



*J. R. Statist. Soc. B* (2017)  
**79**, Part 4, pp. 1119–1141

# Statistical clustering of temporal networks through a dynamic stochastic block model

Catherine Matias

*Centre National de la Recherche Scientifique, Université Pierre et Marie Curie  
and Laboratoire de Probabilités et Modèles Aléatoires, Paris, France*

and Vincent Miele

*Université de Lyon, Université Lyon 1, Centre National de la Recherche Scientifique  
and Laboratoire de Biométrie et Biologie Évolutive, Villeurbanne, France*

[Received June 2015. Revised June 2016]

**Summary.** Statistical node clustering in discrete time dynamic networks is an emerging field that raises many challenges. Here, we explore statistical properties and frequentist inference in a model that combines a stochastic block model for its static part with independent Markov chains for the evolution of the nodes groups through time. We model binary data as well as weighted dynamic random graphs (with discrete or continuous edges values). Our approach, motivated by the importance of controlling for label switching issues across the different time steps, focuses on detecting groups characterized by a stable within-group connectivity behaviour. We study identifiability of the model parameters and propose an inference procedure based on a variational expectation–maximization algorithm as well as a model selection criterion to select the number of groups. We carefully discuss our initialization strategy which plays an important role in the method and we compare our procedure with existing procedures on synthetic data sets. We also illustrate our approach on dynamic contact networks: one of encounters between high school students and two others on animal interactions. An implementation of the method is available as an R package called *dynsbm*.

**Keywords:** Contact network; Dynamic random graph; Graph clustering; Stochastic block model; Variational expectation–maximization

## 1. Introduction

Statistical network analysis has become a major field of research, with applications as diverse as sociology, ecology, biology and the Internet. General references on statistical modelling of random graphs include Kolaczyk (2009) and Goldenberg *et al.* (2010) and Snijders (2011). Whereas static approaches were developed as early as in the 1960s (mostly in the field of sociology), the literature concerning dynamic models is much more recent. Modelling discrete time dynamic networks is an emerging field that raises many challenges and we refer to Holme (2015) for a most recent review.

An important part of the literature on static network analysis is dedicated to clustering methods, with both aims of taking into account the intrinsic heterogeneity of the data and summarizing these data through node classification. Among clustering approaches, community detection methods form a smaller class of methods that aim at finding groups of highly connected nodes.

*Address for correspondence:* Catherine Matias, Université Pierre et Marie Curie, Case courvieux 188, 4 place Jussieu, 75252 Paris Cedex 05, France.  
E-mail: catherine.matias@math.cnrs.fr

Our focus here is not only on community detection but also more generally on node classification based on connectivity behaviour, with a particular interest on model-based approaches (see for example Matias and Robin (2014)). When considering a sequence of snapshots of a network at different time steps, these static clustering approaches will give rise to classifications that are difficult to compare through time and thus difficult to interpret. Important to note is that label switching between two successive time steps may not be solved without an extra assumption, e.g. that most of the nodes do not change group across two different time steps. However, to our knowledge, this kind of assumption has never been discussed in the literature. In this work, we are interested in statistical models for discrete time dynamic random graphs, with the aim of providing a node classification varying with time, while controlling for label switching issues across the different time steps. Our answer to this challenge will be to focus on the detection of groups that are characterized by stable within-group connectivity behaviour. We believe that this is particularly suited to dynamic contact networks.

Stochastic block models (SBMs) form a widely used class of statistical (and static) random graph models that provide a clustering of the nodes. SBMs introduce latent (i.e. unobserved) random variables on the nodes of the graph, taking values in a finite set. These latent variables represent the node groups and interaction between two nodes is governed by these corresponding groups. The model includes (but is not restricted to) the specific case of community detection, where within-group connections have higher probability than across-group connections. Combining SBMs with a Markov structure on the latent part of the process (the nodes classification) is a natural way of ensuring a smooth evolution of the groups across time and has already been considered in the literature. Yang *et al.* (2011) considered undirected, either binary or finitely valued, discrete time dynamic random graphs. The static aspect of the data is handled through SBMs, whereas its dynamic aspect is as follows. For each node, its group membership forms a Markov chain, independent of the values of the other node memberships. There, only the group membership is allowed to vary across time whereas connectivity parameters among groups stay constant through time. Yang *et al.* (2011) proposed a method to infer these parameters (either on line or off line), based on a combination of Gibbs sampling and simulated annealing. For binary random graphs, Xu and Hero (2014) proposed to introduce a state space model through time on (the logit transform of) the probability of connection between groups. In contrast with previous work, both group membership and connectivity parameters across groups may vary through time. As such, we shall see below that this model has a strong identifiability problem. Their (on-line) iterative estimation procedure is based on alternating two steps: a label switching method to explore the space of node group configurations, and the (extended) Kalman filter that optimizes the likelihood when the group memberships are known. Note that neither Yang *et al.* (2011) nor Xu and Hero (2014) proposed to infer the number of clusters. Bayesian variants of these dynamic SBMs may be found for instance in Ishiguro *et al.* (2010) and Herlau *et al.* (2013).

Surprisingly, we noted that the above-mentioned methods were evaluated on synthetic data sets in terms of the averaged value over the time steps of a clustering quality index computed at a fixed time step. Naturally, those indices do not penalize for label switching and two classifications that are identical up to a permutation have the highest quality index value. Computing an index for each time step, the label switching issue between different time steps disappears and the classification task becomes easier. Indeed, such criteria do not control for a smoothed recovery of groups along different time points. It should also be noted that the synthetic experiments from these works were performed under the dynamic version of the binary affiliation SBM, which has non-identifiable parameters. The affiliation SBM, also known as the planted partition model, corresponds to the case where the connectivity parameter matrix has only two different values: a

diagonal value that drives within-group connections and an off-diagonal value for across-group connections. In particular, the label switching issue between different time steps may not be easily removed in this particular case.

Other approaches for model-based clustering of dynamic random graphs do not rely directly on SBMs but rather on variants of the SBM. We mention the random subgraph model that combines SBMs with the *a priori* knowledge of a nodes partition (inducing subgraphs), by authorizing the group proportions to differ in the different subgraphs. A dynamic version of the random subgraph model that builds on the approach of Xu and Hero (2014) appears in Zreik *et al.* (2016). Detection of persistent communities has been proposed in Liu *et al.* (2014) for directed and dynamic graphs of call counts between individuals. Here the static underlying model is a time- and degree-corrected SBM with Poisson distribution on the call counts. Group memberships are independent through time instead of Markov, but smoothness in the classification is obtained by imposing that within-group expected call volumes are constant through time. Inference is performed through a heuristic greedy search in the space of group memberships. Only real data sets and no synthetic experiments have been explored in this latter work.

Another very popular statistical method for analysing static networks is based on latent space models. Each node is associated with a point in a latent space and probability of connection is higher for nodes whose latent points are closer (Hoff *et al.*, 2002). In Sarkar and Moore (2005) a dynamic version of the latent space model is proposed, where the latent points follow a (continuous state space) Markov chain, with transition kernel given by a Gaussian perturbation on the current position with zero mean and small variance. Latent position inference is performed in two steps: a first initial guess is obtained through multi-dimensional scaling. Then, non-linear optimization is used to maximize the model likelihood. The work by Xu and Zheng (2009) is very similar, adding a clustering step on the nodes. Finally, Heaukulani and Ghahramani (2013) relied on Markov chain Monte Carlo methods to perform a Bayesian inference in a more complicated set-up where the latent positions of the nodes are not independent.

Mixed membership models (Airoldi *et al.*, 2008) are also explored in a dynamic context. The work by Xing *et al.* (2010) relies on a state space model for the evolution of the parameters of the priors of both the mixed membership vector of a node and the connectivity behaviour. Inference is carried out through a variational Bayes expectation–maximization (EM) algorithm e.g. Jordan *et al.* (1999).

This non-exhaustive bibliography on model-based clustering methods for dynamic random graphs shows both the importance and the huge interest in the topic.

In the present work, we explore statistical properties and frequentist inference in a model that combines an SBM for its static part with independent Markov chains for the evolution of the node groups through time. Our approach aims to achieve both interpretability and statistical accuracy. Our set-up is very close to those of Yang *et al.* (2011) and Xu and Hero (2014), the first and main difference being that we allow for both group memberships and connectivity parameters to vary through time. By focusing on groups that are characterized by a stable within-group connectivity behaviour, we can ensure parameter identifiability and thus valid statistical inference. Indeed, whereas Yang *et al.* (2011) used the strong constraint of fixed connectivity parameters through time, Xu and Hero (2014) entirely relaxed this assumption at the (not acknowledged) cost of a label switching issue between time steps. Second, we model binary data as well as weighted random graphs, whether they are dense or sparse, with discrete or continuous edges. Third, we propose a model selection criterion to choose the number of clusters. To simplify the notation, we restrict our model to undirected random graphs with no self-loops but easy generalizations would handle directed data sets and/or including self-loops.

The paper is organized as follows. Section 2.1 describes the model and sets the notation. Section 2.2 gives intuition about the identifiability issues that are raised by authorizing both group memberships and connectivity parameters to vary freely with time. This was not pointed out by Xu and Hero (2014) although they worked in this context. The section motivates our focus on groups that are characterized by a stable within-group connectivity behaviour. Section 2.3 then establishes our identifiability results. To our knowledge, it is the first dynamic random graph model where parameter identifiability (up to label switching) is discussed and established. Then, Section 3 describes a variational expectation–maximization (VEM) procedure for inferring the model parameters and clustering the nodes. The VEM procedure works with a fixed number of groups and an integrated classification likelihood (Biernacki *et al.*, 2000) criterion is proposed for estimating the number of groups. We also discuss initialization of the algorithm—an important but rarely discussed step, in Section 3.2. Synthetic experiments are presented in Section 4. There, we discuss classification performances without neglecting the label switching issue that may occur between time steps. In Section 5, we illustrate our approach with the analysis of real life contact networks: a data set of encounters between high school students and two other data sets of animal interactions. We believe that our model is particularly suited to handling this type of data. We mention that the methods are implemented in an R package that is available from <http://lbbe.univ-lyon1.fr/dynsbm>. On-line supplementary materials complete the paper.

## 2. Set-up and notation

### 2.1. Model description

We consider weighted interactions between  $N$  individuals recorded through time in a set of data matrices  $\mathbf{Y} = (Y^t)_{1 \leq t \leq T}$ . Here  $T$  is the number of time points and, for each value  $t \in \{1, \dots, T\}$ , the adjacency matrix  $Y^t = (Y_{ij}^t)_{1 \leq i \neq j \leq N}$  contains real values measuring interactions between individuals  $i, j \in \{1, \dots, N\}^2$ . Without loss of generality, we consider undirected random graphs without self-loops, so that  $Y^t$  is a symmetric matrix with no diagonal elements.

