The dynamic stochastic topic block model for dynamic networks with textual edges

Marco Corneli, Charles Bouveyron, Pierre Latouche, Fabrice Rossi

The dynamic stochastic topic block model for dynamic networks with textual edges

Marco Corneli[1], Charles Bouveyron[2], Pierre Latouche [1] & Fabrice Rossi [1]

[1] *Laboratoire SAMM, EA 4543, Université Paris 1 Panthéon-Sorbonne.*
E-mail: Marco.Corneli@malix.univ-paris1.fr
[2] *Laboratoire J.A. Dieudonné, UMR CNRS 7351 Equipe Asclepios, INRIA*
Sophia-Antipolis Université Côte d'Azur, Nice, France

Abstract. The present paper develops a probabilistic model to cluster the nodes of a dynamic graph, accounting for the content of textual edges as well as their frequency. Vertices are clustered in groups which are homogeneous both in terms of interaction frequency and discussed topics. The dynamic graph is considered stationary on a latent time interval if the proportions of topics discussed between each pair of node groups do not change in time during that interval. A classification variational expectation-maximization (C-VEM) algorithm is adopted to perform inference. A model selection criterion is also derived to select the number of node groups, time clusters and topics. Experiments on simulated data are carried out to assess the proposed methodology. We finally illustrate an application to the Enron dataset.

Keywords. Dynamic random graph, model based clustering, stochastic block model, non homogeneous Poisson process, topic modeling, latent Dirichlet allocation.

1 Introduction

One of the main goals in network analysis consists in clustering the nodes of a graph into groups of homogeneous interactivity behavior. The clustering techniques can be used to

study various types of data recorded, namely the presence/absence of interactions between nodes, the frequency of such interactions, the number of neighbors of nodes, etc. However, the increasing volume of communication in social networks such as Linkedin, Twitter and Facebook, has being motivating researches on new techniques accounting for both the graph connectivity and the textual contents on the edges. When dealing with time evolving networks, it is of interest to be able to detect deep changes in the graph structure (structural changes) that can affect either the groups composition or the way existing groups interact. As shown in this paper, a joint analysis of both the text contents and the interaction dynamics can provide important insights.

1.1 Statistical approaches for dynamic network analysis

The interactions between nodes are assumed to occur over the time interval $[0, T]$, each interaction being represented by a triplet (i, j, u) if i connects with j at time $u \leq T$. Such datasets are considered in Guigourès et al. (2012, 2015); Corneli et al. (2017) to develop probabilistic models to group the vertices into time invariant groups and to detect change points in the graph structure.

Although this continuous time approach has the advantage of preserving time information (e.g. the exact order in which interactions occur), statistical models in dynamic network analysis are usually in discrete time: a time partition up to time T is considered and interactions are aggregated on the time intervals of such partition to obtain a sequence of static graphs. In the binary case, for example, two nodes are connected if an interaction between them occurs in the corresponding time frame. Notice that, following this approach, a dynamic graph is synonymous of sequence of static graphs. In such a framework, several clustering methods have been proposed, based on the stochastic block model (SBM, Wang and Wong, 1987; Nowicki and Snijders, 2001). This model assumes that the vertices are

clustered in hidden groups and that the probability of interactions between two nodes only depends on the clusters they belong to. Yang et al. (2011) proposed a dynamic extension of SBM, allowing nodes to switch from their cluster at time t to another cluster at time $t+1$, according to a transition probability matrix. Hence, the stochastic process that assigns one node to a group, at each time step, is an homogeneous Markov chain. An alternative approach, based on non-homogeneous Markov chains is proposed in Xu and Hero III (2013). The two approaches described so far are generalized in Matias and Miele (2016). Moreover, in their paper, they also show that restrictions on the connectivity behaviour of groups are needed to ensure parameter identifiability. Two dynamic extensions of SBM, relying on conditional non-homogeneous Poisson processes (NHPPs) were independently developed by Matias et al. (2015) and Corneli et al. (2016a). The former introduced conditionally independent NHPPs to count interactions between all pair of nodes in a dynamic graph. Nodes are clustered in hidden, not time-varying groups and the intensity functions of the NHPPs only depend on the groups of the corresponding pair of nodes. The authors relied on a variational expectation-maximization algorithm (VEM) to cluster vertices and proposed two non parametric techniques to estimate the intensity functions of the NHPPs. In order to avoid over-fitting problems, a further hypothesis is introduced in Corneli et al. (2016a). They assume that the Poisson intensity functions associated with each pair of nodes are piecewise constant on hidden time clusters that are common to the whole graph. In that paper, the inference procedure to cluster both nodes and time intervals relies on a greedy maximization of the exact-ICL (see Biernacki et al., 2000; Côme and Latouche, 2015). It also allows them to select the number of clusters and time clusters.

We finally review some important contributions to cluster analysis (and sometimes change point detection) in dynamic graphs based on probabilistic models alternative to SBM. The dynamic random subgraph model (dRSM, Zreik et al., 2016) extends the RSM model (Jernite et al., 2014) to uncover time varying clusters of nodes within subgraphs

provided a priori. The generalized hierarchical random graph model (GHRG, Peel and Clauset, 2014) decomposes the vertices of a graph into a series of nested groups, whose relationships are represented in a dendrogram where the original nodes are the leaves and the probability of interaction between two nodes is located at their lowest common ancestor. Moreover, the authors developed a statistical test to detect structural changes in the dynamic network based on a sliding window of fixed length and the posterior Bayes factor (Aitkin, 1991). The temporal exponential random graph model (TERGM) of Hanneke et al. (2010) generalizes the exponential random graph model (ERGM) (see Robins et al., 2007, for instance), which is often considered in real applications. In this framework, the evolution of the graph snapshots is modeled through a Markov chain whose transition probabilities depend on some user-defined functions. A similar technique is adopted by Krivitsky and Handcock (2014) who introduced an hypothesis of separability (i.e. conditional independence) between appearing and disappearing connections in two consecutive snapshots of a dynamic graph. This assumption justifies the name STERGM (separable TERGM) and allows the model to gain in ease of specification and tractability. Finally, the popular latent position model (LPM, Hoff et al., 2002) and latent position cluster model (LPCM, Handcock et al., 2007) were also extended by Sarkar and Moore (2005); Friel et al. (2016); Sewell and Chen (2015, 2016) to deal with dynamic, binary or weighted interactions. In a recent work, Durante et al. (2016) allow the node coordinates to evolve in continuous time, via nested Gaussian processes, in order to account for non stationarity in real networks.

