

Dynamic Mixed Membership Stochastic Block Model for Weighted Labeled Networks

Anonymous Author(s)

ABSTRACT

Most real-world networks evolve over time. Existing literature proposes models for dynamic networks that are either unlabeled or assumed to have a single membership structure. On the other hand, a new family of Mixed Membership Stochastic Block Models (MMSBM) allows to model static labeled networks under the assumption of mixed-membership clustering. In this work, we propose to extend this later class of models to infer dynamic labeled networks under a mixed membership assumption. Our approach takes the form of a temporal prior on the model's parameters. It relies on the single assumption that dynamics are not abrupt. We show that our method significantly differs from existing approaches, and allows to model more complex systems –dynamic labeled networks. We demonstrate the robustness of our method with several experiments on both synthetic and real-world datasets. A key interest of our approach is that it needs very few training data to yield good results. The performance gain under challenging conditions broadens the variety of possible applications of automated learning tools –as in social sciences, which comprise many fields where small datasets are a major obstacle to the introduction of machine learning methods.

CCS CONCEPTS

• **Information systems** → **Clustering**; *Social networks*; *Social recommendation*; **Data mining**; *Collaborative filtering*; *Network data models*; • **Mathematics of computing** → *Probabilistic inference problems*; **Bayesian networks**.

KEYWORDS

Block models, Dynamic networks, Clustering, Recommender systems

ACM Reference Format:

Anonymous Author(s). 2018. Dynamic Mixed Membership Stochastic Block Model for Weighted Labeled Networks. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 9 pages. <https://doi.org/XXXXXX.XXXXXXX>

1 INTRODUCTION

Dynamic networks are powerful tools to visualize and model interactions between different entities that can evolve over time. The

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00
<https://doi.org/XXXXXX.XXXXXXX>

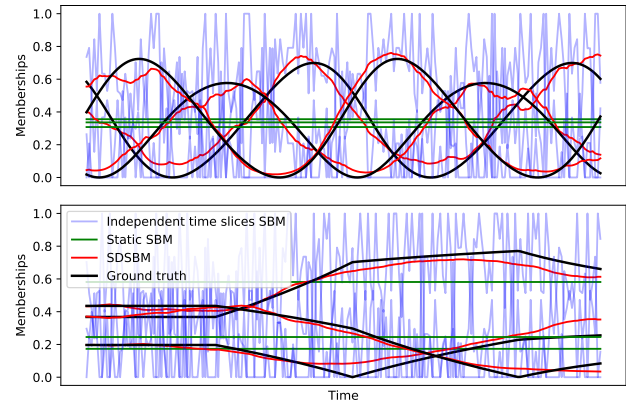


Figure 1: Users' attachment to groups can vary over time — A music listener could cyclically prefer Rock, Jazz or Pop music (top), or listen to either of these without any specific pattern (bottom). For 200 epochs containing only 5 observations each, our approach (in red) infers any smooth dynamic membership pattern, and does it more accurately than static models (in green [9]) and models that consider each time slice independently (in blue [22]).

network's nodes represent the interacting entities, and ties between these nodes represent an interaction. In many real-world situations, ties strength can vary over time –on music streaming websites for instance, users' affinity with various musical genres vary over time [14, 23]. The network is said dynamic.

Now, every interaction does not have the same significance. A music listener might like both Rock and Jazz, but favor one over the other. This user's ties to musical genres do not have the same intensity; each tie is associated to a number, representing the strength of the interaction. The network is said to be dynamic and valued.

Finally, simple dynamic valued networks might not be fit to grasp the complexity of a given situation. A music listener may have different opinions on musical genres; they can like it, dislike it, be bored of it, prefer to listen to some only in the morning, at night, etc. Each of these relations can be represented by their own tie in the network, each associated to their own strength. The network is said to be dynamic, valued and labeled.

Such networks are high-dimensional objects, whose direct inference is a difficult problem. Several ways to achieve this task have been proposed; the Stochastic Block Models (SBM) family is one of the most popular approaches [5, 10, 12]. The underlying assumption is that some sets of nodes behave in a similar way regarding other sets of nodes. They can be grouped in a cluster; instead of modeling every edge for every node in a network, only the edges

between clusters are modeled, making the inference task much more tractable. Each cluster is associated to labeled edges, and each node is associated to a cluster. A variant of SBM that allows more expressive power is called the Mixed-Membership SBM (MMSBM), where each node can belong to several clusters with different proportions [2, 9, 22, 24]. A major advantage of the SBM family is that they yield readily interpretable results (interpretable clusters), unlike most existing neural-network approaches for labeled networks [6].

The goal of this paper is to infer networks that are dynamic, valued and labeled by using a mixed-membership SBM. Our contribution consists in the extension of a broad class of MMSBM by using a temporal Dirichlet prior. We first conduct a careful review of dynamic network inference literature, and we could not find any prior work addressing the task at hand. Although some previous works tackle similar problems, they are not fit for the task at hand. Besides, our approach is conceptually much simpler than those proposed in the literature. Last but not least, our dynamic extension is readily pluggable into most existing MMSBM for labeled and valued networks.

We develop an EM optimization algorithm that scales linearly with the size of the dataset, demonstrate the effectiveness of our approach on both synthetic and real-world datasets, and further detail a possible application scenario.

2 BACKGROUND

2.1 Notations

We consider a network of I nodes and O labels. All the clustering models discussed in this section can be represented by a restricted set of parameters: a matrix $\theta(t) \in \mathbb{R}^{I \times K}$ accounts for the membership of each of I nodes to each of K possible clusters, and a block-interaction matrix $p(t) \in \mathbb{R}^{K \times K \times O}$ represents the probability for each of K clusters to yield each of O labels. Both $\theta(t)$ and $p(t)$ can vary over time; in the following, this temporal dependence is implicit ($x(t) := x$) unless specified otherwise. The network is said to be unlabeled when $O = 1$, and binary (as opposed to valued) if an edge can only exist or not exist.

2.2 Dynamic unlabeled networks - Single-membership

Single-membership SBMs consider a membership matrix such as $\theta \in \{0; 1\}^{I \times K}$: each membership vector equals 1 for one cluster, and 0 everywhere else –“hard” clustering in the literature. In [28, 29], the authors proposed to model a binary unlabeled dynamic network using a label-switching inference method, optimized by a Sequential Monte Carlo algorithm [13]. Both the membership and the interaction matrices can vary over time, thus supposing two independent underlying Markov processes [13].

In [21, 30], the authors propose to model a binary dynamic and unlabeled network. The cluster interaction matrix p can vary over time while keeping the memberships θ static. The entries of p are drawn from a Dirichlet distribution and expressed as a Chinese Restaurant Process. This process converges to a Dirichlet distribution, and allows to infer a potentially infinite number of clusters. This model is therefore non-parametric and inferred using an MCMC algorithm.