We assume that the  $N$  individuals are split into  $Q$  latent (unobserved) groups that may vary through time, as encoded by the random variables  $\mathbf{Z} = (Z_i^t)_{1 \leq t \leq T, 1 \leq i \leq N}$  with values in  $\mathcal{Q}^{NT} := \{1, \dots, Q\}^{NT}$ . This process is modelled as follows. Across individuals, random variables  $(Z_i)_{1 \leq i \leq N}$  are independent and identically distributed. Now, for each individual  $i \in \{1, \dots, N\}$ , the process  $Z_i = (Z_i^t)_{1 \leq t \leq T}$  is an irreducible, aperiodic stationary Markov chain with transition matrix  $\pi = (\pi_{qq'})_{1 \leq q, q' \leq Q}$  and initial stationary distribution  $\alpha = (\alpha_1, \dots, \alpha_Q)$ . When no confusion occurs, we may alternatively consider  $Z_i^t$  as a value in  $\mathcal{Q}$  or as a random vector  $Z_i^t = (Z_{i1}^t, \dots, Z_{iQ}^t) \in \{0, 1\}^Q$  constrained to  $\sum_q Z_{iq}^t = 1$ .

Given latent groups  $\mathbf{Z}$ , the time varying random graphs  $\mathbf{Y} = (Y^t)_{1 \leq t \leq T}$  are independent, the conditional distribution of each  $Y^t$  depending only on  $Z^t$ . Then, for fixed  $1 \leq t \leq T$ , random graph  $Y^t$  follows an SBM. In other words, for each time  $t$ , conditionally on  $Z^t$ , random variables  $(Y_{ij}^t)_{1 \leq i < j \leq N}$  are independent and the distribution of each  $Y_{ij}^t$  depends on only  $Z_i^t$  and  $Z_j^t$ . For now, we assume a very general parametric form for this distribution on  $\mathbb{R}$ . Following Ambroise and Matias (2012), to take into account possible sparse weighted graphs, we explicitly introduce a Dirac mass at 0, denoted by  $\delta_0$ , as a component of this distribution. More precisely, we assume that

$$Y_{ij}^t | \{Z_{iq}^t Z_{jl}^t = 1\} \sim (1 - \beta_{ql}^t) \delta_0(\cdot) + \beta_{ql}^t F(\cdot, \gamma_{ql}^t), \quad (1)$$

where  $\{F(\cdot, \gamma), \gamma \in \Gamma\}$  is a parametric family of distributions with no point mass at 0 and densities (with respect to Lebesgue or counting measure) denoted by  $f(\cdot, \gamma)$ . This could be the Gaussian

family with unknown mean and variance, the truncated Poisson family on  $\mathbb{N} \setminus \{0\}$  (leading to a 0-inflated or 0-deflated distribution on the edges of the graph), a finite space distribution on  $M$ -values (a case which comprises non-parametric approximations of continuous distributions through discretization into a finite number of  $M$  bins), etc. The binary case is encompassed in this set-up with  $F(\cdot, \gamma) = \delta_1(\cdot)$ , namely the parametric family of laws is reduced to a single point, the Dirac mass at 1 and conditional distribution of  $Y_{ij}^t$  is simply a Bernoulli  $\mathcal{B}(\beta_{ij}^t)$  distribution. In what follows and in contrast with the ‘binary case’, we shall call the ‘weighted case’ any set-up where the set of distributions  $F$  is parameterized and not reduced to a single point. Here, the sparsity parameters  $\beta^t = (\beta_{ql}^t)_{1 \leq q, l \leq Q}$  satisfy  $\beta_{ql}^t \in [0, 1]$ , with  $\beta^t \equiv 1$  corresponding to the particular case of a complete weighted graph. As a result of considering undirected graphs, the parameters  $\beta_{ql}^t$  and  $\gamma_{ql}^t$  moreover satisfy  $\beta_{ql}^t = \beta_{lq}^t$  and  $\gamma_{ql}^t = \gamma_{lq}^t$  for all  $1 \leq q, l \leq Q$ . For the moment, SBM parameters may be different across time points. We shall return to this point in the next sections. The model is thus parameterized by

$$\theta = (\pi, \beta, \gamma) = (\pi, \{\beta^t, \gamma^t\}_{1 \leq t \leq T}) = (\{\pi_{qq'}\}_{1 \leq q, q' \leq Q}, \{\beta_{ql}^t, \gamma_{ql}^t\}_{1 \leq t \leq T, 1 \leq q \leq l \leq Q}) \in \Theta,$$

and we let  $\mathbb{P}_\theta$  denote the probability distribution on the whole space  $\mathcal{Q}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}$ . We also let  $\phi(\cdot; \beta, \gamma)$  denote the density of the distribution given by expression (1), namely

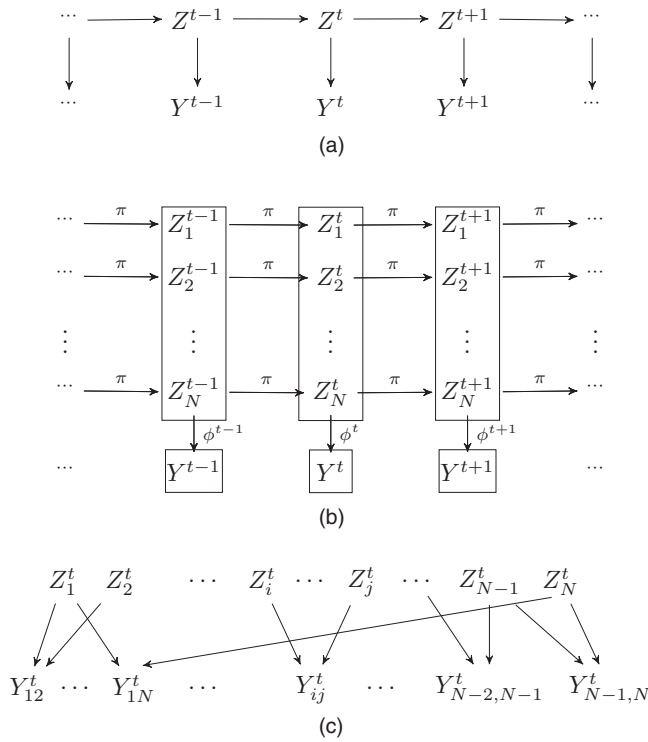
$$\forall y \in \mathbb{R}, \quad \phi(y; \beta, \gamma) = (1 - \beta) \mathbf{1}\{y = 0\} + \beta f(y, \gamma) \mathbf{1}\{y \neq 0\},$$

where  $\mathbf{1}\{A\}$  is the indicator function of set  $A$ . With some abuse of notation and when no confusion occurs, we shorten  $\phi(\cdot; \beta_{ql}^t, \gamma_{ql}^t)$  to  $\phi_{ql}^t(\cdot)$  or  $\phi_{ql}^t(\cdot; \theta)$ . Directed acyclic graphs describing the dependence structure of the variables in the model with different levels of detail are given in Fig. 1. Note that the model assumes that the individuals are present at any time in the data set. An extension that covers the case where some nodes are not present at every time point is given in section S.5 in the on-line supplementary materials and used in analysing the animal data sets from Section 5.2.

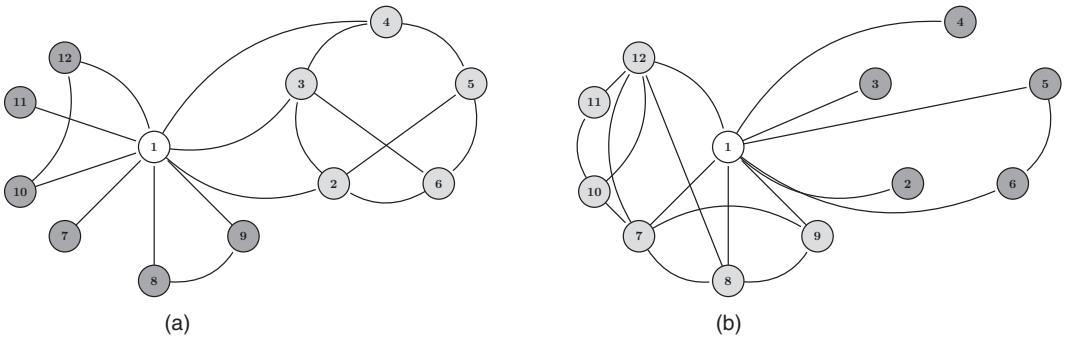
## 2.2. Varying connectivity parameters versus varying group membership

In this section, we give some intuition into why it is not possible to let both connectivity parameters and group membership vary through time without entering label switching issues between time steps. For this, we consider the toy example in Fig. 2.

Fig. 2 shows a graph between  $N = 12$  nodes at two different time points  $t_1$  and  $t_2$ . Node 1 is a hub (namely a highly connected node), nodes 2–6 form a community at time  $t_1$  (they tend to form a clique) and are peripheral nodes at time  $t_2$  and finally nodes 7–12 are peripheral at time  $t_1$  and become a community at time  $t_2$ . In observing those two graphs (without the clusters indicated by the shading of the nodes), there are at least two possible statistical interpretations relying on a clustering with  $Q = 3$  groups. The first (which is illustrated in Fig. 2) is to consider that the three different groups at stake are hubs (in white), a community (light grey) and peripheral nodes (dark grey) and that the nodes 2–6 change group from a community to peripheral group between time  $t_1$  and  $t_2$  whereas nodes 7–12 change from peripheral group to a community between those same time points (node 1 stays a hub, in white, for both time points). Another point of view would rather be to consider that nodes 2–6 stayed in the same group that was organized as a community at time  $t_1$  and is now characterized by peripheral behaviour at time  $t_2$ , whereas nodes 7–12 also stayed in the same group, behaving peripherally at time  $t_1$  and now as a community at time  $t_2$ . Obviously neither of these two interpretations is better than the other. Without adding constraints on the model, the label switching phenomenon will randomly output one of these two interpretations (clustering at time  $t_2$  is the same when permuting light grey and dark grey



**Fig. 1.** Dependence structures of the model: (a) general view corresponding to a hidden Markov model structure; (b) details on latent structure organization corresponding to  $N$  different independent and identically distributed Markov chains  $Z_i = (Z_i^t)_{1 \leq t \leq T}$  across individuals; (c) details for fixed time point  $t$  corresponding to an SBM structure



**Fig. 2.** Connectivity parameters or group membership variation—a toy example: (a)  $t = t_1$ ; (b)  $t = t_2$

colours). In this context, it is thus impossible to recover group membership trajectories. We formalize these ideas in the next section through the concept of parameter identifiability.