1.2 Statistical approaches for the joint analysis of texts and networks

Among probabilistic methods for text analysis, the latent Dirichlet allocation (LDA, Blei et al., 2003) is quite popular. The basic idea of LDA is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. The topic proportions are assumed to follow a Dirichlet distribution. The author-topic (AT, Steyvers et al., 2004; Rosen-Zvi et al., 2004) and the author-recipient-topic (ART McCallum et al., 2005) models partially extend LDA to deal with textual networks. Although providing authorships and information about recipients, these models do not account for the graph structure, e.g. the way vertices are connected. A first attempt to take into account the graph structure, along with the textual content of edges is due to Zhou et al. (2006). The authors propose two community-user topic (CUT) models: CUT1, modeling the communities based on the graph structure only and the CUT2, modeling the communities based on the textual information alone. More recently, Pathak et al. (2008) extended the ART model by introducing the community-author-recipient-topic (CART) model. In this context, authors and recipients are assigned to latent communities and they are clustered by CART based on homogeneity criteria, both in terms of graph structure and textual content. Interestingly, the nodes are allowed to belong to multiple communities and each pair of nodes is associated with a specific topic. Although flexible, the models illustrated so far rely on Gibbs sampling for the inference procedure, which can be prohibitive when dealing with large networks. An alternative model, that can be fitted via variational EM inference, is the topic-link LDA (Liu et al., 2009) performing both community detection and topic modeling. This model employs a logistic transformation based on topic proportions as well as author latent features. A family of 4 topic-user-community models was proposed by Sachan et al. (2012). These models, accounting for multiple com-

munity/topic memberships, discover topic-meaningful communities in graphs with different types of edges. This is of particular interest in social networks like Twitter where different types of interactions exist: follow, tweet, re-tweet, etc.

In order to overcome the limitations of previous methods in terms of scalability and flexibility, Bouveyron et al. (2016) proposed the stochastic topic block model (STBM) along with an inference procedure. This approach can exhibit node partitions that are meaningful both regarding the graph structure and the topics, in directed and undirected graphs. The graph structure analysis relies on SBM, allowing the model to recover a large variety of topological structures (see Latouche et al., 2012, for SBM clustering properties) whereas the textual analysis relies on LDA, allowing the model to characterize the construction of documents. The inference procedure is based on an original classification variational EM algorithm.

1.3 Goals and outline of this paper

In this paper, we aim at analyzing a dynamic graph, i.e. a sequence of static graphs, where interactions between nodes involve text data. The starting point is the STBM model of Bouveyron et al. (2016) and we extend it to the dynamic framework. Data are aggregated over time intervals defined at hand and clusters of time intervals with specific parameters are introduced in the graphical model. An inference procedure is derived allowing to retrieve clusters of nodes with homogeneous connection profiles involving both the interactions patterns and the topics discussed. The procedure also allows us to uncover clusters of time intervals. Finally, a model selection criterion is developed.

The model proposed, called dSTBM, is introduced in Section 2. The model inference and model selection are discussed in Section 3. Section 4 focuses on experiments on simulated and real data to highlight the main features of the proposed approach (model and

inference).

2 The dynamic STBM (dSTBM)

In the first part of this section we detail a generative model for the interactions between nodes of a dynamic graph. Then, in the second part, we describe a generative model for the textual content associated with graph edges. The last part of this section links the proposed methodology to the existing literature.

2.1 Dynamic modeling of edges

A dynamic graph consisting in instantaneous interactions between M nodes, over the time interval $[0, T]$, is considered. Interactions are directed and self loops are not allowed. In a block modeling perspective, nodes are assumed to belong to Q hidden groups $\mathcal{A}_1, \ldots, \mathcal{A}_Q$, whose number has to be estimated (see Section 3). Let Y be an hidden M-vector denoting node memberships ($Y_i = q$ iff node i is in cluster $\mathcal{A}_q$). A multinomial probability distribution is attached to Y

$$p(Y|\rho) = \prod_{q=1}^{Q} \rho_q^{|\mathcal{A}_q|},$$

where $\rho_q := \mathbb{P}\{Y_i = q\}$, $\sum_{q=1}^{Q} \rho_q = 1$ and $|\mathcal{A}_q|$ is the number of nodes in cluster $\mathcal{A}_q$. In the following, the zero-one notation ($Y_{iq} = 1$ if node i is in cluster $\mathcal{A}_q$, zero otherwise) will be used interchangeably, when no confusion arises. Interactions from node i to node j are assumed to be counted by a non homogeneous Poisson process (NHPP) $\{ID_{ij}(t)\}_{t \leq T}$ whose intensity function, $\lambda_{ij}(t)$, positive and integrable on $[0, T]$, only depends on the clusters of the two nodes

$$ID_{ij}(t)|Y_{iq}Y_{jr} = 1 \sim \mathcal{P}\left(\int_0^t \lambda_{qr}(u)du\right),$$

for $t \leq T$. The $M \times (M-1)$ NHPPs, associated with all different pairs (i, j), are assumed to be independent conditionally on Y.

As in Corneli et al. (2016a), we switch to a discrete time framework (see Section 1.1) introducing a partition of the interval $[0, T]$ in U subintervals, $I_u := [t_{u-1}, t_u[$, where

$$0 = t_0 < t_1 < \cdots < t_U = T. \tag{1}$$

The increments of each counting process on the considered time partition can be computed

$$D_{iju} := ID_{ij}(t_u) - ID_{ij}(t_{u-1}), \qquad \forall (i, j, u) \tag{2}$$

and stored in the $M \times M \times U$ tensor $D = \{D_{iju}\}_{i,j,u}$. Hence, we focus on the number of interactions from i to j taking place over the time interval I_u. The time intervals $I_1, \ldots, I_U$ are assigned to L disjoint hidden time clusters $\mathcal{C}_1, \ldots, \mathcal{C}_L$ whose number has to be estimated. Hence, each cluster contains a certain number of time intervals, not necessarily adjacent and an hidden U-vector X is introduced to label memberships to time clusters: $X_u = l$ if and only if I_u belongs to cluster $\mathcal{C}_l$. We stress that the time intervals of the user defined partition (1) are known whereas the time clusters are not observed and have to be estimated. Then, X is assumed to follow a multinomial distribution

$$p(X|\delta) = \prod_{l=1}^{L} \delta_l^{|\mathcal{C}_l|},$$

where $\delta_l := \mathbb{P}\{X_u = l\}$, $\sum_{l=1}^{L} \delta_l = 1$ and $|\mathcal{C}_l|$ denotes the number of time intervals in $\mathcal{C}_l$.

The following assumption is made: the intensity functions are stepwise constant on each time cluster $\mathcal{C}_l$, such that

$$D_{iju}|Y_{iq}Y_{jr}X_{ul} = 1 \sim \mathcal{P}(\Delta_u \lambda_{qrl}),$$

where Δ_u denotes the size of I_u. In the rest of this paper, the grid in (2) is assumed to be regular to simplify the notation. This means that $\Delta_u = \Delta$ and the time intervals $\{I_u\}_u$

have a constant size. It is also possible to consider intervals with different sizes as is Corneli et al. (2015). A $Q \times Q \times L$ tensor $\Lambda = \{\lambda_{qrl}\}_{q,r,l}$ is finally introduced and the complete-data likelihood of the model described is given by

$$p(D, Y, X|\Lambda, \rho, \delta) = p(D|Y, X, \Lambda)p(Y|\rho)p(X|\delta), \tag{3}$$

where the random vectors Y and X are independent and

$$p(D|Y, X, \Lambda) \propto \prod_{q,r}^{Q} \prod_{l}^{L} \left(\Delta\lambda_{qrl}\right)^{S_{qrl}} \exp(-\Delta\lambda_{qrl}P_{qrl}), \tag{4}$$

with

$$\begin{aligned} S_{qrl} &:= \sum_{j \neq i}^{M} \sum_{u=1}^{U} Y_{iq}Y_{jr}X_{ul}D_{iju} \\ P_{qrl} &:= \sum_{j \neq i}^{M} \sum_{u=1}^{U} Y_{iq}Y_{jr}X_{ul}. \end{aligned} \tag{5}$$

Notice that Δ is a time scale factor and can be set equal to one without loss of generality, indeed when $\Delta \neq 1$, we can safety define $\tilde{\lambda}_{qrl} = \Delta\lambda_{qrl}$ and reduce to the previous case.

