8064	0.8227	0.8157

To further explore the performance of the proposed approach, we also set up a comparison with the alternative inferential procedures for dynamic SBMs available in literature. In particular, we considered the MCMC method developed by Yang et al. (2011) in the Bayesian framework and the variational approach of Matias and Miele (2017). Estimation algorithms for these approaches are freely available: the R code for the variational approach of Matias and Miele (2017) is available at the CRAN website (`https://cran.r-project.org/web/packages/dynsbm/index.html`); the MATLAB code for the MCMC approach of Yang et al. (2011) is available on the author's website (i.e. `http://homepage.cs.uiowa.edu/~tyng/codes/`).

To make the comparison, we simulated data according to the design introduced in Section 4.1, focusing on a subset of values for the reciprocity parameter, that is, $\rho = -2.5, -1.5, 0, 1.5, 2.5$. It is worth recalling that both competing methods are based on the independence assumption between response variables $Y_{ij}^{(t)}$ and thus they do not allow for reciprocity. In this regard, when simulating data with $\rho = 0$, we are estimating the true model and we are directly able to compare the quality of results of the proposed approach to those of the main alternatives. On the contrary, when $\rho \neq 0$, we aim at evaluating how the lack of reciprocity affects the clustering accuracy of such alternatives. The quality of results is again assessed by considering the average ARI across simulations; these results are reported in Table 6. In this table, we also report the average ARI obtained under the proposed M_C and M_U specifications appearing in Table 4 to facilitate the comparison.

Comparing the results obtained when adopting the approach of Matias and Miele (2017) and Yang et al. (2011) with those derived when estimating M_C and M_U, we observe that the proposed approach always outperforms the alternatives, especially when n and ρ are small. The superiority of the proposed approach with respect to the competitors is evident when reciprocity plays a role in the simulation of the data, but also when considering the independence model ($\rho = 0$). As regards the approach suggested by Matias and Miele (2017), these findings may be due, at least partially, to the initialization strategy adopted by the authors which, as stated in Section 3.1.1, does not admit an initial time-varying clustering of the units. As already mentioned, we tried to implement a similar strategy for our approach

Table 6: *Average ARI for the approach of Matias and Miele (2017) (MM) and of Yang et al. (2011) (Y) under different scenarios.*

	$n = 20$				$n = 50$				$n = 100$			
ρ	M_C	M_U	MM	Y	M_C	M_U	MM	Y	M_C	M_U	MM	Y
-2.5	0.5840	0.5838	0.0770	-0.0078	0.9694	0.9695	0.2805	0.5669	0.9984	0.9984	0.9461	0.9549
-1.5	0.6512	0.6508	0.0877	-0.0091	0.9723	0.9721	0.3440	0.6730	0.9989	0.9988	0.9596	0.9638
0	0.7766	0.7749	0.1304	-0.0005	0.9871	0.9873	0.8015	0.8723	0.9998	0.9998	0.9859	0.9886
1.5	0.9074	0.9062	0.5098	0.6331	0.9988	0.9987	0.9718	0.9857	1.0000	1.0000	0.9998	0.9997
2.5	0.9731	0.9727	0.9178	0.9256	1.0000	1.0000	0.9725	0.9996	1.0000	1.0000	0.9887	1.0000

and empirical findings showed a poor performance of the algorithm in recovering the true clustering. As regards the MCMC approach of Yang et al. (2011), the low quality of the results may be possibly due to an inadequate choice of the prior.

In conclusion, it is worth to notice that the gap in terms of clustering accuracy of our proposal with respect to the alternative approaches is mostly evident for smaller values of ρ. As stated before, these lead to a more sparse data matrix. Therefore, we may conclude that, unlike the alternatives, our proposal permits to properly recognize the clustering structure of the data, even when dealing with sparse data.

5. Empirical applications

In this section, we describe the application of the proposed methodology to two benchmark datasets in the network literature: the Newcomb Fraternity network and the Enron email network.

5.1. Newcomb Fraternity network

The network at issue, described in Newcomb (1961), consists of 14 network snapshots on preference rankings (coded from 1 to 16) from 17 students. Data were collected longitudinally over 15 weeks between 1953 and 1956 (although data from week 9 are missing) among students living in an off-campus (fraternity) house at the University of Michigan. For the purpose of the analysis, we considered the binary socio-matrices $\boldsymbol{Y}^{(t)}$ derived from these data which are freely available as part of the R package `networkDynamic` (Butts et al., 2016). In each network snapshot, $Y_{ij}^{(t)} = 1$ if student i states a ranking for student j equal to 8 or less at time occasion t; $Y_{ij}^{(t)} = 0$ otherwise.