The main problem with the previous example comes from the possibility of arbitrarily relabelling the groups between two time steps. We mentioned in Section 1 that a natural idea would be that most of the individuals should not change groups between successive time steps. However, imposing constraints on the transition matrix  $\pi$  (for example it has large diagonal elements) is useless because estimation would then be unfeasible. Indeed, without imposing 0-

values on the off-diagonal elements of  $\pi$  (i.e.  $Z_i^t$  does not depend on  $t$ ), it can happen that there is no labelling of the groups that ensures that most individuals stay in the same group. Thus it is not always possible to label the groups so that, between two successive time steps, estimation of the transition parameters would be constrained to have large diagonal elements. Thus we choose to focus our attention on groups that are characterized through their stable within-group connectivity parameter. This choice is reminiscent of work on detection of persistent communities (Liu *et al.*, 2014), except that we do not restrict our attention to communities (i.e. groups of highly connected individuals). In this toy example, this corresponds to the first interpretation rather than the second. Other choices could be made and we believe that this one is particularly suited to model social networks or contact data where the groups are defined as structures exhibiting stable within-group connectivity behaviour and individuals may change groups through time (see Section 5 for applications on real data sets).

### 2.3. Parameter identifiability

Recall that, with discrete latent random variables, identifiability can only be obtained up to label switching on the node groups  $\mathcal{Q}$ . For any permutation  $\sigma$  in  $\mathfrak{S}_{\mathcal{Q}}$  (the set of permutations on  $\mathcal{Q}$ ) and any  $\theta \in \Theta$ , we define

$$\sigma(\theta) := (\{\pi_{\sigma(q)\sigma(q')}\}_{1 \leq q, q' \leq Q}, \{\beta_{\sigma(q)\sigma(l)}^t, \gamma_{\sigma(q)\sigma(l)}^t\}_{1 \leq t \leq T, 1 \leq q \leq l \leq Q}).$$

It should be noted that, here, the permutation  $\sigma$  acts *globally*, meaning that it is the same at each time point  $t$ . Now, if we let  $\mathbb{P}_{\theta}^Y$  denote the marginal of  $\mathbb{P}_{\theta}$  on the set of observations  $\mathbf{Y}$ , identifiability of the parameterization, up to label switching, means that

$$\forall \theta, \tilde{\theta} \in \Theta, \quad \mathbb{P}_{\theta}^Y = \mathbb{P}_{\tilde{\theta}}^Y \Rightarrow \exists \sigma \in \mathfrak{S}_{\mathcal{Q}}, \theta = \sigma(\tilde{\theta}).$$

Without additional constraints on the transition matrix  $\pi$  or on the parameters  $(\beta, \gamma)$ , the parameters may not be recovered up to label switching. However, it could be that the static SBM part of the parameter is recovered up to *local* label switching. Local label switching on the SBM part of the parameter is the weaker property

$$\forall \theta, \tilde{\theta} \in \Theta, \quad \mathbb{P}_{\theta}^Y = \mathbb{P}_{\tilde{\theta}}^Y \Rightarrow \exists \sigma_1, \dots, \sigma_T \in \mathfrak{S}_{\mathcal{Q}}^T, \forall t, (\beta^t, \gamma^t) = \sigma_t(\tilde{\beta}^t, \tilde{\gamma}^t).$$

This property is not satisfactory since clustering in models that satisfy only local identifiability of the SBM part of the parameter prevents us from obtaining a picture of the evolution of the groups across time.

A formal example of the fact that, if both  $Z^t$  and  $(\beta^t, \gamma^t)$  may vary through time, then the parameter cannot be identified up to label switching without additional constraints is given in section S.1 in the on-line supplementary materials. We stress that this example implies that a dynamic affiliation SBM (or planted partition model) does not have identifiable parameters and groups may not be recovered consistently across time. This is an important point as previous researchers have tried to recover groups from this type of synthetic data sets and evaluated their estimated classification in a non-natural way.

As a consequence and following the ideas that were developed in Section 2.2, we choose to impose the following constraints on the parameter  $\theta$ :

$$\forall q \in \mathcal{Q}, \forall t, t' \in \{1, \dots, T\}, \quad \begin{cases} \text{binary case, } \beta_{qq}^t = \beta_{qq}^{t'} := \beta_{qq}, \\ \text{weighted case, } \gamma_{qq}^t = \gamma_{qq}^{t'} := \gamma_{qq}. \end{cases} \quad (2)$$

Under this condition, we focus on groups that are characterized by stable within-group

connectivity behaviour ( $\beta_{qq}$  or  $\gamma_{qq}$  is constant with time). This constraint could in fact be weakened as follows:

$$\forall q \in Q, \forall t, t' \in \{1, \dots, T\}, \exists l \in \{1, \dots, T\}, \begin{cases} \text{binary case, } \beta_{ql}^t = \beta_{ql}^{t'}, \\ \text{weighted case, } \gamma_{ql}^t = \gamma_{ql}^{t'}. \end{cases} \quad (3)$$

In this condition, the group  $l$  that helps to characterize the group  $q$  between the two different time points  $t$  and  $t'$  may depend on  $q$ ,  $t$  and  $t'$ . Such a constraint may be useful if groups are not characterized by stable within-group connectivity but rather by their connectivity to at least one specific other group. For estimation, this group  $l$  needs to be known in advance (for each  $q$ ,  $t$  and  $t'$ ) which requires more complex *a priori* modelling of the data. In what follows, we choose to restrict our attention to constraint (2) only but our theoretical results remain valid under constraint (3). We prove below that these constraints, combined with the same conditions that are used for identifiability in the static case, are sufficient to ensure identifiability of the parameterization in our dynamic set-up.

*Assumption 1* (weighted case). We assume that

- (a) for any  $t \geq 1$ , the  $Q(Q+1)/2$  values  $\{\gamma_{ql}^t, 1 \leq q \leq l \leq Q\}$  are distinct and
- (b) the family of distributions  $\mathcal{F} = \{f(\cdot, \gamma), \gamma \in \Gamma\}$  is such that all elements  $f(\cdot, \gamma)$  have no point mass at 0 and the parameters of finite mixtures of distributions in  $\mathcal{F}$  are identifiable, up to label switching.

Assumption 1 is the condition that ensures identifiability of static weighted SBMs (see theorem 12 in Allman *et al.* (2011)). Note that it does not impose any constraint on the sparsity parameters  $\beta_{ql}^t$  in the weighted case. In particular and for parsimony, these may be chosen identical (to some  $\beta^t$  or some constant  $\beta$ ) or set to two different values, e.g.  $\beta_{qq}^t = \beta_{\text{in}}^t$  and  $\beta_{ql}^t = \beta_{\text{out}}^t$  whenever  $q \neq l$  at each time point (or even constant with time).

*Proposition 1.* Considering the distribution  $\mathbb{P}_\theta^Y$  on the set of observations and assuming constraint (2), the parameter  $\theta = (\pi, \beta, \gamma)$  satisfies the following conditions.

- (a) *Binary case:*  $\theta$  is generically identified from  $\mathbb{P}_\theta^Y$ , up to label switching, as soon as  $N$  is not too small with respect to  $Q$ .
- (b) *Weighted case:* under additional assumption 1, the parameter  $\theta$  is identified from  $\mathbb{P}_\theta^Y$ , up to label switching, as soon as  $N \geq 3$ .

Generic identifiability means ‘up to excluding a subset of zero Lebesgue measure of the parameter set’. We refer to Allman *et al.* (2009, 2011) for more details. In particular for the binary case, assuming that the matrix of Bernoulli parameters  $\beta$  has distinct rows is a generic constraint (meaning that it removes a subset of zero Lebesgue measure of the parameter set). As we do not specify the whole generic constraint that is needed here, we do not stress that one either. But the reader should have it in mind in the binary set-up. Finally, note that the condition on the number of nodes  $N$  being not too small in the binary case is given precisely in theorem 2 from Allman *et al.* (2011). The particular affiliation case (planted partition) is not covered by these results and is further discussed in section S.2 in the on-line supplementary materials.

*Proof.* The proof combines the approaches of Leroux (1992) for proving identifiability of hidden Markov model parameters and Allman *et al.* (2011) that studies identifiability for (static) SBMs.

First, we fix a time point  $t \geq 1$  and consider the marginal distribution  $\mathbb{P}_\theta(Y^t)$ . According to



theorems 1 and 2 (binary case with  $Q = 2$  and  $Q \geq 3$  respectively) and theorem 12 (weighted case) in Allman *et al.* (2011) on parameter identifiability in static SBMs, there is a permutation  $\sigma_t$  on the group labels  $\mathcal{Q}$  such that we can identify  $(\beta^t, \gamma^t)$  as well as the marginal distribution  $\alpha$ , up to this permutation. This result stands generically in the binary case only.

Now, for two different time points  $t$  and  $t'$ , we use constraint (2) and the assumption of distinct parameter values to identify the parameters  $\{(\beta^t, \gamma^t), t \geq 1\}$  up to a (common) permutation  $\sigma$  on  $\mathcal{Q}$ . Indeed, in the binary case, assuming that the within-groups Bernoulli parameters satisfy  $\beta_{qq}^t = \beta_{qq}^{t'}$  and that the set  $\{\beta_{qq}^t; 1 \leq q \leq Q\}$  contains  $Q$  distinct values (a generic constraint) suffices to obtain a global permutation  $\sigma$ , not depending on time  $t$ , up to which  $\{(\beta^t, \gamma^t), t \geq 1\}$  are identified. The same applies in the weighted case, by assuming equality between the parameter  $\gamma_{qq}^t = \gamma_{qq}^{t'}$  for any  $t$  and  $t'$ .

It remains to identify the transition matrix  $\pi$  (up to the same permutation  $\sigma$ ). We fix an edge  $(i, j)$  and, following Leroux (1992), consider the bivariate distribution  $\mathbb{P}_\theta(Y_{ij}^t, Y_{ij}^{t+1})$ . This is given by

$$\mathbb{P}_\theta(Y_{ij}^t, Y_{ij}^{t+1}) = \sum_{q_1, q_2, l_1, l_2 \in \mathcal{Q}} \alpha_{q_1} \alpha_{l_1} \pi_{q_1 q_2} \pi_{l_1 l_2} \phi_{q_1 l_1}^t(Y_{ij}^t) \phi_{q_2 l_2}^{t+1}(Y_{ij}^{t+1}). \quad (4)$$

Note that Teicher (1967) has proved the equivalence between parameter identifiability for the mixtures of a family of distributions and parameter identifiability for the mixtures of finite products from this same family. For clarity, we develop his proof adapted to our context. We thus write

$$\mathbb{P}_\theta(Y_{ij}^t, Y_{ij}^{t+1}) = \sum_{q_2, l_2 \in \mathcal{Q}} \left\{ \sum_{q_1, l_1 \in \mathcal{Q}} \alpha_{q_1} \alpha_{l_1} \pi_{q_1 q_2} \pi_{l_1 l_2} \phi_{q_1 l_1}^t(Y_{ij}^t) \right\} \phi_{q_2 l_2}^{t+1}(Y_{ij}^{t+1}).$$

As the mixtures from the family  $\{\phi_{q_l}^{t+1}, 1 \leq q \leq l \leq Q\}$  have identifiable parameters (assumption 1, part (b)), we can identify the mixing distribution

$$\sum_{q_2, l_2 \in \mathcal{Q}} \left\{ \sum_{q_1, l_1 \in \mathcal{Q}} \alpha_{q_1} \alpha_{l_1} \pi_{q_1 q_2} \pi_{l_1 l_2} \phi_{q_1 l_1}^t(Y_{ij}^t) \right\} \delta_{(\beta_{q_2 l_2}^{t+1}, \gamma_{q_2 l_2}^{t+1})}(\cdot).$$

Now, applying again this identifiability at time  $t$  and constraint (1), we can identify the whole mixing distribution

$$\sum_{q_2, l_2 \in \mathcal{Q}} \sum_{q_1, l_1 \in \mathcal{Q}} \alpha_{q_1} \alpha_{l_1} \pi_{q_1 q_2} \pi_{l_1 l_2} \delta_{(\beta_{q_1 l_1}^t, \gamma_{q_1 l_1}^t)}(\cdot) \otimes \delta_{(\beta_{q_2 l_2}^{t+1}, \gamma_{q_2 l_2}^{t+1})}(\cdot).$$

This proves that the mixture given by equation (4) has identifiable components. From this mixture and the fact that we have already identified the parameters  $(\beta, \gamma)$  up to a global permutation, we can extract the set of coefficients  $\{\alpha_q^2 \pi_{qq}^2, 1 \leq q, q' \leq Q\}$  that corresponds to the components  $\phi_{qq}^t \phi_{q'q'}^{t+1}$  in equation (4). As we have also already obtained the values  $\{\alpha_q, 1 \leq q \leq Q\}$ , this now identifies the parameters  $\{\pi_{qq'}, 1 \leq q, q' \leq Q\}$ . This concludes the proof.