2.2 Dynamic modeling of documents

The model described in the previous section can easily be extended to deal with textual communication networks, by assuming that a directed interaction characterizing the pair (i, j) corresponds to a document sent from i to j. With the previous notations, D_{iju} is the number of documents sent from i to j over the time interval I_u and more generally $ID_{ij}(t)$ is the number of documents sent from i to j up to time t. The documents counted by D_{iju} are considered as a unique document obtained by concatenation and N_{iju} denotes the number of words of such document. In the following, a dictionary containing V words

will be considered and each word in a document is extracted from the dictionary: W_n^{iju} will denote the n-th word (in the aggregated document) sent from i to j during the time interval I_u and, using a zero-one notation, $W_{nv}^{iju} = 1$ if the word W_n^{iju} is the v-th in the dictionary, 0 otherwise.

In line with the LDA model (Blei et al., 2003), a list of K *topics* is introduced and each word of a document is associated with one topic through a latent N_{iju}-vector, noted Z^{iju}. In more details, $Z_n^{iju} = k$ if and only if the word W_n^{iju} is associated with the k-th topic. For each pair of clusters $(\mathcal{A}_q, \mathcal{A}_r)$ and a time cluster $\mathcal{C}_l$, a vector of topic proportions $\theta_{qrl} := (\theta_{qrlk})_{k \leq K}$ is assumed to follow a Dirichlet distribution

$$\theta_{qrl} \sim \mathrm{Dir}(\alpha = (\alpha_1, \ldots, \alpha_K)),$$

such that $\sum_{k=1}^{K} \theta_{qrlk} = 1$. Hence, the n-th word in the document associated with the triplet (i, j, I_u), namely W_n^{iju}, is extracted from the latent topic k according to the following conditional probability distribution

$$\mathbf{P}(Z_{nk}^{iju} = 1 | D, Y, X, \theta) = \prod_{q,r}^{Q} \prod_{l}^{L} \theta_{qrlk}^{Y_{iq} Y_{jr} X_{ul}}$$

corresponding to a multinomial distribution of parameter θ_{qrl}. The following full conditional distribution is obtained

$$p(Z | D, Y, X, \theta) = \prod_{q,r}^{Q} \prod_{l=1}^{L} \prod_{k=1}^{K} \theta_{qrlk}^{\sum_{j \neq i}^{M} \sum_{u=1}^{U} \sum_{n=1}^{N_{iju}} Y_{iq} Y_{jr} X_{ul} Z_{nk}^{iju}}, \tag{6}$$

where the exponent counts the total occurrences, in the dynamic graph, of words associated with the k-th topic, sent from cluster $\mathcal{A}_q$ to cluster $\mathcal{A}_r$, during the time cluster $\mathcal{C}_l$ and $Z := (Z^{iju})_{i,j,u}$. Given Z, the word W_n^{iju} is finally assumed to be drawn from a multinomial distribution

$$W_n^{iju} | Z_{nk}^{iju} = 1 \sim \mathcal{M}(1, \beta_k = (\beta_{k1}, \ldots, \beta_{kV})).$$

Hence, β denotes a $K \times V$ matrix whose k-th line is β_k. Notice that, unlike the topic proportions θ, the matrix β depends neither on node clusters nor on time clusters. In particular, this means that the mean number of occurrences of each word in each topic is time invariant. Denoting by $W = (W^{iju})_{i,j,u}$ the whole set of documents appearing in the dynamic network, the following conditional distribution is obtained by independence

$$p(W|Z, D, \beta) = \prod_{k=1}^{K} \prod_{v=1}^{V} \beta_{kv}^{\sum_{j \neq i}^{M} \sum_{u=1}^{U} \sum_{n=1}^{N_{iju}} W_{nv}^{iju} Z_{nk}^{iju}}, \tag{7}$$

where the exponent counts the total occurrences, in the dynamic graph, of the v-th word of the dictionary associated with the k-th topic.

The complete-data conditional distribution for the textual part of the model is finally obtained by conditioning

$$p(W, Z, \theta|D, Y, X, \beta) = p(W|Z, D, \beta)p(Z|D, Y, X, \theta)p(\theta)$$

and the joint distribution of the whole dSTBM model is

$$p(D, Y, X, W, Z, \theta|\Lambda, \rho, \delta, \beta) = p(W, Z, \theta|D, Y, X, \beta)p(D, Y, X|\Lambda, \rho, \delta).$$

A graphical representation of the dynamic STBM can be seen in Figure 1.

2.3 Link with existing models

First of all, let us clarify the relation between dSTBM and LDA. Assuming that Y and X are known, the set of documents W can be reorganized such that $W = (\tilde{W}_{qrl})_{qrl}$ where

$$\tilde{W}_{qrl} = \{W^{iju} | Y_{iq} Y_{jr} X_{ul} = 1\}$$

is the set of all documents sent from any vertex in $\mathcal{A}_q$ to any vertex in $\mathcal{A}_r$, during the time cluster $\mathcal{C}_l$. By marginalization over Z, it can easily be seen that each word W_n^{iju} has a

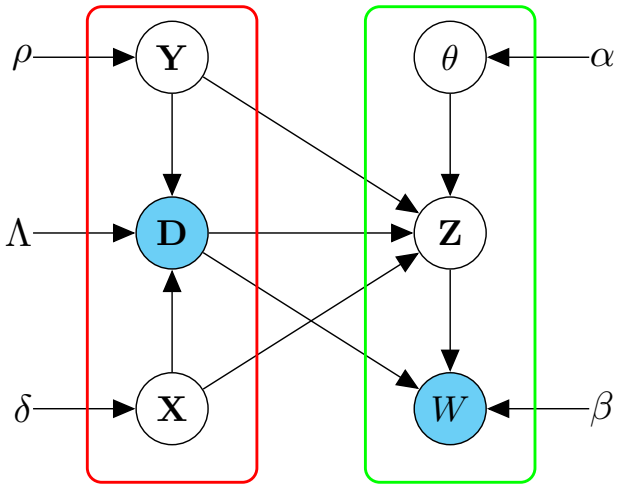


Figure 1: Graphical representation of the dynamic STBM model (dSTBM).

mixture distribution over topics which only depends on the clusters of i and j and the time cluster of I_u. As a consequence, all words in $\tilde{W}_{qrl}$ share the same mixture distribution over topics and removing the knowledge of (q, r, l), $\tilde{W}_{qrl}$ can be seen as one of $Q^2 \times L$ independent documents. This means that, if the pair (X, Y) is known, the generative model described so far is the one of a LDA model with $Q^2 \times L$ independent documents. Each documents has its own vector of topic proportions and shares a matrix β of word probabilities.