In such a context, we are interested in measuring the degree of cohesion of the network and characterizing relations between students. In this respect, we estimated the proposed dynamic SBM with $k = 1, \ldots, 5$, considering the different model specifications illustrated in Section 2.2 and corresponding to different hypotheses on reciprocity.

As stated in Section 3.1.1, for each k, we run a k-means algorithm on the rows of the extended data matrix obtained via row-concatenation of the adjacency matrices $\boldsymbol{Y}^{(t)}$. To reduce the risk of spurious solutions due to local maxima, we considered 100 random initializations of this algorithm. The parameters obtained from the resulting partitions were used as initial values for the VEM algorithm.

Table 7 reports the value of the approximate log-likelihood function obtained at convergence of the VEM algorithm, $\mathcal{J}(\hat{\boldsymbol{\theta}})$, and the value of ICL used to select the optimal number of latent blocks. From these results, we observe that the ICL criterion leads to selecting $k = 3$ latent blocks, regardless the chosen model specification. Furthermore, when using this criterion also for identifying the optimal model specification, the results reported in Table 7 confirm a better fit of M_C with respect to both M_I and M_U.

Table 7: *Newcomb Fraternity data. Value of $\mathcal{J}(\hat{\boldsymbol{\theta}})$ at convergence and ICL for different choices of k and different model specifications.*

		Number of Hidden States (k)				
		1	2	3	4	5
$\mathcal{J}(\cdot)$	M_I	-2639.50	-2479.92	-2124.15	-2108.67	-2048.43
	M_C	-2575.21	-2285.40	-2087.65	-2066.32	-2019.02
	M_U	-2575.21	-2282.19	-2083.08	-2048.86	-1999.98
ICL	M_I	-2643.63	-2505.36	-2181.28	-2214.34	-2212.51
	M_C	-2583.46	-2313.45	-2148.91	-2175.67	-2187.71
	M_U	-2583.46	-2318.44	-2165.19	-2195.14	-2225.80

To verify whether model M_C is truly optimal, when compared to M_I and M_U, we considered the approximate LR test introduced in Section 3.2 for the chosen number of latent blocks ($k = 3$). Results are reported in Table 8. In such a context, the observed values of the test statistics R_I, R_{CI}, and R_C have to be compared against a χ^2 distribution with $k(k+1)/2 = 6$, 1, and $k(k+1)/2 - 1 = 5$ degrees of freedom, respectively.

Table 8: *Newcomb Fraternity data. Approximate LR test statistic for different choices of k.*

	Number of Hidden States (k)				
	1	2	3	4	5
R_I	128.59	94.11	82.15	109.26	70.03
R_{CI}	128.59	73.69	73.00	91.63	41.33
R_C	0.00	20.42	9.16	17.63	28.70

Considering the results reported in Table 8 with $k = 3$, we may observe that the LR test statistic R_I is statistically significant ($p(\chi^2_6 > R_I) < 0.001$) and, therefore, leads to prefer M_U with respect to M_I. A significant test statistic is also observed when comparing M_I against M_C: $p(\chi^2_1 > R_{CI}) < 0.001$. On the other hand, considering the observed value of R_C, we conclude that the assumption of constant reciprocity, H_C, cannot be rejected based on the observed data as $p(\chi^2_5 > R_C) = 0.102$. These findings are therefore in line with the results based on the comparison of the ICL values reported in Table 7.

The estimated parameters under this model specification are shown in the following. In particular, Table 9 reports the estimated value of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and ρ together with the corresponding standard errors obtained via the parametric bootstrap procedure illustrated in Section 3.1.3.