A conceptually novel way of modeling dynamic unlabeled networks under the single membership assumption has been proposed in [16, 17]. The authors propose to model the cluster interaction and membership matrices as Poisson processes, that explicitly represent the continuous temporal dependency without slicing the dataset into episodes. The method allows to infer varying membership and interaction matrices for dynamic binary or valued networks, but their results have shown that allowing both to vary simultaneously leads to identifiability and label switching issues [8]. This conclusion seems reasonable, given none of these SBM algorithms can reach a global optimum of the likelihood function. A model where both θ and p can vary over time is more likely to get stuck into a local optimum.

Finally, we mention the existence of SBM variants that account for dynamic degree-correction [25] or that enforce a scale-free characteristic [26].

All these methods consider unlabeled networks, and consider a hard clustering which does not allow for as much expressive power as the Mixed-Membership approaches.

2.3 Dynamic unlabeled networks - Mixed-membership

Mixed-membership SBMs consider a membership matrix such as $\theta \in \mathbb{R}^{I \times K}$, where each membership vector θ_i has a L_1 norm of 1. Literature also refer to it as “soft” clustering.

Similar to [21, 30], a method for inferring dynamical binary unlabeled networks has been proposed in [7]. The membership vector of each piece of information is drawn from a Chinese Restaurant Process (CRP) according to the number of times a node has already been associated to each cluster before. The process yields a distribution over an infinity of available clusters. The formulation as a CRP arises naturally, because the prior on membership vectors is typically a Dirichlet distribution –a CRP naturally converges to a draw from a Dirichlet distribution [3]. The block-interaction matrix p does not vary over time. The article shows a complexity analysis that suggest the methods runs with a complexity of $O(N^2)$ which makes it unfit for large-scale real-world applications.

The work the most closely related to ours is [27]. This seminal work proposed the dMMSB as a way to model dynamic binary unlabeled networks using a variational algorithm [15]. To do so, the authors modify the original MMSBM [2] to consider a logistic normal distribution as a prior on the membership vectors θ_i . This choice allows to model correlations between membership vectors’ evolution [1]. The membership vectors are then embedded in a state space model, that is a space where one can define a linear transition between two time points for a given variable. The authors define such trajectory for the membership vectors as a linear function of the previous time point. The trajectory is estimated and smoothed using a Kalman Filter. This approach is the most closely related to ours, as it models the temporal dependency using a prior distribution over memberships, noted $P(\theta)$. [27] has been extended to consider $P(\theta)$ as a logistic normal mixture prior [11], which improves its expressive power.

However, this model and related extensions are not fit for the task at hand. First, it considers unlabeled and binary networks [15], and extension to labeled and valued networks, if possible, is not

trivial. The proposed optimization algorithm requires a loop until convergence at each EM iteration, making it unable to handle large datasets. Besides, the clusters interaction matrix p must remain static for the approach to work, which we alleviate here. And most importantly, it assumes a linear transition between time steps in the state space, while we do not assume any kernel function in our proposed approach.

2.4 Static labeled networks - Mixed-membership

Recent years saw a rise of Bayesian methods for inferring static valued labeled networks using MMSBM variants [9, 18–20, 22]. Here, nodes are associated a type a ; each type of node has its own layer in a multipartite network, and its associated set of available clusters through a membership matrix θ^a . Nodes from one layer belong to every cluster corresponding to this layer with mixed proportions. The interaction between clusters of all layers then yields a distribution over O possible output labels. This results in a static MMSBM for labeled networks. Our contribution extends all of [9, 18–20, 22] to the dynamic case. These works will be further detailed in Section 3.

Note that in [22], the authors consider a temporal slicing of the data and consider each slice as independent from the others; a time slice is considered as a node in a tripartite network. We will compare our approach to this modeling later.

We use these works as a base model, that we couple to a dynamic prior distribution on parameters. Our work focuses on making the prior probability of both θ and p time-dependent, in order to model these parameters dynamics. We provide a ready-to-use temporal plug-in for each of the works presented in this section. It applies to dynamical, valued and labeled networks in a mixed-membership context, and inference is conducted with a scalable variational EM algorithm.

3 DYNAMIC LABELED MMSBM

3.1 Base model

For clarity, we choose to consider the **simplest form** of a labeled MMSBM: each of I nodes is associated to each of K clusters in mixed proportions, and each of K clusters is in turn associated to O labels (SIMSBM(1) in [20]). As shown in [20], our demonstration on SIMSBM(1) trivially extends to [9] (SIMSBM(1,1)) [22] (SIMSBM(2,1)) and [19] (SIMSBM(2)) –and in general any model formulated as SIMSBM(x,y,...).

We consider a set of I nodes that can be associated to O possible labels on a discrete time interval, or epoch, written t . We assume that, at each time step, each of I nodes belongs to a mixture of K available clusters, each of which are in turn yield a probability distribution over O labels. The membership of each of I nodes to each of the K possible clusters at time t is encoded in the membership matrix $\theta^{(t)} \in \mathbb{R}^{I \times K}$. One vector $\theta_i^{(t)}$ represents the probability that i belongs to any of the K clusters at time t , and is normalized as:

$$\sum_{k \in K} \theta_{i,k}^{(t)} = 1 \quad \forall i, t \quad (1)$$

The probability for each of K clusters to be associated to each of O labels at time t is encoded in the matrix $p^{(t)} \in \mathbb{R}^{K \times O}$. An entry

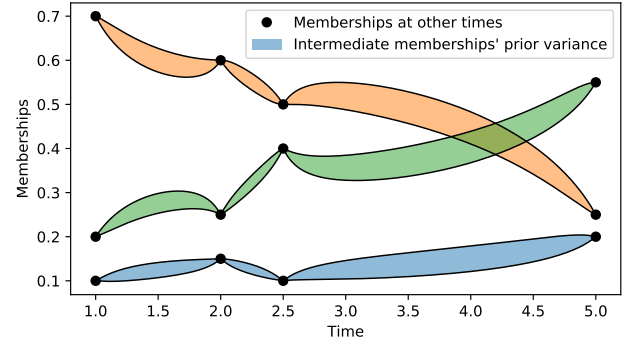


Figure 2: Prior probability’s variance on memberships at all times according to the temporal neighbourhood –Variance of the prior over a membership entry (filled curves, we represented 3 such entries as illustration) as a function of time, given some temporal neighbours (black dots). This illustration considers an averaging kernel as $\kappa(t, t') = \frac{1}{|t-t'|}$. When inferring a parameter $x^{(t)}$ at a time t , the variance of its prior probability $P(x^{(t)})$ depends on t relative to the temporal neighbours. Here for instance, the variance is null at $t = 2$ because $\kappa(t, t')$ diverges, and so does $\alpha^{(t)}$, hence the variance collapsing to 0.