More generally we can highlight the following relations between dSTBM and some of the existing models mentioned so far.

1. **Single time cluster** $(L = 1)$**.** In this case both Λ and θ are constant in time and dSTBM reduces to STBM (Bouveyron et al., 2016).

2. **Single topic** $(K = 1)$**.** When a single topic is used in the whole network, there is no additional information that can be extrapolated relying on text analysis. In this

case, dSTBM reduces to the dSBM model (Corneli et al., 2016a).

3. **Single cluster** ($Q = 1$)**.** When all vertices are clustered in a single group, the set of documents can be reorganized as $W = (\tilde{W}_l)_{l \leq L}$ corresponding to L documents. Each one corresponds to a time cluster and has its own topic proportions $(\theta_l)_{l \leq L}$. This could be seen as an original dynamic extension of the LDA model (Blei et al., 2003) in which the topic proportions evolve in time. From a generative point of view, we stress that only L i.i.d. topic proportion vectors, $\theta_1, \ldots, \theta_L$, are generated. With respect to the original time partition (1), *all* documents sent in time intervals belonging to the same time cluster share the same (previously) extracted topic proportion parameter. Notice that the dynamic approach described so far is completely different from the one adopted by Blei and Lafferty (2006). In that paper, sequentially organized corpus of documents are taken into account and both the Dirichlet parameter (α) and the topic parameter (β) change in time according to (unit-root) autoregressive models combined with multinomial-logit probabilities. Hence, from a generative point of view, at each time step t, a *new* vector of topic proportions is drawn based on α_t.

4. **Case** $Q = L = 1$**.** In line with the previous case, the set W can now be considered as a single document with its own topic proportions. The dSTBM model reduces in this case to the LDA model.

5. **Case** $K = L = 1$**.** In presence of a single topic discussed in the whole network (i.e. text analysis is useless), with Λ constant in time, the dSTBM model reduces to SBM with weighted Poisson distributed links (see e.g. Nouedoui and Latouche, 2013).

3 Estimation

This section focuses on the inference procedure adopted to learn the model parameters and provide estimates for X, Y and Z. In the last part of the section, a model selection criterion is developed to select Q, L and K.

3.1 Variational inference

Let us assume for now that the number of clusters (Q), time clusters (L) and the number of topics (K) are known.

Consider the following complete-data integrated log-likelihood

$$\log p(D, Y, X, W|\Lambda, \rho, \delta, \beta) = \log \sum_{Z} \int_{\theta} p(D, Y, X, W, Z, \theta|\Lambda, \rho, \delta, \beta) d\theta. \tag{8}$$

We aim at maximizing it with respect to the model parameters $(\Lambda, \rho, \delta, \beta)$ and the hidden label vectors (Y, X). Unfortunately, (8) is not tractable due to the sum over all possible values of Z inside the logarithm. Nonetheless, a variational decomposition of the above log-likelihood can be employed to obtain a lower bound which can be directly maximized. Adopted in Blei et al. (2003) and Bouveyron et al. (2016), this approach gives

$$\begin{aligned} \log p(D, Y, X, W|\zeta) = \mathcal{L}(R(\cdot); D, Y, X, W, \zeta) \\ + \mathrm{KL}(R(\cdot)||p(\cdot|D, Y, X, W, \zeta)) \end{aligned} \tag{9}$$

where $\zeta := \{\Lambda, \rho, \delta, \beta\}$, $R(\cdot)$ is a variational distribution over the pair (Z, θ),

$$\mathcal{L}(R(\cdot); D, Y, X, W, \zeta) := \mathbf{E}_{R(Z,\theta)} \left[\log \frac{p(D, Y, X, W, Z, \theta|\zeta)}{R(Z, \theta)} \right] \tag{10}$$

and $\mathrm{KL}(\cdot)$ denotes the Kullback-Leibler divergence between the approximate and the true posterior distribution of the pair (Z, θ) given the data and the model parameters

$$\mathrm{KL}(R(\cdot)||p(\cdot|D, Y, X, W, \zeta)) := -\mathbf{E}_{R(Z,\theta)} \left[\log \frac{p(Z, \theta|D, Y, X, W, \zeta)}{R(Z, \theta)} \right].$$

Notice that, since the left hand side of (9) does not depend on $R(\cdot)$, when maximizing the lower bound $\mathcal{L}$ with respect to $R(\cdot)$, the KL divergence is necessarily minimized. When performing variational inference, a common choice to approximate the true posterior distribution of latent variables (e.g. Daudin et al., 2008), consists in assuming that $R(\cdot)$ factorizes over the latent variables. In this case, this leads to

$$R(Z, \theta) = R(Z)R(\theta) = R(\theta) \prod_{j \neq i}^{M} \prod_{u=1}^{U} \prod_{n=1}^{N_{iju}} R(Z_n^{iju}).$$

Hence, since the integrated likelihood in (8) cannot be directly maximized, the idea is to replace it with the lower bound $\mathcal{L}$ and maximize it with respect to the model parameters (Λ, π, δ, β), the approximate posterior distribution $R(Z, \theta)$ in the above equation and the hidden vectors Y and X. Furthermore, as it can be seen in the graphical model in Figure 1, the full joint distribution of the dSTBM model can be decomposed into two parts. The component represented by the red rectangle does *not* depend on the pair (Z, θ). As a consequence, the lower bound defined in (10), can be split into two parts also

$$\mathcal{L}(R(\cdot); D, Y, X, W, \zeta) = \tilde{\mathcal{L}}(R(\cdot); D, Y, X, W, \beta) + \log p(D, Y, X|\Lambda, \rho, \delta), \tag{11}$$

where

$$\tilde{\mathcal{L}}(R(\cdot); D, Y, X, W, \beta) := \mathbf{E}_{R(Z,\theta)} \left[\log \frac{p(W, Z, \theta|D, Y, X, \beta)}{R(Z, \theta)} \right]. \tag{12}$$

Note that the joint distribution $p(D, Y, X|\Lambda, \rho, \delta)$ appeared for the first time in (3) and corresponds to the dynamic SBM part of the model. Furthermore, given Y and X, the first term on the right hand side of (11) only involves the pair $(R(\cdot), \beta)$ while the second term only involves (Λ, ρ, δ). Hence, the maximization algorithm that is detailed in the next section consists in alternating the following two steps, up to convergence

1. **VEM step.** For a given pair (Y, X), the lower bound $\mathcal{L}$ is maximized with respect to the pair $(R(\cdot), \beta)$, involving $\tilde{\mathcal{L}}$ and the triplet (Λ, ρ, δ) involving the dSBM complete-data likelihood.