### 3. Inference algorithm

#### 3.1. General description

As usual with latent variables, the log-likelihood  $\log\{\mathbb{P}_\theta(\mathbf{Y})\}$  contains a sum over all possible latent configurations  $\mathbf{Z}$  and thus may not be computed except for small values of  $N$  and  $T$ . A classical solution is to rely on an EM algorithm (Dempster *et al.*, 1977), which is an iterative procedure that finds local maxima of the log-likelihood. The use of the EM algorithm relies on the computation of the conditional distribution of the latent variables  $\mathbf{Z}$  given the observed

variables  $\mathbf{Y}$ . However, in the context of SBMs, this distribution does not have a factored form and thus may not be computed efficiently. A classical solution is to rely on variational approximations of the EM algorithm: the VEM algorithm (see for instance Jordan *et al.* (1999)). These approximations were first proposed in the context of SBMs in Daudin *et al.* (2008) and later developed in many directions, such as on-line procedures (Zanghi *et al.*, 2008, 2010) or Bayesian VEM algorithms (Latouche *et al.*, 2012). We refer to Matias and Robin (2014) for more details about the VEM algorithm (in particular a presentation of the EM algorithm viewed as a special instance of the VEM algorithm) and its comparison with other estimation procedures in SBMs. Note that convergence properties of VEM algorithms are discussed in full generality in Gunawardana and Byrne (2005) and in the special case of SBMs in Celisse *et al.* (2012) and Bickel *et al.* (2013).

### 3.1.1. Variational expectation–maximization for dynamic stochastic block models

In our context of dynamic random graphs, we start by writing the complete-data log-likelihood of the model

$$\begin{aligned} \log\{\mathbb{P}_\theta(\mathbf{Y}, \mathbf{Z})\} &= \sum_{i=1}^N \sum_{q=1}^Q Z_{iq}^1 \log(\alpha_q) + \sum_{t=2}^T \sum_{i=1}^N \sum_{1 \leq q, q' \leq Q} Z_{iq}^{t-1} Z_{iq'}^t \log(\pi_{qq'}) \\ &\quad + \sum_{t=1}^T \sum_{1 \leq i < j \leq N} \sum_{1 \leq q, l \leq Q} Z_{iq}^t Z_{jl}^t \log\{\phi(Y_{ij}^t; \beta_{ql}^t, \gamma_{ql}^t)\}. \end{aligned} \quad (5)$$

We now explore the dependence structure of the conditional distribution  $\mathbb{P}_\theta(\mathbf{Z}|\mathbf{Y})$ . First, note that it can be easily deduced from the directed acyclic graph of the model (Fig. (1a)) that

$$\mathbb{P}_\theta(\mathbf{Z}|\mathbf{Y}) = \mathbb{P}_\theta(Z^1|Y^1) \prod_{t=2}^T \mathbb{P}_\theta(Z^t|Z^{t-1}, Y^t).$$

However, the distribution  $\mathbb{P}_\theta(Z^t|Z^{t-1}, Y^t) = \mathbb{P}_\theta\{(Z_i^t)_{1 \leq i \leq N} | Z^{t-1}, Y^t\}$  cannot be further factored. Indeed, for any  $i \neq j$ , the variables  $Z_i^t$  and  $Z_j^t$  are not independent when conditioned on  $Y^t$ . Our variational approximation naturally considers the following class of probability distributions  $\mathbb{Q} := \mathbb{Q}_\tau$  parameterized by  $\tau$ :

$$\begin{aligned} \mathbb{Q}_\tau(\mathbf{Z}) &= \prod_{i=1}^N \mathbb{Q}_\tau(Z_i) = \prod_{i=1}^N \mathbb{Q}_\tau(Z_i^1) \prod_{t=2}^T \mathbb{Q}_\tau(Z_i^t | Z_i^{t-1}) \\ &= \prod_{i=1}^N \left\{ \prod_{q=1}^Q \tau(i, q)^{Z_{iq}^1} \right\} \times \prod_{t=2}^T \prod_{1 \leq q, q' \leq Q} \tau(t, i, q, q')^{Z_{iq}^{t-1} Z_{iq'}^t}, \end{aligned}$$

where, for any values  $(t, i, q, q')$ , we have  $\tau(i, q)$  and  $\tau(t, i, q, q')$  both belong to the set  $[0, 1]$  and are constrained by  $\sum_q \tau(i, q) = 1$  and  $\sum_{q'} \tau(t, i, q, q') = 1$ . This class of probability distributions  $\mathbb{Q}_\tau$  corresponds to considering independent laws through individuals, whereas, for each  $i \in \{1, \dots, N\}$ , the distribution of  $Z_i$  under  $\mathbb{Q}_\tau$  is the distribution of a Markov chain (through time  $t$ ), with inhomogeneous transition  $\tau(t, i, q, q') = \mathbb{Q}_\tau(Z_i^t = q' | Z_i^{t-1} = q)$  and initial distribution  $\tau(i, q) = \mathbb{Q}_\tau(Z_i^1 = q)$ .

We shall need the marginal components of  $\mathbb{Q}_\tau$ , namely  $\tau_{\text{marg}}(t, i, q) := \mathbb{Q}_\tau(Z_i^t = q)$ . These quantities are computed recursively by

$$\begin{aligned} \tau_{\text{marg}}(1, i, q) &= \tau(i, q) \text{ and } \forall t \geq 2, \\ \tau_{\text{marg}}(t, i, q) &= \sum_{q'=1}^Q \tau_{\text{marg}}(t-1, i, q') \tau(t, i, q', q). \end{aligned}$$

Note also that all the values  $\tau_{\text{marg}}(t, i, q)$  depend on the initial distribution  $\tau(i, q)$ . The entropy of  $\mathbb{Q}_\tau$  is denoted by  $\mathcal{H}(\mathbb{Q}_\tau)$ . Using this class of probability distributions on  $\mathcal{Q}^\mathbb{N}$ , the VEM algorithm is an iterative procedure to optimize the criterion

$$\begin{aligned}
 J(\theta, \tau) &:= \mathbb{E}_{\mathbb{Q}_\tau}[\log\{\mathbb{P}_\theta(\mathbf{Y}, \mathbf{Z})\}] + \mathcal{H}(\mathbb{Q}_\tau) \\
 &= \sum_{i=1}^N \sum_{q=1}^Q \tau(i, q) [\log(\alpha_q) - \log\{\tau(i, q)\}] \\
 &\quad + \sum_{t=2}^T \sum_{i=1}^N \sum_{1 \leq q, q' \leq Q} \tau_{\text{marg}}(t-1, i, q) \tau(t, i, q, q') [\log(\pi_{qq'}) - \log\{\tau(t, i, q, q')\}] \\
 &\quad + \sum_{t=1}^T \sum_{1 \leq i < j \leq N} \sum_{1 \leq q, l \leq Q} \tau_{\text{marg}}(t, i, q) \tau_{\text{marg}}(t, j, l) \log\{\phi_{ql}^t(Y_{ij}^t)\}. \tag{6}
 \end{aligned}$$

It consists of iterating the following two steps. At the  $k$ th iteration, with current parameter value  $(\tau^{(k)}, \theta^{(k)})$ , we do the following steps.

- (a) VE step: compute  $\tau^{(k+1)} = \arg \max_\tau J(\theta^{(k)}, \tau)$ .
- (b) M-step: compute  $\theta^{(k+1)} = \arg \max_\theta J(\theta, \tau^{(k+1)})$ .

*Proposition 2.* The value  $\hat{\tau}$  that maximizes in  $\tau$  the function  $J(\theta, \tau)$  satisfies the fixed point equation

$$\forall t \geq 2, \forall i \geq 1, \forall q, q' \in \mathcal{Q}, \quad \hat{\tau}(t, i, q, q') \propto \pi_{qq'} \prod_{j=1}^N \prod_{l=1}^Q \phi_{q'l}^t(Y_{ij}^t) \hat{\tau}_{\text{marg}}(t, j, l')$$

where ‘ $\propto$ ’ means ‘proportional to’ (the constants are obtained by the constraints on  $\tau$ ). Moreover, the value  $(\hat{\pi}, \hat{\beta})$  that maximizes in  $(\pi, \beta)$  the function  $J(\theta, \tau)$  satisfies

$$\begin{aligned}
 \forall (q, q') \in \mathcal{Q}^2, \quad \hat{\pi}_{qq'} &\propto \sum_{t=2}^T \sum_{i=1}^N \tau_{\text{marg}}(t-1, i, q) \tau(t, i, q, q'), \\
 \forall t, \forall q \neq l \in \mathcal{Q}, \quad \hat{\beta}_{ql}^t &= \frac{\sum_{i,j} \tau_{\text{marg}}(t, i, q) \tau_{\text{marg}}(t, j, l) \mathbf{1}_{Y_{ij}^t \neq 0}}{\sum_{i,j} \tau_{\text{marg}}(t, i, q) \tau_{\text{marg}}(t, j, l)}, \\
 \forall q \in \mathcal{Q}, \quad \hat{\beta}_{qq} &= \frac{\sum_t \sum_{i,j} \tau_{\text{marg}}(t, i, q) \tau_{\text{marg}}(t, j, q) \mathbf{1}_{Y_{ij}^t \neq 0}}{\sum_{t,i,j} \tau_{\text{marg}}(t, i, q) \tau_{\text{marg}}(t, j, q)}.
 \end{aligned}$$

The proof of this result is immediate and has been omitted. Note that we have given a formula with constant (through time) values  $\beta_{qq}$  for any group  $q \in \mathcal{Q}$ . Whereas this assumption is an identifiability requirement in the binary set-up, it is not necessary in the weighted case. In the weighted case, we use it only for parsimony. The corresponding formula when this parameter is not assumed to be constant may be easily obtained.