Table 9: *Newcomb Fraternity data. Estimates and estimated standard errors (se) for the* $\boldsymbol{\alpha}, \boldsymbol{\beta}, \rho$ *parameters under the constant reciprocity assumption for the dynamic SBM with* $k = 3$ *hidden states.*

	$\hat{\alpha}_{11}$	$\hat{\alpha}_{12}$	$\hat{\alpha}_{13}$	$\hat{\alpha}_{22}$	$\hat{\alpha}_{23}$	$\hat{\alpha}_{33}$	$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{23}$	$\hat{\rho}$
Estimates	1.203	-2.055	-1.103	-1.069	-0.718	1.738	2.760	0.737	-1.151	1.044
SE	0.211	0.162	0.149	0.120	0.154	3.450	0.208	0.208	0.198	0.114

Table 10: *Newcomb Fraternity data. Estimates for the matrix of conditional dyad probabilities* $\hat{\boldsymbol{\Psi}}(u_1, u_2)$.

	$u_1 = 1, u_2 = 1$		$u_1 = 1, u_2 = 2$		$u_1 = 1, u_2 = 3$	
	0	1	0	1	0	1
0	0.026	0.085	0.257	0.520	0.373	0.259
1	0.085	0.804	0.033	0.189	0.124	0.244
	$u_1 = 2, u_2 = 2$		$u_1 = 2, u_2 = 3$		$u_1 = 3, u_2 = 3$	
	0	1	0	1	0	1
0	0.495	0.170	0.539	0.083	0.010	0.055
1	0.170	0.166	0.263	0.115	0.055	0.881

The result in Table 9 suggest the presence of significant mutual relations between students, irrespective to the cluster they belong to ($\hat{\rho} = 1.044$). Regarding the remaining parameters, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, interpretation may be based on the description provided in Section 2.2. Just to give an idea, according to the results in Table 9, students in block 1 are likely to declare a non-reciprocated friendship with nodes belonging to the same block ($\hat{\alpha}_{11} = 1.203$), while null within-group relations are mainly observed for students belonging to block 2 ($\hat{\alpha}_{22}$ is negative). A non-significant effect is observed for $\hat{\alpha}_{33}$. On the other hand, $\beta_{u_1 u_2}$ may be interpreted in terms of asymmetric relations between nodes in blocks u_1 and u_2; see equation (4). So, for instance, the value of $\hat{\beta}_{12}$ leads to the conclusion that the tendency of students in the second block to have an asymmetric relation with students in the first block is significantly higher than the opposite relation. We also report in Table 10 the estimated matrix of conditional dyad probabilities corresponding to the above parameters.

To conclude, Table 11 shows the estimated initial and transition probabilities of the hidden Markov chain. As can be observed, cluster 2 is the most likely at the beginning of the observation period ($\hat{\lambda}_2 = 0.48$). Moreover, estimated transitions show quite persistent hidden states, with students' memberships that mainly remain unchanged during time.

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

	Estimates				SE			
u	$\hat{\lambda}_u$	$\hat{\pi}_{1\|u}$	$\hat{\pi}_{2\|u}$	$\hat{\pi}_{3\|u}$	$\hat{\lambda}_u$	$\hat{\pi}_{1\|u}$	$\hat{\pi}_{2\|u}$	$\hat{\pi}_{3\|u}$
1	0.348	1.000	0.000	0.000	0.116	0.000	0.000	0.000
2	0.476	0.014	0.959	0.027	0.132	0.017	0.030	0.023
3	0.176	0.000	0.000	1.000	0.090	0.030	0.070	0.100

5.2. Enron email network

The second example is based on a dynamic network derived from the Enron corpus (Priebe et al., 2005), consisting of a large set of email messages that was made public during the legal investigation concerning the Enron corporation.

The dataset [1] concerns 184 Enron employees; we considered communications recorded between April, 2001, and March, 2002, and we built an email network for each month, so that the dynamic network has 12 time points. Clearly, this represents a simplification of the data structure as information about the exact time the emails were sent are available. However, for illustrative purposed, we decided to analyze the network on discrete time. 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 analyzed time window, with $i = 1, \ldots, 183$, $j = i+1, \ldots, 184$ and $t = 1, \ldots, 12$. Following Xu and Hero (2014) we made no distinction between emails sent "to", "cc", or "bcc".

The interest is in understanding the evolution of dyadic relations between users (email exchange) over time, and analyzing reciprocity between these relations. We are also interested in defining groups characterized by similar communication profiles. To this extent, we estimated a dynamic SBM with a varying number of blocks ($k = 1, \ldots, 7$), by adopting the initialization strategy described in Section 3.1.1. Results are reported in Table 12. In this application, ICL values lead to selecting a model with $k = 6$ hidden states for all considered parameterizations. Based on the same index, we select the unconstrained model M_U, with reciprocity parameters depending on the latent blocks.