2. **Classification step.** The lower bound $\mathcal{L}$ is maximized in a greedy fashion with respect to the pair (Y, X).

This algorithm, alternating a variational EM routine with a clustering step, was first used in Bouveyron et al. (2016) and is built upon the Classification-EM (CEM) algorithm (Celeux and Govaert, 1991).

3.2 Maximization of the lower bound

In this section, the updating formulas for $R(Z, \theta)$ and the model parameters $(\Lambda, \rho, \delta, \beta)$ are provided by the following propositions. At the end of the section, we discuss the maximization with respect to the pair (Y, X).

Maximization of $\mathcal{L}$ with respect to $R(Z, \theta)$. The updating formulas corresponding to the E step of the VEM algorithm are given in the following two propositions.

Proposition 1. *The VEM update step for distribution $R(Z_n^{iju})$ is given by*

$$R(Z_n^{iju}) = \mathcal{M}(Z_n^{iju}; 1, \phi_n^{iju} = (\phi_{n1}^{iju}, \ldots, \phi_{nK}^{iju}))$$

where

$$\phi_{nk}^{iju} \propto \left(\prod_{v=1}^{V} \beta_{kv}^{W_{nv}^{iju}} \right) \prod_{q,r=1}^{Q} \prod_{l=1}^{L} \exp \left(\psi(\gamma_{qrlk}) - \psi(\sum_{k'=1}^{K} \gamma_{qrlk'}) \right)^{Y_{iq} Y_{jr} X_{ul}}, \qquad \forall (n, k)$$

where ϕ_{nk}^{iju} is the approximate posterior probability of word W_n^{iju} being in topic k and $\psi(\cdot)$ denotes the digamma function.

Proof. In Appendix A.1. □

Proposition 2. *The VEM update step for distribution $R(\theta)$ is given by*

$$R(\theta) = \prod_{q,r=1}^{Q} \prod_{l=1}^{L} \mathrm{Dir}(\theta_{qrl}; \gamma_{qrl} = (\gamma_{qrl1}, \ldots, \gamma_{qrlK}))$$

where

$$\gamma_{qrlk} = \alpha_k + \sum_{j \neq i}^{M} \sum_{u=1}^{U} \sum_{n=1}^{N_{iju}} Y_{iq} Y_{jr} X_{ul} \phi_{nk}^{iju}, \qquad \forall (q, r, l).$$

Proof. In Appendix A.2 □

Maximization of $\mathcal{L}$ with respect to the model parameters. The following proposition provides the estimates of the model parameters $(\beta, \Lambda, \rho, \delta)$ obtained through maximizing the lower bound in (10). The lower bound $\tilde{\mathcal{L}}$ in (12) is computed in the appendix.

Proposition 3. *The estimates of (β, Λ, ρ) and δ are given by*

$$\beta_{kv} \propto \sum_{j \neq i}^{M} \sum_{u=1}^{U} \sum_{n=1}^{N_{iju}} W_{nv}^{iju} \phi_{nk}^{iju}, \qquad \forall (k, v) \tag{13}$$

$$\lambda_{qrl} = \frac{S_{qrl}}{P_{qrl}}, \qquad \forall (q, r, l) \tag{14}$$

$$\rho_q \propto |\mathcal{A}_q|, \qquad \forall q, \tag{15}$$

$$\delta_l \propto |\mathcal{C}_l|, \qquad \forall l, \tag{16}$$

where S_{qrl} and P_{qrl} were defined in (5).

Proof. In Appendix A.4. □

Maximization of $\mathcal{L}$ with respect to the label vectors Other parameters being fixed, we now attempt to maximize $\mathcal{L}$ with respect to the pair (Y, X). Since this combinatorial problem cannot be attacked directly, due to the huge number of cluster assignments to

test ($Q^M L^U$), a *greedy* search strategy is employed to look for a local maximum. Greedy search methods are quite popular in the network analysis literature. They are employed for community detection problems (Newman and Girvan, 2004; Blondel et al., 2008) or more general clustering purposes, either in static (Côme and Latouche, 2015) or dynamic (Corneli et al., 2016b) graphs.

Consider Y at first and assume that nodes are clustered in Q initial groups (see Section 3.3 for more details about initialization). If node i is currently in cluster $\mathcal{A}_q$, the algorithm assesses the increase/decrease in the lower bound $\mathcal{L}$ due to switching node i to the cluster $\mathcal{A}_r$ for each $r \neq q$. The switch (if any) leading to the highest increase of the lower bound is actually performed and the entire routine is iteratively applied to *all* nodes until no further increase of $\mathcal{L}$ is possible. The maximization with respect to X is performed similarly: nodes are replaced by time sub-intervals I_u and node clusters $\mathcal{A}_q$ by time clusters $\mathcal{C}_l$.

As previously explained, a greedy search is never guaranteed to converge to a global maximum. Hence a good strategy consists in performing several independent greedy maximizations, randomizing over the node/time intervals moving order and finally choosing the values of (Y, X) leading to the highest lower bound.

3.3 Further issues

Initialization. Assuming that Q, L and K are known, the C-VEM algorithm still needs some initial values of (Y, X), in order to provide estimates for the model parameters and the variational posterior distribution $R(Z, \theta)$. Since the EM-like algorithms are only guaranteed to converge to local optima (see e.g. Wu, 1983) it is crucial to provide them with several initializations. The approach proposed in this paper relies on a spectral clustering algorithm (von Luxburg, 2007) applied to proper similarity matrices. The initialization of

Y is considered at first. Recalling the definition of $D = \{D_{iju}\}_{iju}$, we proceed as follows

1. The VEM algorithm for the LDA model (Blei et al., 2003) is applied to the collection of documents exchanged from all pair of nodes in the whole time horizon. Note that these documents correspond to the entries of D and the VEM algorithm provides the majority topic discussed in each document. Hence an $M \times M \times U$ tensor MT (main topic) is obtained, such that $MT_{iju} = k$ if and only if k is the main topic discussed in the document sent form i to j, during the time interval I_u.

2. An $M \times M$ similarity matrix Ξ is obtained as follows

$$\Xi(i,j) = \sum_{u=1}^{U}\sum_{h=1}^{M} \delta(MT_{ihu} = MT_{jhu}) D_{ihu} D_{jhu} + \sum_{u=1}^{U}\sum_{h=1}^{M} \delta(MT_{hiu} = MT_{hju}) D_{hiu} D_{hju}.$$

   The rationale behind the above equation is quite intuitive: if i and j have a common neighbour *and* they talk with it about the same (main) topic, then the similarity between i and j increases. Two terms appear on the right hand side of the equality because we are dealing with directed graphs.

3. The spectral clustering algorithm is applied to the graph Laplacian associated with matrix Ξ. This allows us to cluster nodes in Q groups and to produce an initial estimate of Y.