$p_k^{(t)}(o)$ represents the probability that cluster k is associated to label o at time t , and thus is normalized as:

$$\sum_{o \in O} p_k^{(t)}(o) = 1 \quad \forall k, t \quad (2)$$

Finally, the probability that a node i is associated to label o at time t (i.e. the probability of an edge between i and o at time t) is written:

$$P(i \rightarrow o|t) = \sum_{k \in K} \theta_{i,k}^{(t)} p_k^{(t)}(o) \quad (3)$$

Given a set R^o of observed triplets (i, o, t) , the model’s posterior distribution can be written [9, 20]:

$$\begin{aligned} P(\theta, p|R^o) &\propto P(R^o|\theta, p) \prod_t P(\theta^{(t)}) P(p^{(t)}) \\ &= \prod_{(i,o,t) \in R^o} \sum_{k \in K} \theta_{i,k}^{(t)} p_k^{(t)}(o) \prod_t \left(\prod_i P(\theta_i^{(t)}) \prod_k P(p_k^{(t)}) \right) \end{aligned} \quad (4)$$

Now, before we describe the optimization procedure, we must choose the priors $P(\theta^{(t)})$ and $P(p^{(t)})$.

3.2 Simple Dynamic prior

We formulate the prior distribution over $\theta^{(t)}$ and $p^{(t)}$ under a single assumption: the parameters at a given time do not vary abruptly at small time scales. It means an entry $\theta_{i,k}^{(t_1)}$ should not differ significantly from $\theta_{i,k}^{(t_2)}$ for every t_2 close enough to t_1 . The entries close to a reference time are called **temporal neighbours**.

Our *a priori* knowledge on each entry $\theta_i^{(t)}$ and $p_k^{(t)}$ is that they should not differ significantly from their temporal neighbours. This is a fundamental difference with [27], where the next parameters

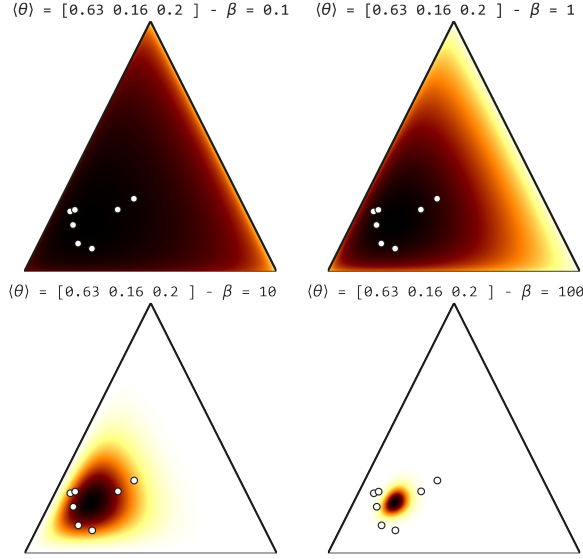


Figure 3: Prior probability on a membership vector for various values of β according to temporal neighbourhood – Darker means higher probability. Projected on a simplex tri-space (each of 3 axes ranges from 0 to 1). The white dots represent the temporal neighbours of the considered 3D vector. Their average is given as $\langle \theta \rangle$ using a uniform weight function $\kappa(t, t')$ for illustration purpose. β controls the variable's prior variance around its neighbours.

values are estimated using from a Kalman Filter that only considers the previous time step. Moreover, the authors assume a linear transition function, while we do not make such hypothesis. An illustration of the proposed approach is given Fig. 2, where the prior probability of a membership over time depends on its temporal neighbours.

3.2.1 Dirichlet distribution. Since each entry $\theta_i^{(t)}$ and $p_k^{(t)}$ is normalized to 1, we consider a Dirichlet distribution as a prior, which naturally yields normalized vectors such that $\sum_n x_n = 1$. It reads:

$$Dir(x|\alpha) = \frac{1}{B(\alpha)} \prod_n x_n^{\alpha_n - 1} \quad (5)$$

where $B(\cdot)$ is the multivariate beta function. In Eq. 5, the vector α is called the concentration parameter and must be provided to the model. This parameter defines the mode and the variance of the Dirichlet distribution.

We consider a concentration parameter as $\alpha = 1 + \beta\alpha_0$, so that when $\beta = 0$ we recover a uniform prior over the simplex, and $\beta \geq 0$ so that the prior has a unique mode. The most frequent value drawn from Eq.5 (or mode) is $\mathbb{M}(x_n) = \frac{\alpha_n - 1}{\sum_n (\alpha_n - 1)} = \frac{\alpha_{0,n}}{\sum_n \alpha_{0,n}}$. We recover a uniform prior for $\beta = 0$; the variance vanishes with $\beta \gg 1$ as $\frac{1}{\beta}$. The effect of various values of β on the prior distribution is illustrated in Fig. 3.

3.2.2 Simple Dynamic prior's mode. Our single constraint is that $\theta_i^{(t)}$ and $p_k^{(t)}$ do not vary abruptly over small time scales. To do so,

we define their prior probability's mode with respect to close temporal neighbours. The hyper-parameter β controls the variance of the prior –that is, how much our constraint should impact the inference procedure. We express the Simple Dynamic prior parameters for $\theta_i^{(t)}$ as:

$$\alpha_{i,k}^{(t,\theta)} = 1 + \beta \underbrace{\left(\frac{\sum_{t' \neq t} \kappa(t, t') \theta_{i,k}^{(t')}}{\sum_{t' \neq t} \kappa(t, t')} \right)}_{\langle \theta_{i,k}^{(t)} \rangle} \quad (6)$$

where $\kappa(t, t')$ is a weight function, and $\alpha^{(t,\theta)}$ corresponds to the concentration parameter for θ at time t . In following experiments, we define the weight function as $\kappa(t, t') = \frac{N_{t'}}{|t - t'|}$, where $N_{t'}$ is the number of observations made at time t' . This way, temporal neighbours' influence decrease as the inverse of temporal distance. We illustrate the influence of this particular kernel function on the prior probability on membership at all times in Fig. 2. In particular we see that with this expression, the prior probability variance goes to 0 when the considered time is very close to a temporal neighbour.

The mode of the prior is then the average value of its temporal neighbours weighted by $\kappa(t, t')$, noted $\langle \theta_{i,k}^{(t)} \rangle$. Note that this holds because $\sum_k \langle \theta_{i,k}^{(t)} \rangle = 1 \forall i, t$. Besides, the prior variance is a decreasing function of β ; when $\beta = 0$ the prior is uniform over the simplex, and when $\beta \rightarrow \infty$ the variance goes to 0, as illustrated Fig. 3. The same reasoning holds for $p_k^{(t)}$, with prior parameters $\alpha_{k,o}^{(t,p)} = 1 + \beta \langle p_k^{(t)}(o) \rangle$, as well as for any other parameter in [9, 19, 20, 22].