Table 12: *Enron data. Value of $\mathcal{J}(\hat{\boldsymbol{\theta}})$ at convergence and ICL for different choices of k and different model specifications.*

		Number of Hidden States (k)						
		1	2	3	4	5	6	7
$\mathcal{J}(\cdot)$	M_I	-28824.73	-24013.02	-22479.96	-21477.72	-20598.79	-19843.34	-19706.03
	M_C	-25215.17	-21745.57	-20652.19	-19957.50	-19340.35	-18619.62	-18577.28
	M_U	-25215.17	-21704.47	-20546.47	-19695.55	-19132.86	-18429.46	-18356.60
ICL	M_I	-28831.19	-24094.53	-22611.46	-21688.13	-20900.32	-20257.84	-20268.75
	M_C	-25228.08	-21840.56	-20794.61	-20183.30	-19655.20	-19074.35	-19142.58
	M_U	-25228.08	-21816.33	-20724.88	-19973.27	-19527.97	-19010.62	-19099.33

Even in this case, we may validate this result by comparing the values of the approximate LR statistics introduced in Section 3.2. In Table 13, we report the observed values of R_I, R_{CI}, and R_C, which have to be compared against a χ^2 distribution with 21, 1, and 20 degrees of freedom, respectively. From this comparison, when $k = 6$, we observe that the hypothesis of absence of reciprocity, H_I, is strongly rejected by both tests based on R_I and R_{CI}. Moreover, the observed value of the test statistic R_C allows us to confirm that the unconstrained model has to be preferred to the other model specifications, due to a very low

[1] Available at `http://cis.jhu.edu/~parky/Enron/enron.html`

p-value: $p(\chi^2_{20} > R_C) < 0.001$. Accordingly, in this application, we conclude that reciprocal relations are statistically significant, and that they depend on the latent blocks of the nodes. The corresponding estimated values of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\rho}$ are reported in Table 14, whereas the estimated initial and transition probabilities are displayed in Table 15. These estimates may be interpreted as illustrated for the Newcomb Fraternity network.

6. Concluding remarks

In this paper, we introduce a class of stochastic block models for dynamic networks where the standard hypothesis of independence between univariate responses is relaxed in favor of less stringent assumptions. In particular, the element of analysis is the set of dyads referred to ordered pairs of units and the assumption of conditional independence between them is considered. Obviously, marginal dependence, due for instance to triangulation effects, is not ruled out but can be, instead, explained in a meaningful way by the latent variables.

Reciprocity plays a central role in this paper. In particular, we propose to parametrically specify dyad probabilities with the aim of measuring mutual effects. By properly specifying the reciprocity structure, we may obtain different parameterizations of the model and, thus, allow for different types of cohesion between nodes. Non-homogeneous, homogeneous, and null reciprocity effects are the possible specifications we can consider.

For this class of models, computation of the full likelihood to obtain parameter estimates becomes progressively unfeasible as the dimension of the network increases. For this reason, we propose to make inference by a variational approximation of the intractable full likelihood function, based on the assumption of posterior independence between the dyads.

We show how the approximate likelihood function, which is exploited to derive parameter estimates, also offers the possibility to define an approximate LR statistic. This allows us to compare different model specifications by means of a proper hypothesis testing procedure. Empirical findings confirm that the χ^2 distribution properly approximate the true distribution of such a test statistic.

The behavior of the proposed approach is evaluated by means of a large scale simulation study, in which different experimental scenarios are considered. Results show very good performance of the proposed approximate LR test in identifying the true data generating process and the effectiveness of our approach, in general, in recovering the true data structure in terms of clustering. The analysis of the Newcomb Fraternity and the Enron datasets provides further insights into the proposed method.