The initialization of X is performed similarly. A $U \times U$ similarity matrix Σ is built such that two time intervals are similar if they share the same majority topic discussed in the whole network

$$\Sigma(u,v) = \sum_{i=1}^{M}\sum_{j=1}^{M} \delta(MT_{iju} = MT_{ijv}) D_{iju} D_{ijv}$$

for all pairs of time intervals (I_u, I_v). The spectral clustering algorithm if finally applied to the graph Laplacian associated with the similarity matrix Σ to produce an initial estimate of X.

Model selection. So far, the parameters Q, L and K were assumed to be known but in real world datasets this assumption is fairly unrealistic. In order to estimate these parameters, we rely on the ICL criterion (Biernacki et al., 2000) to approximate the complete-data integrated log-likelihood in (8). This approach extends the model selection criterion proposed in Bouveyron et al. (2016) to the dynamic framework of the present paper.

Proposition 4. *An integrated classification criterion (ICL) for the dSTBM is*

$$\begin{aligned} ICL_{dSTBM} = \tilde{\mathcal{L}}(R(\cdot); D, Y, X, W, \beta) - \frac{K(V-1)}{2}\log(LQ^2) + \max_{\Lambda,\rho,\delta} \log p(D, Y, X|\Lambda, \rho, \delta) \\ - \frac{LQ^2}{2}\log MU(M-1) - \frac{Q-1}{2}\log M - \frac{L-1}{2}\log U. \end{aligned} \tag{17}$$

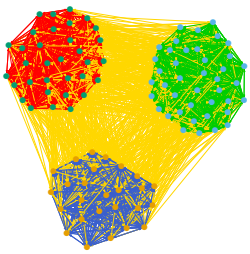
Proof. In Appendix A.5. □

4 Numerical Experiments

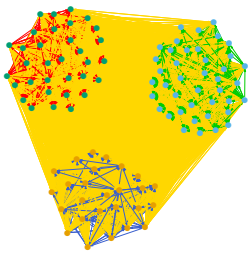
In this section, both dSTBM and the ICL criterion introduced above are tested on simulated data. In order to highlight some peculiarities, dSTBM is tested in three different scenarios and compared with four other models: dSBM (Corneli et al., 2016a), STBM (Bouveyron et al., 2016), SBM using the mixer R package `https://cran.r-project.org/web/packages/mixer/index.html` and LDA using the topicmodels R package `https://cran.r-project.org/web/packages/topicmodels/index.html`

4.1 Simulation setups

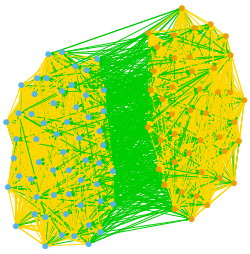
In the following simulation setups, the parameter α_k is assumed to be equal to 1, inducing a uniform distribution over the topic proportions θ_{qrl}. In each setup, 50 dynamic graphs are independently simulated and the messages associated with graph edges are sampled from four texts from BBC news. One text is about the birth of Princess Charlotte, the second



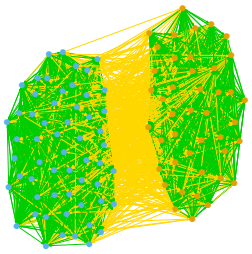
(a) **A**. First time cluster ($\mathcal{C}_1$).



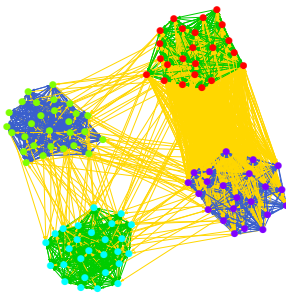
(b) **A**. Second time cluster ($\mathcal{C}_2$).



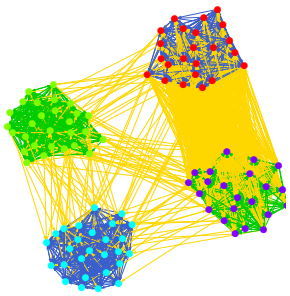
(c) **B**. First time cluster ($\mathcal{C}_1$).



(d) **B**. Second time cluster ($\mathcal{C}_2$).



(e) **C**. First time cluster ($\mathcal{C}_1$).



(f) **C**. Second time cluster ($\mathcal{C}_2$).

Figure 2: Dynamic graphs simulated according to three different setups (A,B and C). The graph on the left (respectively right) hand side of each row is obtained through aggregation of the interactions on the first (second) time cluster.

is about black holes in astrophysics, the third one focuses on UK politics and the fourth on cancer diseases. Each message, associated with one directed interaction, is made of 75 words. We finally stress that, the message sampling procedure adopted in the following scenarios is *not* exactly the one described in the previous sections for dSTBM. Each setup is detailed in the following.

Scenario A. Figures 2a and 2b. Nodes are grouped in three clusters and time intervals in two time clusters. During the first time cluster, the graph exhibits a clear community structure: interactions *within* groups are more frequent than interactions *between* groups. An opposite non-assortative structure characterizes the graph during the second time cluster: interactions between groups are more frequent than interactions within groups. Each group talks about a single topic and a fourth shared topic is associated with the interactions between two different groups. In order to introduce some noise, 10% of interactions within each group are (randomly) associated to the shared topic. In this first scenario the topic proportions do not change in time.

Scenario B. In this second scenario, the dynamic graph maintains a persistent community structure, whereas a structural time change occurs in the topic proportions. Nodes are grouped into two clusters and time intervals into two time clusters. Two topics are taken into account, corresponding to two of the four texts from the BBC news. During the first time cluster, each community talks preferentially about the same topic (in yellow, say T_1) and a second topic T_2 (green) is reserved to the interactions between communities (Figure 2c). During the second time cluster, the two topics have the opposite role. Hence T_2 is used for the intra-community interactions whereas T_1 is discussed between members of different groups (Figure 2d). As in the previous setup, 10% of interactions inside each group is (randomly) associated with the shared topic to introduce some noise.

Scenario C. This third scenario consists in a dynamic graph whose nodes are grouped into four clusters. However, only two of these clusters are real communities, with actors talking preferentially about a unique topic inside the community. The other two clusters form a single community and the topic they discuss about is the only discriminant. Hence, three topics are considered: two clusters use one topic (green), the other two clusters use another topic (blue) and a third topic is used for communications between all different groups (yellow). In order to induce a relevant time structure, the topics used within groups change from a time cluster to another as illustrated in Figures 2e and 2f.

A detailed description of each scenario can be seen in Table 1.