3.2.3 Priors expression. Finally, we give the final log-priors on $\theta_i^{(t)}$ and $p_k^{(t)}$:

$$P(\theta_i^{(t)} | \{\theta_{i,k}^{(t')}\}_{t' \neq t}) \propto \prod_k \theta_{i,k}^{(t) \beta \langle \theta_{i,k}^{(t)} \rangle} \quad (7)$$

$$P(p_k^{(t)}(o) | \{p_k^{(t')}(o)\}_{t' \neq t}) \propto \prod_o p_k^{(t)}(o)^{\beta \langle p_k^{(t)}(o) \rangle}$$

We omitted the normalisation factor for clarity –it does not influence the inference procedure.

3.3 Inference

3.3.1 E step. We develop an EM inference procedure for maximizing the log-posterior distribution defined Eq.8. The expectation step computes the expected probability of a latent variable (here a cluster k) being chosen given each entry of R° . Since such latent variables do not appear in the priors expressions, the expectation step remains unchanged by the introduction of the Simple Dynamic Priors; in general, prior distributions do not intervene in the computation of the expectation step [4]. The E step for such labeled networks has already been discussed on several occasions [9, 19, 22].

It can be derived as:

$$\begin{aligned}
\log P(\theta, p | R^\circ) &\propto \log P(R^\circ | \theta, p) \prod_t \prod_i P(\theta_i^{(t)}) \prod_k P(p_k^{(t)}) \\
&= \sum_{(i,o,t) \in R^\circ} \log \sum_{k \in K} \theta_{i,k}^{(t)} p_k^{(t)}(o) \\
&\quad + \sum_t \sum_i \log P(\theta_i^{(t)}) \sum_k \log P(p_k^{(t)}) \\
&\geq \sum_{(i,o,t) \in R^\circ} \sum_{k \in K} \omega_{i,o}^{(t)}(k) \log \frac{\theta_{i,k}^{(t)} p_k^{(t)}(o)}{\omega_{i,o}^{(t)}(k)} \\
&\quad + \sum_t \sum_i \log P(\theta_i^{(t)}) \sum_k \log P(p_k^{(t)})
\end{aligned} \tag{8}$$

In Eq.8, we introduced a proposal distribution $\omega_{i,o}^{(t)}(k)$, that represents the probability of one cluster allocation k given the observation (i, o, t) . The last line followed from Jensen's inequality assuming that $\sum_k \omega_{i,o}^{(t)}(k) = 1$. We notice that Jensen's inequality holds as an equality when:

$$\omega_{i,o}^{(t)}(k) = \frac{\theta_{i,k}^{(t)} p_k^{(t)}(o)}{\sum_{k'} \theta_{i,k'}^{(t)} p_{k'}^{(t)}(o)} \tag{9}$$

which provides us with the expectation of the latent variable k given an observation $(i, o, t) \in R^\circ$, written $\omega_{i,o}^{(t)}(k)$.

Using this expression, we can rewrite the log-likelihood $\log P(R^\circ | \theta, p)$ as [4, 9, 20]:

$$\log P(R^\circ | \theta, p) = \sum_{(i,o,t) \in R^\circ} \sum_{k \in K} \omega_{i,o}^{(t)}(k) \log \frac{\theta_{i,k}^{(t)} p_k^{(t)}(o)}{\omega_{i,o}^{(t)}(k)} \tag{10}$$

3.3.2 M step. Taking back the first line of Eq.8 and substituting with Eq.7 and Eq.10, we get an unconstrained expression of the posterior distribution. We introduce Lagrange multipliers to account for the constraints of Eq.1 ($\phi_i^{(t)}$) and Eq.2 ($\psi_i^{(t)}$), and finally compute the maximization equations with respect to the model's parameters. Starting with the membership matrix entries $\theta_{i,k}^{(t)}$:

$$\begin{aligned}
&\frac{\partial \left(\log P(\theta, p | R^\circ) - \sum_{i',t'} \phi_{i'}^{(t')} (\sum_{k'} \theta_{i',k'}^{(t')} - 1) \right)}{\partial \theta_{i,k}^{(t)}} = 0 \\
&\Leftrightarrow \sum_{o \in \partial(i,t)} \frac{\omega_{i,o}^{(t)}(k)}{\theta_{i,k}^{(t)}} + \frac{\beta \langle \theta_{i,k}^{(t)} \rangle}{\theta_{i,k}^{(t)}} - \phi_i^{(t)} = 0 \\
&\Leftrightarrow \sum_{o \in \partial(i,t)} \omega_{i,o}^{(t)}(k) + \beta \langle \theta_{i,k}^{(t)} \rangle = \phi_i^{(t)} \theta_{i,k}^{(t)} \\
&\Leftrightarrow \sum_{o \in \partial(i,t)} \underbrace{\sum_k \omega_{i,o}^{(t)}(k)}_{=1 \text{ (Eq.9)}} + \beta \underbrace{\sum_k \langle \theta_{i,k}^{(t)} \rangle}_{=1 \text{ (Eq.1)}} = \phi_i^{(t)} \\
&\Leftrightarrow \frac{\sum_{o \in \partial(i,t)} \omega_{i,o}^{(t)}(k) + \beta \langle \theta_{i,k}^{(t)} \rangle}{N_{i,t} + \beta} = \theta_{i,k}^{(t)} \tag{11}
\end{aligned}$$

where $\partial(i, t) = \{o | (i, o, t) \in R^\circ\}$ is the subset of labels associated to both i and t , and $N_{i,t} = |\partial(i, t)|$ is the size of this set. Note that for

$\beta = 0$ we recover the M-step of standard static MMSBM models [9, 19, 20, 22].

The derivation of the M-step for the entries $p_k^{(t)}(o)$ is identical and yields :

$$p_k^{(t)}(o) = \frac{\sum_{(i,t) \in \partial o} \omega_{i,o}^{(t)}(k) + \beta \langle p_k^{(t)}(o) \rangle}{\sum_{(i,o,t) \in R^\circ} \omega_{i,o}^{(t)}(k) + \beta} \tag{12}$$

3.4 Discussion

3.4.1 Plug-in for existing models. Using SDSBM as a temporal extension in existing MMSBM models requires very few changes. In [9, 19, 20, 22], its introduction boils down to adding a term $\beta \langle x \rangle$ to the numerator of maximization equations, and the corresponding normalizing term β to the denominator. This way, our approach is ready-to-use to make these models for modelling dynamic networks.