Table 13: *Enron data. Approximate LR test statistic for different choices of k and different model specifications.*

	Hidden States k						
	1	2	3	4	5	6	7
R_I	7219.13	4617.11	3866.98	3564.33	2931.87	2827.76	2698.86
R_{CI}	7219.13	4534.91	3655.54	3040.43	2516.88	2447.45	2257.49
R_C	0.00	82.20	211.45	523.90	414.99	380.31	441.37

Table 14: *Enron data. Estimates for the* $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\rho}$ *parameters for the dynamic SBM with* $k = 6$ *hidden states.*

$\hat{\alpha}_{11}$	-1.836	$\hat{\alpha}_{26}$	-3.955	$\hat{\alpha}_{66}$	-1.608	$\hat{\beta}_{34}$	12.207	$\hat{\rho}_{15}$	3.708	$\hat{\rho}_{36}$	4.284
$\hat{\alpha}_{12}$	-5.163	$\hat{\alpha}_{33}$	-9.388	$\hat{\beta}_{12}$	-0.775	$\hat{\beta}_{35}$	1.854	$\hat{\rho}_{16}$	3.116	$\hat{\rho}_{44}$	1.971
$\hat{\alpha}_{13}$	-6.652	$\hat{\alpha}_{34}$	-19.373	$\hat{\beta}_{13}$	-1.128	$\hat{\beta}_{36}$	12.949	$\hat{\rho}_{22}$	3.86	$\hat{\rho}_{45}$	3.442
$\hat{\alpha}_{14}$	-6.509	$\hat{\alpha}_{35}$	-7.839	$\hat{\beta}_{14}$	0.007	$\hat{\beta}_{45}$	0.932	$\hat{\rho}_{23}$	6.43	$\hat{\rho}_{46}$	6.546
$\hat{\alpha}_{15}$	-4.501	$\hat{\alpha}_{36}$	-17.269	$\hat{\beta}_{15}$	0.628	$\hat{\beta}_{46}$	2.431	$\hat{\rho}_{24}$	11.32	$\hat{\rho}_{55}$	1.661
$\hat{\alpha}_{16}$	-4.056	$\hat{\alpha}_{44}$	-1.588	$\hat{\beta}_{16}$	0.366	$\hat{\beta}_{56}$	0.136	$\hat{\rho}_{25}$	5.157	$\hat{\rho}_{56}$	2.399
$\hat{\alpha}_{22}$	-4.124	$\hat{\alpha}_{45}$	-4.962	$\hat{\beta}_{23}$	-2.627	$\hat{\rho}_{11}$	2.476	$\hat{\rho}_{26}$	2.584	$\hat{\rho}_{66}$	2.611
$\hat{\alpha}_{23}$	-6.323	$\hat{\alpha}_{46}$	-6.934	$\hat{\beta}_{24}$	-1.730	$\hat{\rho}_{12}$	4.834	$\hat{\rho}_{33}$	4.007		
$\hat{\alpha}_{24}$	-20.06	$\hat{\alpha}_{55}$	-1.945	$\hat{\beta}_{25}$	0.135	$\hat{\rho}_{13}$	-2.671	$\hat{\rho}_{34}$	17.07		
$\hat{\alpha}_{25}$	-5.910	$\hat{\alpha}_{56}$	-2.500	$\hat{\beta}_{26}$	2.180	$\hat{\rho}_{14}$	5.414	$\hat{\rho}_{35}$	5.580		

Table 15: *Enron data. Estimates for the latent Markov model parameters of the dynamic SBM with* $k = 6$ *hidden states.*

u	$\hat{\lambda}_u$	$\hat{\pi}_{1\|u}$	$\hat{\pi}_{2\|u}$	$\hat{\pi}_{3\|u}$	$\hat{\pi}_{4\|u}$	$\hat{\pi}_{5\|u}$	$\hat{\pi}_{6\|u}$
1	0.106	0.842	0.000	0.117	0.000	0.041	0.000
2	0.300	0.003	0.836	0.122	0.000	0.005	0.034
3	0.384	0.004	0.075	0.886	0.008	0.026	0.000
4	0.056	0.000	0.000	0.020	0.973	0.007	0.000
5	0.111	0.018	0.000	0.193	0.005	0.738	0.045
6	0.043	0.000	0.291	0.081	0.000	0.163	0.464

An interesting evolution of the approach here illustrated may be in a exponential random graph (Wang et al., 2009) perspective. This parameterization would allow us to further improve model flexibility, explicitly accounting for triangulations and higher order dependencies between nodes, while providing a clustering of nodes in terms of their social behavior. It would be also possible to allow for categorical responses and for the inclusion of individual covariates with the aim at identifying clustering determinants.