4.2 Benchmark results

The C-VEM algorithm for dSTBM was run on 50 simulated dynamic graphs in each scenario. First, we focus on the clustering produced by the methodology when the numbers of clusters Q, time clusters L and topics K are known. The adjusted rand index (ARI, Rand, 1971) provides a measure of the accuracy of the realised clustering: it ranges from 0, corresponding to a very poor clustering, to 1, when the found partitions are the actual ones. The clustering results for dSTBM, dSBM, and STBM can be seen in Table 2. The clustering measure "edge ARI" is equal to one when the main topic used in each exchanged document is correctly retrieved by the model. We recall that one document is uniquely associated with a triplet (i, j, I_u) in the dynamic graph: source node, destination node and time interval. Hence, the number of exchanged documents coincides with the total degree of the simulated dynamic graph. It follows that the edge ARI defined so far is not available for both dSBM and STBM: the former does not deal with topics, the latter cannot recover information about the interactions taking place at time I_u since this information

Scenario	A	B	C
M	100		
U	100		
Q	3	2	4
L	2		
K	4	2	3
ρ	$(1/Q, \ldots, 1/Q)$		
δ	$(1/L, \ldots, 1/L)$		
Λ on $\mathcal{C}_1$	$\begin{cases} \lambda_{qq1} = 0.03 \\ \lambda_{qr1} = 0.0075 \quad r \neq q \end{cases}$	$\begin{cases} \lambda_{qq1} = 0.03 \\ \lambda_{qr1} = 0.0075 \quad r \neq q \end{cases}$	$\begin{cases} \lambda_{qq1} = \lambda_{141} = \lambda_{411} = 0.03 \\ \lambda_{qr1} = 0.0075 \quad \text{otherwise} \end{cases}$
Λ on $\mathcal{C}_2$	$\begin{cases} \lambda_{qq2} = 0.0075 \\ \lambda_{qr2} = 0.03 \quad r \neq q \end{cases}$	$\begin{cases} \lambda_{qq2} = 0.03 \\ \lambda_{qr2} = 0.0075 \quad r \neq q \end{cases}$	$\begin{cases} \lambda_{qq2} = \lambda_{142} = \lambda_{412} = 0.03 \\ \lambda_{qr2} = 0.0075 \quad \text{otherwise} \end{cases}$
θ on $\mathcal{C}_1$	$\begin{cases} \theta_{1111} = \theta_{2212} = \theta_{3313} = 1 \\ \theta_{qr14} = 1 \quad r \neq q \\ \text{otherwise} \quad 0 \end{cases}$	$\begin{cases} \theta_{1112} = \theta_{2212} = 1 \\ \theta_{qr11} = 1 \quad r \neq q \\ \text{otherwise} \quad 0 \end{cases}$	$\begin{cases} \theta_{1112} = \theta_{3312} = 1 \\ \theta_{2211} = \theta_{4411} = 1 \\ \theta_{qr13} = 1 \quad r \neq q \\ \text{otherwise} \quad 0 \end{cases}$
θ on $\mathcal{C}_2$	$\begin{cases} \theta_{1121} = \theta_{2222} = \theta_{3323} = 1 \\ \theta_{qr24} = 1 \quad r \neq q \\ \text{otherwise} \quad 0 \end{cases}$	$\begin{cases} \theta_{1121} = \theta_{2221} = 1 \\ \theta_{qr22} = 1 \quad r \neq q \\ \text{otherwise} \quad 0 \end{cases}$	$\begin{cases} \theta_{1121} = \theta_{3321} = 1 \\ \theta_{2222} = \theta_{4422} = 1 \\ \theta_{qr23} = 1 \quad r \neq q \\ \text{otherwise} \quad 0 \end{cases}$

Table 1: Parametrization in different setups.

is definitely lost, due to aggregation. However, STBM can cluster the edges of the aggregated graph. Namely, it estimates the main topic used by each pair of nodes during the whole time horizon. Hence, the edge ARI corresponding to STBM can be calculated by assigning to *all* the edges in the dynamic graph associated with the pair (i, j) the main

	Setup A		
Model	node ARI	time ARI	edge ARI
dSTBM	0.99 (0.06)	1 (0)	0.99 (0.06)
dSBM	1 (0)	1 (0)	-
STBM	1 (0)	-	0.66 (0.21)
SBM	0.01 (0.06)	-	-
LDA	-	-	0.73 (0.20)

	Setup B		
Model	node ARI	time ARI	edge ARI
dSTBM	1 (0)	1 (0)	1 (0)
dSBM	0.98 (0.03)	0.00 (0.01)	-
STBM	0.5 (0.5)	-	0.02 (0.03)
SBM	0.99 (0.04)	-	-
LDA	-	-	1 (0)

	Setup C		
Model	node ARI	time ARI	edge ARI
dSTBM	1 (0)	1(0)	1 (0)
dSBM	0.67 (0.05)	0.00 (0.01)	-
STBM	1 (0)	-	0.70 (0.10)
SBM	0.65 (0.04)	-	-
LDA	-	-	0.69 (0.15)

Table 2: Clustering results for dSTBM, dSBM, STBM, SBM and LDA on 50 graphs simulated according to the different setups. The true values of Q, L and K are assumed to be known. The average ARI values are reported, with standard deviations into brackets.

topic estimated for that pair by STBM (in the aggregated graph).

Let us discuss the clustering results of the first setup **A**. Not surprisingly, dSTBM and

dSBM have very similar performances and dSBM is slightly more accurate in clustering nodes (ARI equal to 1 versus ARI equal to 0.99). This small difference however is not very significant and can be explained by the different initializations adopted by the two approaches. As mentioned above, in this scenario the proportion of assigned topics (θ) is constant in time, hence the structural change in the dynamic graphs can be fully detected by dSBM and the analysis of documents does not bring any further information. This is the reason why the time ARI is equal to one for both the approaches: the time structure can be recovered with or without the analysis of documents. Since STBM cannot deal with dynamic graphs, the C-VEM algorithm for this model is run on the static graph obtained by aggregating the interactions on the whole time horizon (September, 2001 - January, 2002). Despite of the structural change (Figures 2a and 2b), the topics used for communications within each community and between communities remain distinct on the whole time horizon. This is the reason why STBM can correctly cluster nodes. Similarly to STBM, the SBM model is run on the aggregated graph. Its performance is poor since the community structure in $\mathcal{C}_1$ and the non-assortative structure in $\mathcal{C}_2$ cancel each other out when aggregating interactions over time. Looking at the edge ARI, when aggregating interactions over time information is lost: this explains the edge ARI of 0.66 for STBM. The edge ARI is slightly better for LDA which is applied to the whole collection of documents (there is no aggregation).

Consider now the second setup **B**. Since the topic proportions are the only time varying parameter, dSBM cannot see any time cluster (null time ARI). Nonetheless, the persistent community structure allows it to recover the actual node partition most of the time (node ARI of 0.98). A similar result can be seen for SBM. Conversely, since each topic is alternatively used for intra and inter community interactions (Figures 2c and 2d), STBM suffers in recovering the actual node partition (node ARI of 0.5). As explained before, the LDA model can be applied to the original set of documents and in this case, not particularly

Scenario C, ICL (dSTBM, $K = 3$)						
Q/L	1	2	3	4	5	6
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	**48**	1	0	0	0
5	0	1	0	0	0	0
6	0	0	0	0	0	0

Table 3: Frequency of selections by ICL for dSTBM (Q, L, K) on 50 simulated graphs in the third scenario **C**. The actual values of (Q, L, K) are $(4, 2, 3)$, respectively. The true value for K is always selected by ICL and it is not reported.

noised, it performs very well.