3.4.2 Flexible dynamic modeling. The prior allows to consider that some parameters are dynamic and that others are not. For instance, when several membership matrices are involved, as in [9, 20, 22]), setting $\beta = 0$ for some makes them time-invariant (or universal). The Simple Dynamic prior also allows to choose whether the block-interaction tensor p is dynamic. Moreover, β does not have to be the same for every membership matrix, or even every entry i of each of them. Finally, β itself can vary over time. To summarize, β allows to control the temporal scale over which parameters may vary. This allows to jointly model universal parameters ($\beta = 0$) and dynamical ones ($\beta \neq 0$).

3.4.3 Tunable temporal dependence. Finally, the choice of the averaging kernel function $\kappa(t, t')$ is important. It allows to choose the range over which the inference of a variable should rely on its temporal neighbours. A formulation as the inverse of time difference seems relevant: the weight of a neighbour appearing at a time δt later should diverge as $\delta t \rightarrow 0$, so that continuity is ensured. Besides, one could control the smoothness of the curve with respect to time by tuning the weight function as $\kappa(t, t') = \frac{N_{t'}}{|t-t'|^a}$ where $a = 1, 2, \dots$ for instance, where $N_{t'}$ is the number of observations in the time slice t' . Overall, the Simple Dynamic prior works by inferring the variables using both microscopic and mesoscopic temporal scales. If a time slice t has few observations but some of its neighbours have a greater number of them, learning the parameters at t is helped mostly by the population of its closest ($\frac{1}{|\Delta t|}$) and most populated ($N_{t'}$) neighbours, and less influenced by further and less populated epochs. This is what is illustrated in Fig. 2.

4 EXPERIMENTS

4.1 Synthetic data

In this section, we develop several situations in which our method (abbreviated SDSBM for Simple Dynamic MMSBM) could be useful¹. Experiments are run for $I = 100$, $K = 3$ and $O = 3$, which are standard testing parameters in the literature of dynamic networks inference [7, 16]. We choose to infer a dynamic membership matrix $\theta^{(t)}$ and to provide a universal block-interaction matrix $p \forall t$. Note that the model yields good performances when p also has to be

¹Code and data available at <https://anonymous.4open.science/r/SDSBM-26DC>

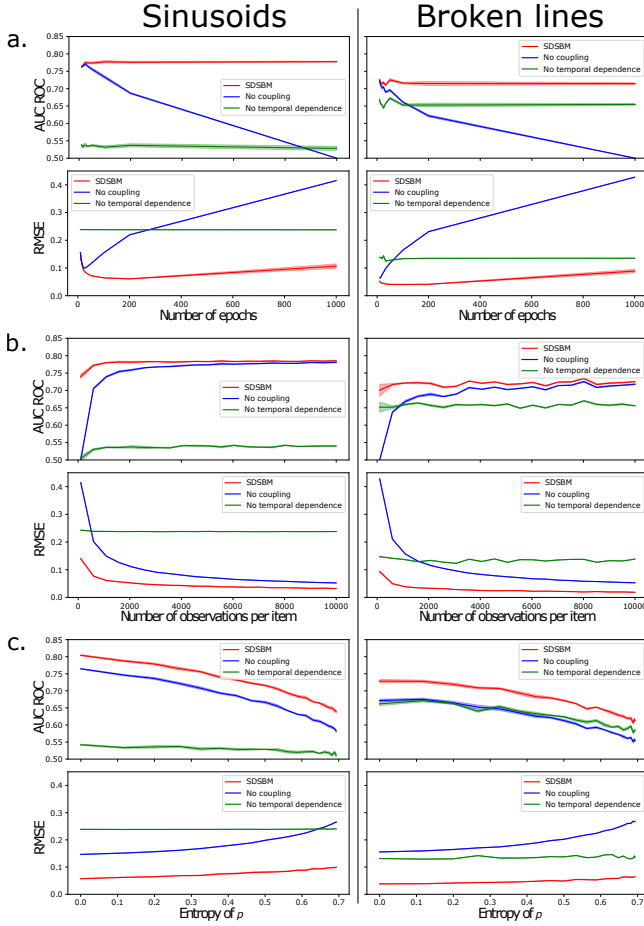


Figure 4: Results on synthetic data – (a.) SDSBM retrieves the correct dynamic memberships and is little influenced by the data slicing. (b.) SDSBM works well on tiny datasets. (c.) SDSBM retrieves dynamic memberships in challenging situations.

inferred, but due to identifiability and label-switching issues raised in [16], there is not unbiased way to assess the correctness of the inferred memberships values. Additional experiments showed that inferring jointly p and θ yields the same prediction accuracy. The expression we use for this matrix p is given Eq.13. We systematically test two variation patterns for θ : a sinusoidal pattern (Fig. 1-top) and a broken-line pattern (Fig. 1-bottom). Each pattern is generated with different coefficient for each item; the memberships still sum to 1 at all times.

$$p = \begin{bmatrix} 1-s & s & 0 \\ 0 & 1-s & s \\ s & 0 & 1-s \end{bmatrix} \quad (13)$$

To the best of our knowledge, the only attempt to model dynamic parameters in labeled valued and dynamic networks using a MMSBM is [22]. In this work, each epoch is modeled independently from the others. We refer to this baseline as the “No coupling” or

“NC” baseline. For reference, we also compare to a baseline that does not consider the temporal dimension and infers a single universal value for each variable (**SIMSBM(1)**) [20]. It has already been assessed that MMSBMs typically outperform state of the art models (NMF, TF, KNN, etc.) in [9] and [20]. Such study is not repeated in the present article, for being out of the scope of our demonstration. We recall that **we purposely chose the simplest form of MMSBM** for clarity; the reported results may therefore be poor in absolute, but underline an improvement over equally simple models (NC and SIMSBM(1)) brought by our Simple Dynamic prior approach.

We systematically perform a 5-folds cross validation. The model is trained on 80% of the data, β is tuned using 10% as a validation set, and the model is evaluated on the 10% left. We choose as metrics the AUC-ROC and the RMSE on the real values of θ (black line in Fig. 1). The procedure is repeated 5 times; the error bars reported in the experimental results represent the standard error over these folds.

4.1.1 SDSBM unveils complex temporal patterns. In Fig. 4a., we consider 1 000 observations for each item $i \in I$ and vary the number of epochs from 10 to 1 000. In the expression of p , s is set to 0.05. For both the sinusoidal and line-broken memberships, the model shows better predictive performances (in terms of AUC ROC) than the proposed baselines. Interestingly, the SDSBM performances remain stable as the number of epochs increases unlike the NC baseline, which means it alleviates a bias of the temporal modeling proposed in [22]. The RMSE with respect to the true parameters remains low over the whole range of tested number of epochs. The RMSE increases as the number of epochs grows because the number of parameters to estimate increases with it; this makes the inference more subject to local variations, which in turns mechanically increases the RMSE. Overall, SDSBM recovers dynamic variations of the membership vectors with a good performance; a sample of the inferred dynamic memberships is shown Fig. 1.