To complete the algorithm’s description, we provide equations to update the parameters  $\tau(i, q)$  and  $\alpha_q$  of initial distributions as well as the connectivity parameter  $\gamma$ . First, optimization of  $J(\theta, \tau)$  with respect to the initialization parameters  $\tau(i, q)$  is a little more involved. By neglecting the dependence on  $\tau(i, q)$  of some terms appearing in criterion  $J$ , we choose to update this value by solving the fixed point equation

$$\forall i \geq 1, \forall q \in \mathcal{Q}, \quad \hat{\tau}(i, q) \propto \alpha_q \prod_{j=1}^N \prod_{l=1}^Q \phi_{ql}^1(Y_{ij}^1)^{\hat{\tau}(j,l)}. \quad (7)$$

Our experiments show that this is a reasonable approximation (Section 4). For completeness, we provide in section S.3 in the on-line supplementary materials the exact equation satisfied by the solution.

Now parameter  $\alpha$  is not obtained from maximizing  $J$  as it is not a free parameter but rather the stationary distribution associated with transition  $\pi$ . Thus,  $\alpha$  is obtained from the empirical mean of the marginal distribution  $\hat{\tau}_{\text{marg}}$  over all data points

$$\forall q \in \mathcal{Q}, \quad \hat{\alpha}_q = \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N \hat{\tau}_{\text{marg}}(t, i, q).$$

Finally, optimization with respect to  $\gamma$  depends on the choice of the parametric family  $\{f(\cdot, \gamma), \gamma \in \Gamma\}$ . We provide explicit formulae for the most widely used families of conditional distributions on the edges (binary or weighted case) in section S.4 in the on-line supplementary materials. More precisely, we give these formulae for Bernoulli, finite space, (zero-inflated or deflated) Poisson and Gaussian homoscedastic distributions.

*Remark 1.* Performing the EM algorithm in a hidden Markov model (Fig. 1(a)) requires the use of forward–backward equations to deal with transition terms  $Z_{iq}^{t-1} Z_{iq}^t$  appearing in the complete-data log-likelihood (5). In our set-up, forward–backward equations are useless and replaced by a variational approximation. Indeed, it can be seen from Fig. 1(b) that the conditional distribution of  $Z_{iq}^{t-1} Z_{iq}^t$  given the data cannot be computed exactly through such forward–backward equations. This is because the variables  $Y^t = \{Y_{ij}^t\}_{i,j}$  depend on all hidden variables  $Z_1^t, \dots, Z_N^t$  and focusing only on  $Z_i^t$  is not sufficient to determine their distribution.

*Remark 2.* Yang *et al.* (2011) derived a VEM procedure in a similar (slightly less general) set-up, but their variational approximation uses independent marginals (through individuals and also time points). As a consequence, the VE step that they derive is more involved than ours (see section 4 in Yang *et al.* (2011)).

### 3.1.2. Model selection

Model selection on the number of groups  $Q$  is an important step. In the case of latent variables, when the true data likelihood may not be easily computed, model selection may be done by maximizing an integrated classification likelihood (ICL) criterion (Biernacki *et al.*, 2000). For any number of groups  $Q \geq 1$ , let  $\hat{\theta}_Q$  be the estimated parameter value with  $Q$  groups and  $\hat{\mathbf{Z}}$  the corresponding maximum *a posteriori* classification at  $\hat{\theta}_Q$ . In our case, the general form of the ICL is given by

$$\text{ICL}(Q) = \log\{\mathbb{P}_{\hat{\theta}_Q}(\mathbf{Y}, \hat{\mathbf{Z}})\} - \frac{1}{2}Q(Q-1) \log\{N(T-1)\} - \text{pen}(N, T, \beta, \gamma), \quad (8)$$

where the first penalization term accounts for transition matrix  $\pi$  and  $\text{pen}(N, T, \beta, \gamma)$  is a penalizing term for the connectivity parameters  $(\beta, \gamma)$ . As the number of parameters  $(\beta, \gamma)$  depends on the specific form of the family  $\{f(\cdot; \gamma), \gamma \in \Gamma\}$ , we provide context-dependent expressions for the ICL in section S.4 in the on-line supplementary materials (along with the expressions of parameter estimates from the M-step for each case considered). Note that the first penalization term accounts for  $N(T-1)$  latent transitions whereas the number of observations corresponding to the SBM part of the parameter in  $\text{pen}(N, T, \beta, \gamma)$  will be different. We refer to Daudin *et al.* (2008) for an expression of the ICL in the static SBM that shows an analogous difference

in penalizing group proportions or connectivity parameters. Note that there are no theoretical results on the convergence of the ICL procedure (nor in simple mixture models nor in the SBM case). However, the criterion shows very good performances on synthetic experiments and is widely used (see Section 4 for experiments in our set-up). Nonetheless we mention that the criterion is not suited to the case of a finite space conditional distribution (see example 2 in section S.4 in the supplementary materials for more details).

### 3.2. Algorithm initialization

All EM-based procedures look for local maxima of their objective function and careful initialization is a key in their success. For the static SBM, VEM procedures often rely on a  $k$ -means algorithm on the adjacency matrix to obtain an initial clustering of the individuals. In our context, the dynamic aspect of the data needs to be properly handled. We choose to initialize our VEM procedure by running  $k$ -means on the rows of a concatenated data matrix containing all the adjacency time step matrices  $Y^t$  stacked in consecutive column blocks. As a result, our initial clustering of the individuals is constant across time (namely  $Z_i^t$  does not depend on  $t$ ). A consequence of this choice is that this initialization works well when the group memberships do not vary too much across time (see Section 4 where we explore different values of transition matrix  $\pi$ ). In practice, real life contact networks will either exhibit nodes that do not change group at all (see Section 5) or nodes that leave a group and then come back to this group. Our initialization is efficient in these cases. Another consequence is that, whereas we would expect the performances of the procedure to increase with the number  $T$  of time steps, we sometimes observe on the contrary a decrease in these performances. This is because increasing  $T$  also increases the probability for an individual to change group at some point in time and thus, starting with a constant-in-time clustering of the individuals, it becomes more difficult to infer the group memberships at each time point correctly (see in Section 4 the difference between results for  $T = 5$  and  $T = 10$ ).

To conclude this section, we mention that initialization is also a crucial point for other methods and we discuss in the next section its effect on the algorithm that was proposed in Yang *et al.* (2011).

## 4. Synthetic experiments

The methods that are presented in this paper are implemented in an R package which is available from <http://lbbe.univ-lyon1.fr/dynsbm>. Although the complexity of the estimation algorithm is  $O(TQ^2N^2)$ , the computation time remains acceptable for networks with a few thousand nodes (see the on-line supplementary Fig. 1).

### 4.1. Clustering performances

In this section, we explore the performances of our method for clustering the nodes across the different time steps. For this, we shall consider two different criteria. We rely on the adjusted Rand index (ARI) (Hubert and Arabie, 1985) to evaluate the agreement between the estimated and the true latent structure. This index is smaller than 1, two identical latent structures (up to label switching) having an ARI equal to 1. Note that it can take negative values and is built on the Rand index with a correction for chance. Now there are two different ways of using the ARI in a dynamic set-up. Following Yang *et al.* (2011) and Xu and Hero (2014) we first consider an averaged value over the different time steps  $1 \leq t \leq T$  of  $\text{ARI}_t$  computed at time  $t$ . In this approach the dynamic set-up may be viewed as a way of improving the node clustering at each

**Table 1.** Bernoulli parameter values in four cases, plus an affiliation example

| <i>Easiness</i>            | $\beta_{11}$ | $\beta_{12}$ | $\beta_{22}$ |
|----------------------------|--------------|--------------|--------------|
| Low–                       | 0.2          | 0.1          | 0.15         |
| Low+                       | 0.25         | 0.1          | 0.2          |
| Medium–                    | 0.3          | 0.1          | 0.2          |
| Medium+                    | 0.4          | 0.1          | 0.2          |
| Medium with<br>affiliation | 0.3          | 0.1          | 0.3          |

time step over a method that would cluster separately the nodes at each time step. However, this averaged index does not say anything about the smooth recovery of group memberships along time. In particular, it is invariant under local switching on the SBM part of the parameter (see Section 2.3). Thus we also consider the global ARI value that compares the clustering of the set of nodes for all time points with the true latent structure. Obviously, good performances for this criterion are more difficult to obtain.

We use synthetic data sets created as follows. We consider binary graphs with  $N = 100$  nodes and  $T \in \{5; 10\}$  different time steps. We assume  $Q = 2$  latent groups with three different values for the transition matrix  $\pi$ :

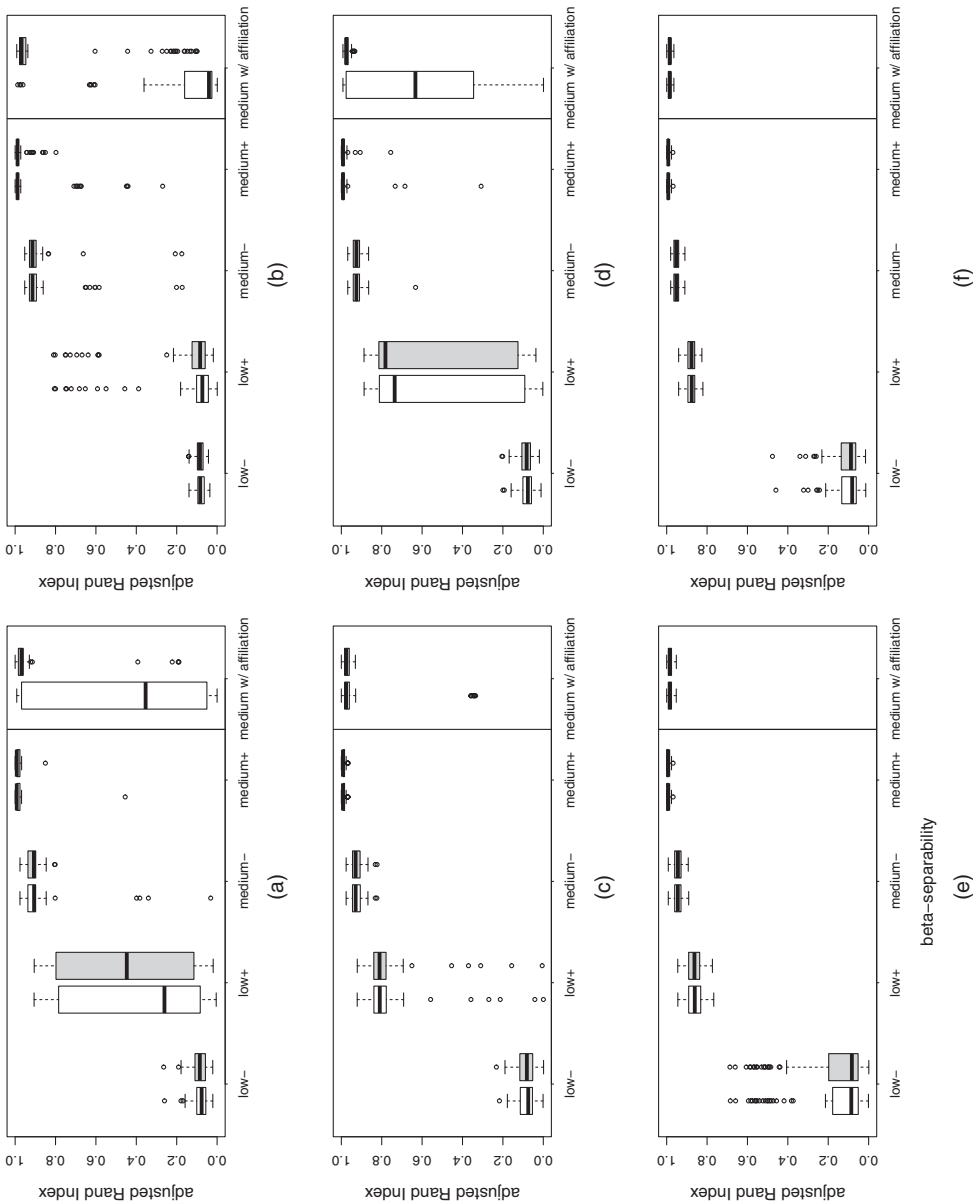
$$\begin{aligned}\pi_{\text{low}} &= \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}; \\ \pi_{\text{medium}} &= \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}; \\ \pi_{\text{high}} &= \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}.\end{aligned}$$