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

Agresti, A., 2013. Categorical Data Analysis, 3rd Edition. Wiley-Interscience, Hoboken, New Jersey.

Amati, V., Lomi, A., Mira, A., 2018. Social network modeling. Annual Review of Statistics and Its Application 5 (1), null.

Ambroise, C., Matias, C., 2012. New consistent and asymptotically normal parameter estimates for random-graph mixture models. Journal of the Royal Statistical Society: Series B 74, 3–35.

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

Biernacki, C., Celeux, G., Govaert, G., 2000. Assessing a mixture model for clustering with the integrated completed likelihood. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 719–725.

Butts, C. T., 2008. A relational event framework for social action. Sociological Methodology 38, 155–200.

Butts, C. T., Leslie-Cook, A., Krivitsky, P. N., Bender-deMoll, S., 2016. networkDynamic: Dynamic Extensions for Network Objects. R package version 0.9.0.
URL `https://CRAN.R-project.org/package=networkDynamic`

Cox, D. R., Hinkley, D. V., 1979. Theoretical statistics. CRC Press.

Daudin, J.-J., Picard, F., Robin, S., 2008. A mixture model for random graphs. Statistics and Computing 18, 173–183.

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

Diggle, P., Heagerty, P., Liang, K.-Y., Zeger, S., 2002. Analysis of Longitudinal Data. Oxford University Press.

Durante, D., Dunson, D. B., 2014. Nonparametric Bayes dynamic modelling of relational data. Biometrika 101, 883–898.

Efron, B., 1979. Bootstrap methods: Another look at the jackknife. Annals of Statistics 7, 1–26.

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

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

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

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

Holland, P. W., Laskey, K. B., Leinhardt, S., 1983. Stochastic blockmodels: first steps. Social Networks 5, 109–137.

Holland, P. W., Leinhardt, S., 1976. Local structure in social networks. Sociological Methodology 7, 1–45.

Hubert, L., Arabie, P., 1985. Comparing partitions. Journal of Classification 2, 193–218.

Lee, N., Priebe, C., 2011. A latent process model for time series of attributed random graphs. Statistical Inference for Stochastic Processes 14, 231–253.

Lusher, D., Koskinen, J., Robins, G., 2013. Exponential random graph models for social networks: theory, methods, and applications. Cambridge University Press.

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

Newcomb, T. M., 1961. The acquaintance process. Holt, Rinehart & Winston.

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

Priebe, C. E., Conroy, J. M., Marchette, D. J., Park, Y., 2005. Scan statistics on Enron graphs. Computational & Mathematical Organization Theory 11, 229–247.

Quintane, E., Conaldi, G., Tonellato, M., Lomi, A., 2014. Modeling relational events a case study on an open source software project. Organizational Research Methods 17, 23–50.

Rand, W. M., 1971. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical association 66 (336), 846–850.

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

Sarkar, P., Moore, A. W., 2006. Dynamic social network analysis using latent space models. In: Advances in Neural Information Processing Systems. pp. 1145–1152.

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

Snijders, T. A., 1996. Stochastic actor-oriented models for network change. Journal of Mathematical Sociology 21, 149–172.

Snijders, T. A., 2001. The statistical evaluation of social network dynamics. Sociological Methodology 31, 361–395.

Snijders, T. A., 2005. Models for longitudinal network data. Models and Methods in Social Network Analysis 1, 215–247.

Snijders, T. A., Nowicki, K., 1997. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. Journal of Classification 14, 75–100.

Snijders, T. A., Van de Bunt, G. G., Steglich, C. E., 2010. Introduction to stochastic actor-based models for network dynamics. Social Networks 32, 44–60.

Vu, D. Q., Hunter, D. R., Schweinberger, M., 2013. Model-based clustering of large networks. The Annals of Applied Statistics 7, 1010.

Wang, P., Sharpe, K., Robins, G. L., Pattison, P. E., 2009. Exponential random graph (p) models for affiliation networks. Social Networks 31, 12–25.

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. pp. 1079–1087.

Xu, K. S., 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., Jin, R., 2011. Detecting communities and their evolutions in dynamic social networks - a Bayesian approach. Machine Learning 82, 157–189.

Zucchini, W., MacDonald, I. L., Langrock, R., 2016. Hidden Markov models for time series: an introduction using R. Vol. 150. CRC press.