4.1.2 SDSBM works with little data. A major problem that arises when considering temporal data is the scarcity of observations, because slicing implies reducing the number of observations in each slice. This concern largely arises in social sciences, where data retrieval cannot be automated and requires tedious human labor. Here, we demonstrate that our method works in challenging conditions, when data is scarce. In Fig. 4b., we vary the number of observations available for each item from 100 to 10 000, distributed over 100 epochs. Thus, in the most challenging situation, there is only one observation per epoch used to determine I dynamic memberships over 3 clusters. In the expression of p , s is set to 0.05. We see Fig. 4b. that for both patterns, the predictive power of SDSBM remains high in such conditions. Moreover, the RMSE on the true dynamic memberships in this case is fairly low, and decreases rapidly as the number of observations increases. When the number of observations is high, the “no coupling” baseline [22] reaches the performances of SDSBM. This is because as the number of observations in each slice goes to infinity, the models need less to rely on temporal neighbours. However, even for 10 000 observations per item (100 observations per epoch), SDSBM yields better results. As an illustration, the results in Fig. 1 have been obtained using only 5 observations per epoch.

4.1.3 SDSBM handles highly stochastic interaction patterns. Finally, we control the deterministic character of the block-interaction matrix p by varying s . We express such character as the mean entropy of p ($S(p)$) with respect to its possible outputs: $\langle S(p) \rangle = \frac{1}{K} \sum_{k \in K} \sum_o p_k(o) \log p_k(o)$. The maximum entropy for the proposed expression of p is reached for $s = 0.5$. We consider 1 000 observations spread over 100 epochs. We show in Fig. 4c. that the predictive performance of all three methods drops as the entropy increases. This is expected, as observations are generated from the true model with a higher variance; each observation becomes less informative about the generative process as s grows. However, the RMSE on the real parameters inferred using SDSBM remains low at the maximum entropy, meaning the model recovers the correct membership parameters.

4.2 Real-world data

Finally, we demonstrate the validity of our approach on real-world data to argue for its usefulness and scalability. SDSBM builds on previous works on labeled MMSBM and shares the same linear complexity $\mathcal{O}(|R^o|)$ with $|R^o|$ the size of the dataset [9]. For our experiments, we consider the recent and documented datasets from [14]. The **Reddit** dataset (10.000 users, 984 subreddits they contribute to, ~670k observations), the **LastFm** dataset (980 users, 1000 songs users listened to, ~1.3M observations) and the Wikipedia (**Wiki**) dataset (8227 users, 1000 pages users edited, ~157k observations). The goal is to predict over time what subreddit a user will contribute to, what songs they will listen to, and what Wikipedia pages they will edit. The Reddit and Wikipedia datasets contain 1 month of data; we slice them in 1 day long temporal intervals. The LastFm dataset spans over approximately 5 years; we slice it into periods of 3 days each. In addition, we build an additional dataset (**Epi**) about historical epigraphy data². The dataset is made of 117.000 latin inscriptions comprising one or several of 18 social status (slave, soldier, senator, etc.) and its location as one of 62 possible regions, along with an estimated datation spanning from 100BC to 400AD. The goal is to guess the region where a status has been found, with respect to time. The goal is to recover statuses diffusion in roman territories. We slice this dataset in epochs of one year each.

Evaluation is again conducted using a 5-folds cross validation with 80% of training data, 10% of validation data and 10% of testing data for each fold. For each pair (i, o_{true}) in the test set, we query the probability for every output o given i and build the confusion matrix by comparing them to o_{true} . In Table 1, we present the results of our method compared to the proposed baselines for various metrics: AUC-ROC (**ROC**), Average Precision (**AP**) and Normalized Coverage Error (**NCE**). The first two metrics evaluate how well models assign probabilities to observations, and the latter evaluates the order in which possible outputs are ranked.

Overall, we see that our method exhibits a greater predictive power, except for the Reddit dataset where the static SBM performs as well as SDSBM. We explain this by the lack of significant temporal variation over the considered interval. This could be expected, since the dataset comprises roughly 80% of repeated actions [14], meaning that users do not significantly explore new communities over a month. This result shows that SDSBM also works well in

Table 1: Numerical results on real-world datasets — Metrics abbreviations stand for the area under the ROC curve (ROC), the Average Precision (AP), the Normalized Coverage Error (NCE). Metrics for models stand for Simple Dynamic SDM (SDSBM), No Coupling baseline (NC) and the classical static mixed membership SBM (SIMSBM(1)). Overall, our approach allows for a higher predictive power.

		ROC	AP	NCE
Epi	SDSBM	0.9025(11)	0.3700(17)	0.1151(11)
	NC	0.8420(22)	0.3435(36)	0.1582(19)
	SIMSBM(1)	0.8597(12)	0.2141(16)	0.1451(13)
Lastfm	SDSBM	0.8942(8)	0.0168(1)	0.1284(11)
	NC	0.8393(5)	0.0157(2)	0.1785(7)
	SIMSBM(1)	0.8647(5)	0.0115(2)	0.1493(4)
Wiki	SDSBM	0.9759(2)	0.0659(9)	0.0459(3)
	NC	0.9092(7)	0.0608(10)	0.1195(8)
	SIMSBM(1)	0.9576(7)	0.0622(4)	0.0565(8)
Reddit	SDSBM	0.9803(3)	0.4295(54)	0.0312(3)
	NC	0.8508(5)	0.3598(17)	0.1846(7)
	SIMSBM(1)	0.9798(2)	0.4269(40)	0.0322(3)

the static case. On the other datasets, SDSBM performs better often by a large margin, especially for the ROC-AUC, meaning that SDSBM is efficient at distinguishing classes from each other. We also checked that our initial assumption on the smooth temporal variation of the parameters holds. For every real-world dataset, the absolute average change in membership between two consecutive time slices is systematically less than 0.005 ± 0.01 . The absolute variation after 10 time slices is less than 0.03 ± 0.04 . Therefore the inferred parameters do not vary abruptly over time. We recall that the model used here is deliberately simplistic for demonstration; low metrics do not mean the Simple Dynamic prior does not work, but instead that it should be coupled to a more complex model (any of SIMSBM(x,y,...)).