These three cases correspond respectively to *low*, *medium* and *high* group stability. Namely, in the first case, individuals are more likely to change group across time, resulting in a more difficult problem from the point of view of the initialization of our algorithm (see Section 3.2). The stationary distribution in those three cases is  $\alpha = (\frac{1}{2}, \frac{1}{2})$  so the two groups have similar proportions.

As for the Bernoulli parameters  $\beta$ , we explore four different cases (Table 1) representing different levels of difficulty, plus a specific example of affiliation for which we recall that the parameters are not identifiable in the dynamic setting. We note that the affiliation case satisfies the separability condition that was established for sparse planted partition models (see Mossel *et al.* (2014) for more details). This means that static reconstruction of the groups is conjectured to be possible (and we consider that this static problem corresponds to medium difficulty).

For each combination of  $(\pi, \beta)$ , we generate 100 data sets, estimate their parameters, cluster their nodes and report in Fig. 3 boxplots of a global and of an averaged ARI value.

Fig. 3 confirms that it is more difficult to obtain a smooth recovery of the groups (measured through the global ARI) than a local recovery (measured through the averaged ARI), the former values being globally smaller than the latter. In particular in the affiliation model, we observe that, whereas the averaged ARI is quite good (all values close to 1), the global ARI can be low (e.g. with low group stability or medium group stability and  $T = 10$  time points). However, in the identifiable cases, we obtain good performances for this global index (values above 0.8) when group stability is not too low or when connectivity parameters are sufficiently well separated



**Fig. 3.** Boxplots of global ARI ( $\square$ ) and averaged ARI ( $\blacksquare$ ) in various set-ups (in each panel, from left to right, the results correspond to  $\beta = \text{low-}, \text{low+}, \text{medium-}, \text{medium+}$  and the affiliation case respectively): (a) low group stability ( $\pi = \pi_{low}$ ),  $T = 5$  time points; (b) low group stability,  $T = 10$ ; (c) medium group stability ( $\pi = \pi_{medium}$ ),  $T = 5$ ; (d) medium group stability,  $T = 10$ ; (e) high group stability ( $\pi = \pi_{high}$ ),  $T = 5$ ; (f) high group stability,  $T = 10$

(medium  $\beta$ -values). As expected, the clustering performances increase (i.e. ARI values increase) with group stability (from  $\pi$  low to high) and with a better separation between the group connectivity behaviours (from low- to medium+ easiness). When increasing the number of time points from five to 10, clustering indices tend to be slightly larger, exhibiting a smaller variance. However, this is not always so: for instance, with low or medium group stability and  $\beta = \text{low+}$ , we observe that the performances decrease from five to 10 time points (smaller ARI values). We believe that this is due to the initialization of our procedure: with  $T = 10$  time points, it is more likely that the group memberships differ from their initial value. As we use as a starting point a constant with time value for these memberships, our algorithm is farther from the optimal value.

Mean-squared errors (MSEs) for estimation of the transition parameter  $\pi$  are given in the on-line supplementary Fig. 2. We show MSEs for  $\pi$  only, as the MSEs for  $(\beta, \gamma)$  are strongly correlated with the clustering results. This figure shows that, when groups are not globally recovered, the MSE values may be high (up to 15%). However, in most of the cases, these MSEs are rather small (less than 2%) so that the dynamics of the group memberships is captured.

Now, we compare our results with those of other procedures. The models of Yang *et al.* (2011) and Xu and Hero (2014) are the closest to our set-up. Since Xu and Hero (2014) obtained performances that were comparable with those of Yang *et al.* (2011), we focus on the latter here. (In fact, Xu and Hero's method is faster, with slightly lower clustering performances than that of Yang *et al.* (2011).) Thus, we use the off-line version of the algorithm that was proposed in Yang *et al.* (2011) (MATLAB code is available from [homepage.cs.uiowa.edu/~tymg/codes](http://homepage.cs.uiowa.edu/~tymg/codes)). We ran their code on the same set-up as above. When relying on default values of the algorithm, the results obtained are very poor, with ARI values that are smaller than  $10^{-2}$  in general (the data are not shown). We note that Yang *et al.* (2011) did not discuss initialization and simply proposed to start with a random partition of the nodes, which proves to be a bad strategy. To make fair comparisons, we thus decided to combine their algorithm with our initialization strategy. Results are presented in the on-line supplementary Fig. 3.

From Fig. 3 in the supplementary material we can see that, putting apart our initialization strategy, our procedure outperforms that of Yang *et al.* (2011) (they globally have much lower ARI). Indeed, their method obtains good performances only in a few cases:  $(\pi_{\text{high}}, \beta \in \{\text{medium+}; \text{medium with affiliation}\}, T \in \{5, 10\})$ ,  $(\pi_{\text{high}}, \beta \in \{\text{low+}; \text{medium-}\}, T = 5)$  and  $(\pi_{\text{medium}}, \beta \in \{\text{medium+}; \text{medium with affiliation}\}, T = 5)$ . In all these cases, we can see that the method's performances are due to very good initialization. Now, when the true classification is farther from initialization, the performances considerably drop. In particular, for intermediate cases (e.g. medium group stability or high group stability with  $T = 10$ ), we can see that our method still succeeds in obtaining a good partition (Fig. 3) whereas this is not so for the method of Yang *et al.* (2011) (supplementary Fig. 3).

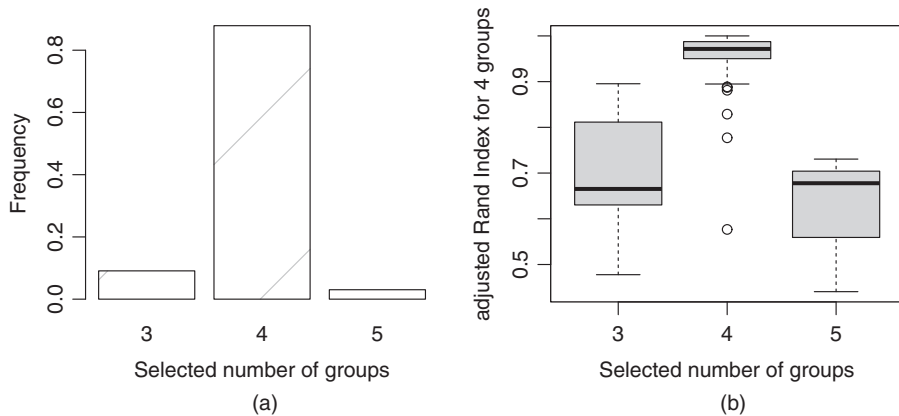
#### 4.2. Model selection

We simulate a binary dynamic data set with  $Q = 4$  groups, the transition matrix between states satisfies  $\pi_{qq} = 0.91$  and  $\pi_{ql} = 0.03$  for  $q \neq l$ . Bernoulli parameters are chosen as follows: we draw independent and identically distributed random variables  $\{\epsilon_{ql}\}_{1 \leq q \leq l \leq 4} \in [-1, 1]$  and then choose

$$\begin{aligned} \forall q \in Q, \quad \beta_{qq} &= 0.4 + 0.1\epsilon_{qq}, \\ \forall q \neq l \in Q^2, \quad \beta_{ql} &= 0.1 + 0.1\epsilon_{ql}. \end{aligned}$$

We generate 100 data sets under this model and estimate the number of groups relying on the ICL criterion. Results are presented in Fig. 4. We observe that the correct number of groups





**Fig. 4.** Estimation of the number of groups via the ICL criterion: (a) frequency of the selected number of groups; (b) ARI of the classification obtained with four groups depending on the number of groups selected

is recovered in 88% of the cases (Fig. 4(a)). Moreover, Fig. 4(b) shows that, when ICL selects only three groups, the ARI of the classification with four groups is rather low (less than 80%). This shows that, in those cases, classification with four groups is not the correct classification, so the VEM algorithm seems responsible for bad results (the optimum has not been reached) more than the penalization term.

## 5. Revealing social structure in dynamic contact networks

Dynamic network analysis has recently emerged as an efficient method for revealing social structure and organization in humans and animals. Indeed, many studies are now beyond the analysis of static networks and take advantage of longitudinal data in the long term, for instance during days or years of observations, that allow for constructing dynamic social networks. In particular, contact networks built from field observations of association between animals or from sensor-based measurements are now currently available in ecology or sociology. In this section, we show that our statistical approach is a suitable tool to analyse dynamic contact networks from the literature.

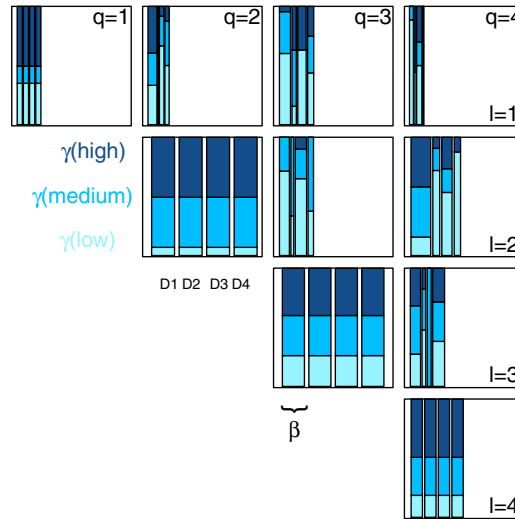
### 5.1. Encounters among high school students

Describing face-to-face contacts in a population (in our case, a classroom) can play an important role in

- (a) understanding whether there is a peculiar non-random mixing of individuals that would be a sign for a social organization and
- (b) predicting how infectious diseases can spread, by studying the cross-link between the contact dynamics and the disease dynamics.