As an illustration of what SDSBM has to offer, we plot in Fig. 5 a possible visualization of the memberships evolution in time for the epigraphy dataset. On the left and on the right, we show the items that are considered in the visualization. The time goes from left (100BC) to right (500AD), and the flows represent the membership transfers between epochs. The grey bars represent the clusters. We manually annotated them by looking at their composition. From this figure, we can recover several historical facts: military presence in Rome was scarce for most of the times considered; Italy concentrates less military presence as time goes (due to its spread over the now extended empire), until the 3rd century crisis that led to its re-militarization; most of the slaves that have been accorded an inscription are located in Italia throughout time; the religious functions (Augusti) are evenly spread on the territory at all times; the libertii (freed slaves) inscriptions are essentially present in Rome and in Italy, etc. Obviously, dedicated works are needed in order to support these illustrative claims, and we believe SDSBM can provide such extended comprehension of these processes.

²Clauss-Slaby repository, <http://www.manfredclauss.de/fr/index.html>

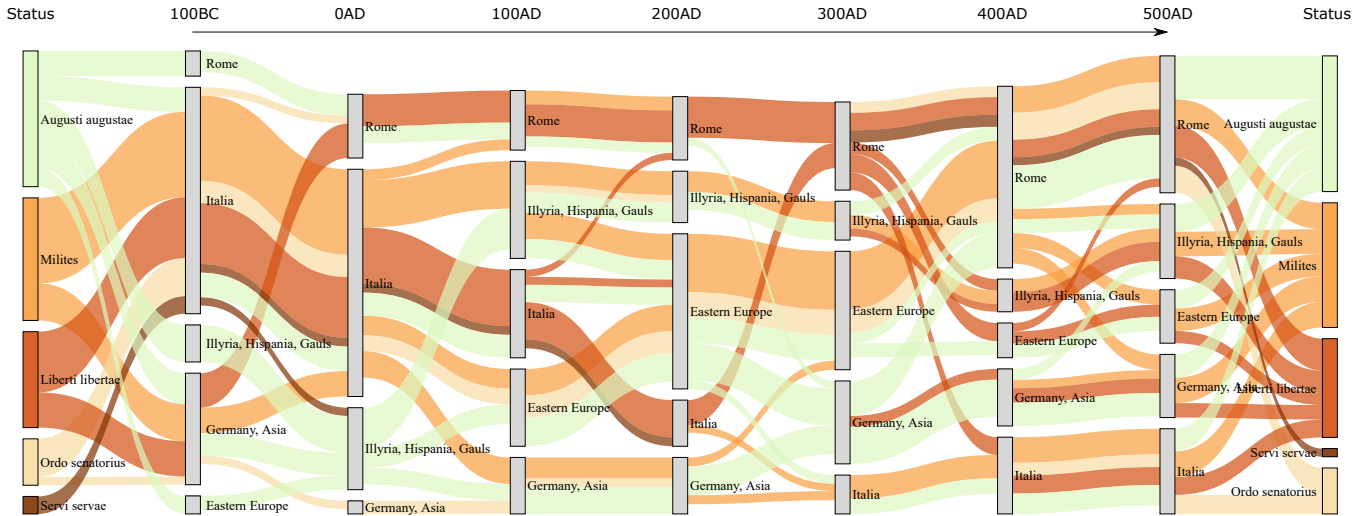


Figure 5: Geographic evolution of status distribution from latin graves (100BC - 500AD) — We applied the SDSBM to the Epigraphy dataset. We recall that our goal is to predict a roman region (e.g. Illyria, Hispania, etc.) given a status (e.g. Slave, Senator, etc.) and a year. We plot the temporal evolution of statuses membership to the five manually labeled clusters (in gray). For clarity, we removed small membership transfers from the data, which explains why the total cluster's population may vary from one time to another. This plot allows to visualize some global historical trends about the evolution of the Roman Empire (e.g. 3rd century crisis, spread of military presence in Europe, Italy demilitarization, etc.).

5 CONCLUSION

We introduced a simple way to model time in dynamic valued and labeled networks by assuming a dynamic Mixed-Membership SBM. Our method consists in defining the Simple Temporal prior, ready to plug into any of a whole class of existing static MMSBMs. Time is considered under the single assumption that network's ties do not vary abruptly.

We assessed the performance of the proposed method by defining the SDSBM and testing it in several controlled situations on synthetic datasets. In particular, we show that our prior allows stable performances with respect to the dataset slicing, and that it works well under challenging conditions (small amounts of data or high entropy blocks interaction matrix). We tested SDSBM on large scale real-world datasets and showed it yields better results than two proposed baselines. Finally, we illustrate an application interest on a dataset of Latin inscriptions that indirectly narrates the evolution of the Roman Empire.

We discussed the advantages of using our approach: uneven slicing of observations in time, heterogeneous dynamic time-scales, time-dependent blocks-interaction matrix, and informativeness of the prior; exploring these directions on real-world data may help retrieving meaningful clusters on useful applications. On a further note, we believe that a key interest of our approach is the small amount of data needed to get satisfactory performances. This point is fundamental to a number of social sciences, and we believe our approach could ease the incorporation of automated learning methods in these fields.

REFERENCES

- [1] Amr Ahmed and Eric P. Xing. 2007. Seeking The Truly Correlated Topic Posterior - on tight approximate inference of logistic-normal admixture model. *Proc. 11th*