As a first step, it is therefore mandatory to find an appropriate model to analyse these contacts and we propose to use our dynamic SBM to achieve this step.

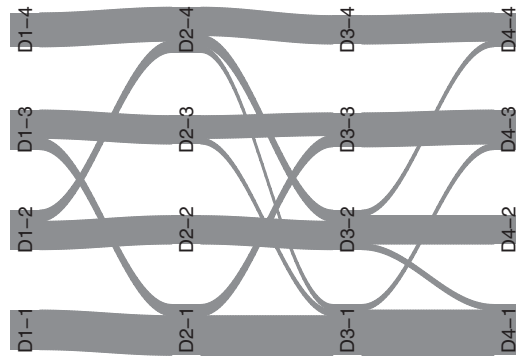
The data set consists of face-to-face encounters of high school students (measured through the use of wearable sensors) of a class from a French high school (see Fournet and Barrat (2014) for a complete description of the experiment). In this class called 'PC' (as students focus on physics and chemistry), interactions were recorded during 4 days (Tuesday–Friday) in



**Fig. 5.** Summary of the interaction parameters  $\hat{\beta}$  and  $\hat{\gamma}$  estimated by our model with  $Q = 4$  groups on the data set of interactions in the PC class (Fournet and Barrat, 2014): in each cell  $(q, l)$  with  $1 \leq q \leq l \leq 4$ , there are  $T = 4$  bar plots corresponding to the four measurements (Tuesday–Friday); each bar plot represents the distribution of the parameter  $\gamma_{ql}^t$  for the three categories of interaction frequency (low, medium and high); the width of each bar plot is proportional to the sparsity parameter  $\beta_{ql}^t$ ; we recall that, when considering the diagonal cells  $(q, q)$ , parameters do not depend on  $t$  anymore

December 2011. We kept only the 27 (out of 31) students who appear every day, i.e. that have at least one interaction with another student during each of the 4 days. Interaction times were aggregated by days to form a sequence of four different networks. These are undirected and weighted networks, the weight of an interaction between two individuals being the number of interactions between these two individuals divided by the number of time points for which at least two individuals interacted; thus a non-negative real number that we call *interaction frequency*. After examination of the distribution of these weights, we choose to discretize these data into  $M = 3$  bins (see example 2 in section S.4 in the on-line supplementary materials) corresponding to low, medium and high interaction frequency. As already mentioned, our model selection criterion is not fitted to this case (see section S.4 in the supplementary materials for more details). We thus choose to rely instead on the ‘elbow’ method, applied to the complete-data log-likelihood. It consists of identifying a change of slope in the curve that represents this complete-data log-likelihood for different values of  $Q$ . The method selects  $Q = 4$  groups (see supplementary Fig. 4) and we now present the results that were obtained with our model fitted with  $Q = 4$  groups.

We observe that groups 2 and 3 are composed of students who are likely to interact (i.e.  $\hat{\beta}_{22}$  and  $\hat{\beta}_{33}$  are close to 1; Fig. 5). Furthermore, the frequency of their interactions inside their groups is higher than in the rest of the network ( $\hat{\gamma}_{qq}(\text{low}) < \hat{\gamma}_{qq}(\text{medium}) < \hat{\gamma}_{qq}(\text{high})$  for  $q = 2, 3$  in Fig. 5). These two groups form two communities such as defined in Fortunato (2010). Moreover, we observe that both groups include a certain number of individuals (3 and 4 respectively) who permanently stay in the group over time (Fig. 6). These individuals may play the role of ‘social attractors’ or ‘core leaders’ around which the other students are likely to gravitate. Group 4 displays a similar pattern of community structure, with much less interaction (intermediate value of  $\hat{\beta}_{44}$ ) but also a significant level of interaction with group 2 (Fig. 5). Interestingly, groups 2 and 4 also exchange students over time (see the fluxes between groups in Fig. 6) and this could



**Fig. 6.** Alluvial plot showing the dynamics of the group membership estimated by our model on the data set of interactions in the PC class (Fournet and Barrat, 2014): each curve is a flux that represents the move of one or more students from a group to another group ( $D_i-k$  indicates group  $k$  for day  $i$ ); the thickness of the curves is proportional to the number of students and the total height represents the 27 students

reflect some co-operation or affinity between the students of these two groups. Group 1 is quite stable over time (seven permanent members; see Fig. 6) and is characterized by a low rate of interaction inside and outside the group (low  $\hat{\beta}_{1q}^t$ -values in Fig. 5). It clearly gathers isolated students, but this does not mean that they do not interact with any student; they usually do so, but with a small number of partners. Therefore, we not only reveal evolving communities (such as in Yang *et al.* (2011)) but we also highlight the dynamics of aloneness inside this class.

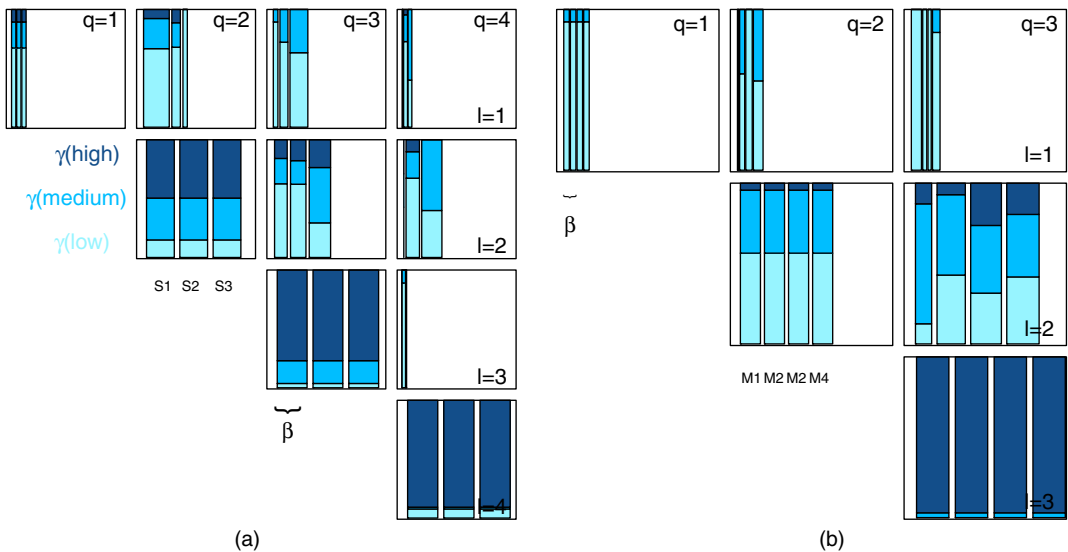
We now investigate whether gender differences may help in (*a posteriori*) explaining or refining the interaction patterns that we reveal. We first note that group 3 is exclusively composed of male students (on-line supplementary Fig. 5). This observation along with the previous conclusions suggests that group 3 may be a closed or exclusive male community. Meanwhile, some of these male students move to group 1 which is partly composed of a ‘backbone’ of female students who stay in group 1 (supplementary Fig. 5). Moreover, we clearly observe that female students are likely to stay in their group (most of the moves between groups are realized by males), and that a majority of them are in low interacting groups 1 and 4. But no female students move between these two groups, which supports a clear dichotomy pattern in the female organization with respect to male organization. In summary, we show evidence for some gender homophily (see Fournet and Barrat (2014) for a precise definition), i.e. gender is a key factor for explaining the dynamics of the interactions between these young adults.

Lastly, we note that both details captured by our model (say  $\beta$  and  $\gamma$ ) are often convergent or correlated in this case, but we note that studying this network with a binary model (i.e. not considering the interaction frequency) does not allow us to capture interesting structure (the data are not shown). Therefore, the presence or absence of an interaction as well as its frequency are important and require explicit modelling such as in our approach.

## 5.2. Social interactions between animals

Interactions between animals are dynamic processes. How and why the topology of the network changes (or not) over time is of primary interest to understand animal societies. Here we analyse two data sets of animal contact networks: the first dealing with migratory birds (*sparrows*) and the second with Indian equids (*onagers*). Both data sets are analysed with the extended model presented in section S.5 of the on-line supplementary materials.

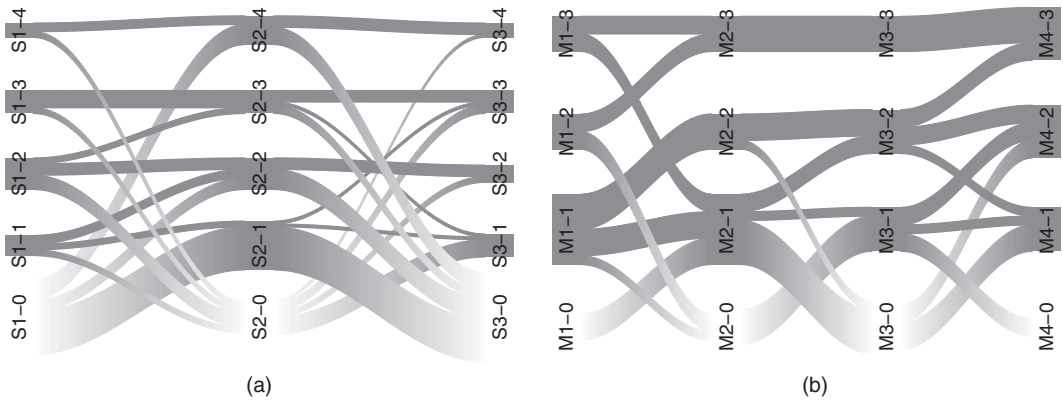
Sparrows were captured and marked during the winters of 2010–2012 in a small area (the University of California, Santa Cruz Arboretum; Shizuka *et al.* (2014)). During these three



**Fig. 7.** Summary of the interaction parameters  $\hat{\beta}$  and  $\hat{\gamma}$  estimated by our model with (a)  $Q = 4$  groups on the data set of sparrows (Shizuka *et al.*, 2014) and (b)  $Q = 3$  groups on the data set of onagers (Rubenstein *et al.*, 2015) (same principle as in Fig. 5)

seasons, Shizuka and colleagues recorded bird interactions (into *flocks*, i.e. individuals in the same place at the same time) and they aggregated their observations by season. They observed 69 birds in total, but there was a significant turnover of birds due to mortality and recruitment and only some of these 69 birds are present at each season (31, 46 and 27 birds in the first, second and third seasons respectively). The data set is therefore composed of  $T = 3$  undirected and weighted networks, with specified presence or absence of nodes at each of the three time steps. Edges are weighted by the number of times that pairs of birds have been seen together at the same place and time (if 0, there is no edge). Shizuka *et al.* (2014) identified reassembly of same communities (as defined previously) across seasons despite the turnover of birds. This stability is due to social preferences across years between individuals that rejoin the community in the same area of the site (Shizuka *et al.*, 2014). Our model is a perfect candidate to fit these observations: indeed, constraints from equation (2) are appropriate in this case where the communities keep existing over time (and therefore the parameters remain stable over time) but the membership is evolving (in particular, due to the presence or absence of birds in the three seasons). As previously, we discretized the edge weights into  $M = 3$  bins (low, medium and high interaction level) and we selected  $Q = 4$  groups with the ‘elbow’ method (the data are not shown). Examination of the estimated parameters  $\hat{\beta}$  and  $\hat{\gamma}$  (Fig. 7(a)) reveals that groups 2, 3 and 4 are clear communities (with different intragroup behaviours) that eventually correspond to those revealed by Shizuka *et al.* (2014). For most of that, our method proposes to gather peripheral birds into group 1. Clearly, we observe some stability across years with individuals staying in communities 2, 3 and 4 over time (see the horizontal fluxes in Fig. 8(a)) and that are joined by incoming birds (see fluxes from the fake group 0 of absent birds in Fig. 8). All these observations confirm the analysis in Shizuka *et al.* (2014) and demonstrate that our modelling approach is particularly suited to such data sets.