- Conf. on Art. Int. and Stat.* 2 (2007), 19–26. <https://proceedings.mlr.press/v2/ahmed07a.html>
- [2] Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg, and Eric P. Xing. 2008. Mixed Membership Stochastic Blockmodels. *J. Mach. Learn. Res.* 9 (jun 2008), 1981–2014. <https://dl.acm.org/doi/10.5555/1390681.1442798>
- [3] Richard Arratia, A. D. Barbour, and Simon Tavaré. 1992. Poisson Process Approximations for the Ewens Sampling Formula. *The Annals of Applied Probability* 2, 3 (1992), 519–535.
- [4] Christopher Bishop. 2006. *Pattern Recognition and Machine Learning*. Springer, 450–455 pages.
- [5] S. Cobo-López, A. Godoy-Lorite, and J. Duch. 2018. Optimal prediction of decisions and model selection in social dilemmas using block models. *EPJ Data Sci* 7(48) (2018). <https://doi.org/10.1140/epjds/s13688-018-0175-3>
- [6] Feng-Lei Fan, Jinjun Xiong, Mengzhou Li, and Ge Wang. 2021. On Interpretability of Artificial Neural Networks: A Survey. *IEEE Trans. on Radiation and Plasma Med. Sci.* 5, 6 (2021), 741–760. <https://doi.org/10.1109/TRPMS.2021.3066428>
- [7] Xuhui Fan, Longbing Cao, and Richard Yi Da Xu. 2015. Dynamic Infinite Mixed-Membership Stochastic Blockmodel. *IEEE Trans. on Neur. Net. and Learn. Sys.* 26, 9 (2015), 2072–2085. <https://doi.org/10.1109/TNNLS.2014.2369374>
- [8] Thorben Funke and Till Becker. 2019. Stochastic block models: A comparison of variants and inference methods. *PloS one* (2019), 1–40. <https://doi.org/10.1371/journal.pone.0215296>
- [9] Antonia Godoy-Lorite, Roger Guimerà, Cristopher Moore, and Marta Sales-Pardo. 2016. Accurate and scalable social recommendation using mixed-membership stochastic block models. *Proc. Natt. Acad. Sci.* 113, 50 (2016), 14207–14212. <https://doi.org/10.1073/pnas.1606316113>
- [10] Roger Guimerà and Marta Sales-Pardo. 2013. A Network Inference Method for Large-Scale Unsupervised Identification of Novel Drug-Drug Interactions. *PLOS Comput. Biol.* 9, 12 (12 2013), 1–9. <https://doi.org/10.1371/journal.pcbi.1003374>
- [11] Qirong Ho, Le Song, and Eric P. Xing. 2011. Evolving Cluster Mixed-Membership Blockmodel for Time-Evolving Networks. *AISTATS* (2011). <https://proceedings.mlr.press/v15/ho11b.html>
- [12] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. 1983. Stochastic blockmodels: First steps. *Social Networks* 5, 2 (1983), 109–137. [https://doi.org/10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7)
- [13] D. Jin, Z. Yu, P. Jiao, S. Pan, D. He, J. Wu, P. Yu, and W. Zhang. 2021. A Survey of Community Detection Approaches: From Statistical Modeling to Deep Learning. *IEEE Trans. on Knowledge and Data Eng.* 01 (2021), 1–1. <https://doi.org/10.1109/TKDE.2021.3104155>
- [14] Srijan Kumar, Xikun Zhang, and Jure Leskovec. 2019. Predicting Dynamic Embedding Trajectory in Temporal Interaction Networks. In *Proc. 25th ACM SIGKDD (Anchorage, AK, USA) (KDD '19)*. Association for Computing Machinery, New

- York, NY, USA, 1269–1278. <https://doi.org/10.1145/3292500.3330895>
- [15] Clement Lee and Darren J. Wilkinson. 2019. A review of stochastic block models and extensions for graph clustering. *Applied Network Science* 4 (2019), 1–50. <https://doi.org/10.1007/s41109-019-0232-2>
- [16] Catherine Matias and Vincent Miele. 2017. Statistical clustering of temporal networks through a dynamic stochastic block model. *J. Roy. Stat. Soc.: Series B (Stat. Meth.)* 79, 4 (2017), 1119–1141. <https://doi.org/10.1111/rssb.12200>
- [17] C Matias, T Rebafka, and F Villers. 2018. A semiparametric extension of the stochastic block model for longitudinal networks. *Biometrika* 105, 3 (06 2018), 665–680. <https://doi.org/10.1093/biomet/asv016>
- [18] Gaël Poux-Médard, Sergio Cobo-Lopez, Jordi Duch, Roger Guimerà, and Marta Sales-Pardo. 2021. Complex decision-making strategies in a stock market experiment explained as the combination of few simple strategies. *EPJ Data Science* 10 (2021), 1–14. Issue 26. <https://doi.org/10.1140/epjds/s13688-021-00280-z>
- [19] Gaël Poux-Médard, Julien Velcin, and Sabine Loudcher. 2021. Information Interactions in Outcome Prediction: Quantification and Interpretation Using Stochastic Block Models. *Proc. 15th ACM Conf. on Rec. Sys. (RecSys)* (2021), 199–208. <https://doi.org/10.1145/3460231.3474254>
- [20] Gaël Poux-Médard, Julien Velcin, and Sabine Loudcher. 2022. Serialized Interacting Mixed Membership Stochastic Block Model. *2022 IEEE International Conference on Data Mining (ICDM)* (2022). <https://doi.org/10.1109/ICDM54844.2022.00145>
- [21] Xuning Tang and Christopher C. Yang. 2014. Detecting Social Media Hidden Communities Using Dynamic Stochastic Blockmodel with Temporal Dirichlet Process. *ACM Trans. Intell. Syst. Technol.* 5, 2 (2014), 21. <https://doi.org/10.1145/2517085>
- [22] Marc Tarrés-Deulofeu, Antonia Godoy-Lorite, Roger Guimerà, and Marta Sales-Pardo. 2019. Tensorial and bipartite block models for link prediction in layered networks and temporal networks. *Phys. Rev. E* 99 (2019), 032307. Issue 3. <https://doi.org/10.1103/PhysRevE.99.032307>
- [23] Quentin Villermet, Jérémie Poiroux, Manuel Moussallam, Thomas Louail, and Camille Roth. 2021. *Follow the Guides: Disentangling Human and Algorithmic Curation in Online Music Consumption*. Association for Computing Machinery, New York, NY, USA, 380–389. <https://doi.org/10.1145/3460231.3474269>
- [24] Yuchung J. Wang and George Y. Wong. 1987. Stochastic Blockmodels for Directed Graphs. *J. Amer. Statist. Assoc.* 82, 397 (1987), 8–19. <https://doi.org/10.1080/01621459.1987.10478385>
- [25] James D. Wilson, Nathaniel T. Stevens, and William H. Woodall. 2019. Modeling and detecting change in temporal networks via the degree corrected stochastic block model. *Qual. Reliab. Eng. Int.* 35 (2019), 1363–1378. <https://doi.org/10.1002/qre.2520>
- [26] Xunxun Wu, Pengfei Jiao, Yaping Wang, Tianpeng Li, Wenjun Wang, and Bo Wang. 2019. Dynamic Stochastic Block Model with Scale-Free Characteristic for Temporal Complex Networks. *Proc. 24th DASFAA* (2019), 502–518. https://doi.org/10.1007/978-3-030-18579-4_30
- [27] Eric P. Xing, Wenjie Fu, and Le Song. 2010. A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics* 4 (2010), 535–566. <https://doi.org/10.1214/09-AOAS311>
- [28] Kevin S. Xu and Alfred O. Hero. 2013. Dynamic Stochastic Blockmodels: Statistical Models for Time-Evolving Networks. *Soc. Comp., Behav.-Cult. Mod. and Pred.* (2013), 201–210. <https://doi.org/10.1109/JSTSP.2014.2310294>
- [29] Kevin S. Xu and Alfred O. Hero. 2014. Dynamic Stochastic Blockmodels for Time-Evolving Social Networks. *IEEE J. of Selected Topics in Sign. Proc.* 8 (2014), 552–562. https://doi.org/10.1007/978-3-642-37210-0_22
- [30] Tianbao Yang, Yun Chi, Shenghuo Zhu, Yihong Gong, and Rong Jin. 2010. Detecting communities and their evolutions in dynamic social networks —a Bayesian approach. *Machine Learning* 82 (2010), 157–189. <https://doi.org/10.1007/s10994-010-5214-7>

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009