Onagers were observed in the Little Rann of Kutch, which is a desert in Gujarat, India (Rubenstein *et al.*, 2015). Each time a herd (group) was encountered, association between each pair of individuals in the group was recorded. We retained the data association of 23 individuals



**Fig. 8.** Alluvial plot showing the dynamics of the group membership estimated by our model on the data sets of interactions between (a) 69 sparrows (Shizuka *et al.*, 2014) and (b) 23 onagers (Rubenstein *et al.*, 2015) (same principle as in Fig. 6 (with  $Si-k$  and  $Mi-k$  indicating group  $k$  at season or month  $i$ )): a fake group (group 0) gathers absent animals at a specific time step and fuzzy fluxes represent arrival or departure to or from a group from or to group 0 respectively

that were present at least once between February and May 2003 and we aggregated interactions by month. The data set contains therefore  $T = 4$  undirected and weighted networks, with specified presence or absence of nodes each month. Edges are weighted by the number of times that pairs of onagers belong to the same herd (if 0, there is no edge). Again, we discretized the edge weights into  $M = 3$  bins (low, medium and high interaction level) and we selected  $Q = 3$  groups. Visual inspection of the estimated parameters (Fig. 7 (b)) shows that cluster 1 gathers peripheral onagers that can actually stay away from the others because predators have been extirpated from this habitat (and so, no collective protection strategy is required; Rubenstein *et al.* (2015)). Cluster 2 is composed of *followers* onagers which have some interactions between them and much more with those of group 3 whereas onagers in group 3 form a *rich club* community (i.e. a clique of hubs as defined in Colizza *et al.* (2006)), with high values of estimates  $\hat{\beta}_{33}$  and  $\hat{\gamma}_{33}$  (high). This community is evolving over time by integrating one or two onagers during successive months. Interestingly, the social integration process is revealed and is somewhat hierarchical: previously absent onagers (fake group 0 in Fig. 8(b)) are likely to integrate with group 1, onagers of group 1 can possibly move to the followers group (i.e. group 2) and a few followers can be integrated over time with the central rich club community (group 3). Again, the structure of the onagers social network remains persistent over time (see similar conclusions in Rubenstein *et al.* (2015)) and our model is therefore particularly adapted and efficient in this case.

## Acknowledgements

We thank Tianbao Yang for making his code available from his Web page. We sincerely acknowledge the SocioPatterns collaboration for providing the high school data set (<http://www.sociopatterns.org>) and especially Alain Barrat for interesting discussions. This work was performed by using the computing facilities of the Centre de Calcul, Laboratoire de Biométrie et Biologie Évolutive–Pôle Rhône-Alpes de Bioinformatique.

## References

- Airoldi, E., Blei, D., Fienberg, S. and Xing, E. (2008) Mixed-membership stochastic block-models. *J. Mach. Learn. Res.*, 9, 1981–2014.

- Allman, E., Matias, C. and Rhodes, J. (2009) Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.*, **37**, 3099–3132.
- Allman, E., Matias, C. and Rhodes, J. (2011) Parameters identifiability in a class of random graph mixture models. *J. Statist. Planng Inf.*, **141**, 1719–1736.
- Ambroise, C. and Matias, C. (2012) New consistent and asymptotically normal parameter estimates for random-graph mixture models. *J. R. Statist. Soc. B*, **74**, 3–35.
- Bickel, P., Choi, D., Chang, X. and Zhang, H. (2013) Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Ann. Statist.*, **41**, 1922–1943.
- Biernacki, C., Celeux, G. and Govaert, G. (2000) Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattn Anal. Mach. Intell.*, **22**, 719–725.
- Celisse, A., Daudin, J.-J. and Pierre, L. (2012) Consistency of maximum-likelihood and variational estimators in the Stochastic Block Model. *Electron. J. Statist.*, **6**, 1847–1899.
- Colizza, V., Flammini, A., Serrano, M. and Vespignani, A. (2006) Detecting rich-club ordering in complex networks. *Nat. Phys.*, **2**, 110–115.
- Daudin, J.-J., Picard, F. and Robin, S. (2008) A mixture model for random graphs. *Statist. Comput.*, **18**, 173–183.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc. B*, **39**, 1–38.
- Fortunato, S. (2010) Community detection in graphs. *Phys. Rep.*, **486**, 75–174.
- Fournet, J. and Barrat, A. (2014) Contact patterns among high school students. *PLOS One*, **9**, article e107878.
- Goldenberg, A., Zheng, A. X., Fienberg, S. E. and Airoldi, E. M. (2010) A survey of statistical network models. *Found. Trends Mach. Learn.*, **2**, 129–233.
- Gunawardana, A. and Byrne, W. (2005) Convergence theorems for generalized alternating minimization procedures. *J. Mach. Learn. Res.*, **6**, 2049–2073.
- Heaukulani, C. and Ghahramani, Z. (2013) Dynamic probabilistic models for latent feature propagation in social networks. *JMLR Wrkshp Conf. Proc.*, **28**, 275–283.
- Herlau, T., Mørup, M. and Schmidt, M. (2013) Modeling temporal evolution and multiscale structure in networks. *JMLR Wrkshp Conf. Proc.*, **28**, 960–968.
- Hoff, P., Raftery, A. and Handcock, M. (2002) Latent space approaches to social network analysis. *J. Am. Statist. Ass.*, **97**, 1090–1098.
- Holme, P. (2015) Modern temporal network theory: a colloquium. *Eur. Phys. J. B*, **88**, 234.
- Hubert, L. and Arabie, P. (1985) Comparing partitions. *J. Classificn*, **2**, 193–218.
- Ishiguro, K., Iwata, T., Ueda, N. and Tenenbaum, J. B. (2010) Dynamic infinite relational model for time-varying relational data analysis. In *Advances in Neural Information Processing Systems*, vol. 23 (eds J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel and A. Culotta), pp. 919–927. New York: Curran Associates.
- Jordan, M., Ghahramani, Z., Jaakkola, T. and Saul, L. (1999) An introduction to variational methods for graphical models. *Mach. Learn.*, **37**, 183–233.
- Kolaczyk, E. D. (2009) *Statistical Analysis of Network Data: Methods and Models*. New York: Springer.
- Latouche, P., Birmel  , E. and Ambroise, C. (2012) Variational Bayesian inference and complexity control for stochastic block models. *Statist. Modllng*, **12**, 93–115.
- Leroux, B. G. (1992) Maximum-likelihood estimation for hidden Markov models. *Stoch. Processes. Appl.*, **40**, 127–143.
- Liu, S., Wang, S. and Krishnan, R. (2014) Persistent community detection in dynamic social networks. *Lect. Notes Comput. Sci.*, **8443**, 78–89.
- Matias, C. and Robin, S. (2014) Modeling heterogeneity in random graphs through latent space models: a selective review. *ESAIM Proc.*, **47**, 55–74.
- Mossel, E., Neeman, J. and Sly, A. (2014) Reconstruction and estimation in the planted partition model. *Probab. Theor. Reltd Flds*, **162**, 431–461.
- Rubenstein, D. I., Sundaresan, S. R., Fischhoff, I. R., Tantipathananandh, C. and Berger-Wolf, T. Y. (2015) Similar but different: dynamic social network analysis highlights fundamental differences between the fission-fusion societies of two equid species, the onager and Grevy’s zebra. *PLOS One*, **10**, article e0138645.
- Sarkar, P. and Moore, A. W. (2005) Dynamic social network analysis using latent space models. *SIGKDD Explorns Newslett.*, **7**, no. 2, 31–40.
- Shizuka, D., Chaine, A. S., Anderson, J., Johnson, O., Laursen, I. M. and Lyon, B. E. (2014) Across-year social stability shapes network structure in wintering migrant sparrows. *Ecol. Lett.*, **17**, 998–1007.
- Snijders, T. A. (2011) Statistical models for social networks. *A. Rev. Sociol.*, **37**, 129–151.
- Teicher, H. (1967) Identifiability of mixtures of product measures. *Ann. Math. Statist.*, **38**, 1300–1302.
- Xing, E. P., Fu, W. and Song, L. (2010) A state-space mixed membership blockmodel for dynamic network tomography. *Ann. Appl. Statist.*, **4**, 535–566.
- Xu, K. and Hero, A. (2014) Dynamic stochastic blockmodels for time-evolving social networks. *IEEE J. Select. Top. Signl Process.*, **8**, 552–562.
- Xu, A. and Zheng, X. (2009) Dynamic social network analysis using latent space model and an integrated clustering algorithm. In *Proc. 8th Int. Conf. Dependable, Autonomic and Secure Computing*, pp. 620–625. New York: Institute of Electrical and Electronics Engineers.

- Yang, T., Chi, Y., Zhu, S., Gong, Y. and Jin, R. (2011) Detecting communities and their evolutions in dynamic social networks—a Bayesian approach. *Mach. Learn.*, **82**, 157–189.
- Zanghi, H., Ambroise, C. and Miele, V. (2008) Fast online graph clustering via Erdős Rényi mixture. *Pattern Recogn.*, **41**, 3592–3599.
- Zanghi, H., Picard, F., Miele, V. and Ambroise, C. (2010) Strategies for online inference of model-based clustering in large and growing networks. *Ann. Appl. Statist.*, **4**, 687–714.
- Zreik, R., Latouche, P. and Bouveyron, C. (2016) The dynamic random subgraph model for the clustering of evolving networks. *Computational Statist.*, to be published.

#### Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supplementary Materials for: "Statistical clustering of temporal networks through a dynamic stochastic block model"'.  
[http://www.tandfonline.com/supplemental/doi:10.1080/10618600.2016.1201111](#)