The last scenario **C** is the hardest for dSBM. As in the previous case, the topic proportions are the only time varying parameter and the time clusters are not correctly detected by the model (null time ARI). Moreover, two clusters form a single community (Figures 2e and 2f) and they are only discriminated by the used topic. Hence the node ARI is never higher than 0.7 for dSBM (and SBM too). Instead, in contrast with the previous scenario, the inter-community topic (yellow) is never employed for intra-community interactions and STBM can recover the actual node partition. Notice, however, that both STBM and LDA are performing worse than dSTBM in clustering the edges.

4.3 Model Selection

So far, the C-VEM algorithm for dSTBM was run on 50 simulated dynamic graphs for each setup and the actual number of groups Q, time clusters L and topics K were assumed to be known. In real applications, these three parameters must be estimated and this can be done for dSTBM relying on the ICL model selection criterion developed in Proposition 4. In terms of model selection, the third scenario **C** is by far the hardest to deal with, due to the quite sophisticated dynamic graph structure. Hence, we focus on this setup to assess the ICL criterion. The estimates of Q, L and K, provided by ICL for dSTBM, are illustrated in Table 3. The actual number of topics ($K = 3$) is always detected by ICL and it is therefore not reported in the table. Tables with $K \neq 3$ would be full of zeros. As it can be seen, the actual values of Q and L are recovered in 48 out of 50 cases. Notice also that, when ICL fails to recover the actual solution, it selects a model very close to the actual one.

5 Analysis of the Enron scandal

This last section focuses on the famous scandal involving the energy company Enron Corporation. The scandal was publicized in October 2001. Two moths later, USA experienced the largest bankruptcy failure up to that time. The first part of this section describes the Enron data set we used, while the second part illustrates the results obtained through applying the dSTBM model to the dataset.

5.1 Context and data

The Enron communication network is a popular data set containing all e-mail exchanges between 149 employees of the company. The original dataset is available at `http://www.cs.cmu.edu/~./enron/` and cover the time horizon 1999-2002. The time window considered

in the present section spans from September, 3rd, 2001 to January, 28th, 2002, including the following three key dates

1. September, 11th, 2001: the terrorist attacks to the Twin Towers and the Pentagon (USA).

2. October, 31st, 2001: the Securities and Exchange Commission (SEC) opened an investigation for fraud concerning Enron.

3. December, 2nd, 2001: Enron failed for bankruptcy, resulting in more than 4,000 lost jobs.

The selected time window is partitioned in weekly subintervals, thus corresponding to $U = 21$ weeks. As previously explained, the e-mails sent from i to j during each time interval I_u (a week) are aggregated into a single document, obtained by concatenation. Each document is preprocessed in a classical way: words are stemmed, less than three characters words and stop words are removed, punctuation and numbers are ignored. Thus, each week is associated with a graph snapshot and one directed edge from i to j corresponds to the e-mails sent from i to j during the week. The whole dynamic graph is made of 4321 directed edges, corresponding to the same number of exchanged documents. The dictionary associated to these documents contains 49,955 words.

5.2 Results

The VEM algorithm for dSTBM was run on this dataset for all values of Q, K and L varying between 1 and 10. For each value of (Q, K, L) several initializations were tested (see Section 3.3 for further details) and the clustering results associated with the highest value of the ICL criterion were retained. The ICL finally selected nine topics ($K = 9$), four time clusters ($L = 4$) and six node groups ($Q = 6$).

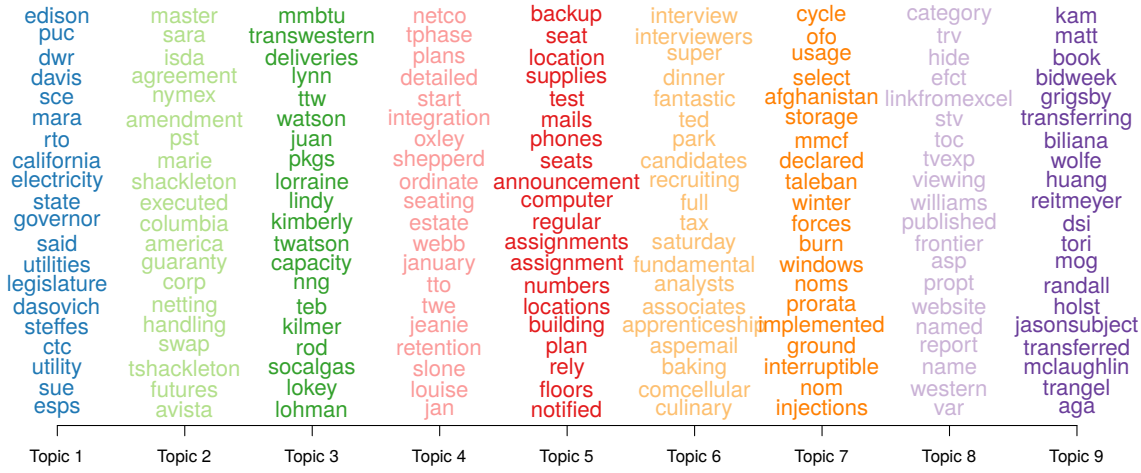


Figure 3: The 20 most representative words for each topic.

Topics. First of all, we discuss is a few details some topics that play a crucial role in the dynamic network, as detailed in the following. Figure 3 shows the most representative words of each topic and can be used in the attempt to understand the main theme of each topic.

a. Topic 1 is related to the California electricity crisis, in which Enron was involved and which almost caused the bankruptcy of the SCE-corp (Southern California Edison Corporation).

b. Topic 3 is a technical topic focusing on gas deliveries (mmBTU are British thermal units).

c. Topic 4 seems to be related to Netco: a set of trading activities bought by the Swiss bank UBS after the Enron bankruptcy.

d. Topic 5 is related to a backup plan developed to face possible work stoppages. In fact, some areas of the Enron Center North building were put aside for recovery purposes and backup seats assignments were announced to employees in November 2001.

e. Topic 7 contains words like "afghanistan" and "taleban" and it is concerned with Enron activities in Afghanistan: Enron and the Bush administration were suspected to work secretly with Talebans before the 9/11 attacks.

f. Topic 8 seems to focus on TRV (trader report viewer), a project allowing traders to share their reports about particular issues. For example, an e-mail dating November, 13, 2001 announced to several employees that a report on West NG (west Virginia natural gas) prices was available. A "link from Excel" was provided in the e-mail.

g. Topic 9 seems to be related to the company trading activities, as the words "book", "transferring" and "bid week" suggest. The bid week, in particular, is the last week of the month when producers try to sell their core production and consumers seek to buy for their core natural gas needs for the upcoming month.

Time structure. In Figure 4, an histogram reports the frequency of exchanged e-mails in the whole network, each rectangle covers one week. Rectangles/weeks of the same color are assigned to the same time cluster by dSTBM. Notice that, although time intervals in the same cluster do *not* have to be adjacent in dSTBM, the clustering reported in Figure 4 clearly detects four segments of adjacent time intervals and three corresponding change points, one for each color switch. It is worth to notice that the last two change points occur some days after the two key dates mentioned at the beginning of the present section and they are represented in the figure by two vertical lines, blue and red, respectively.

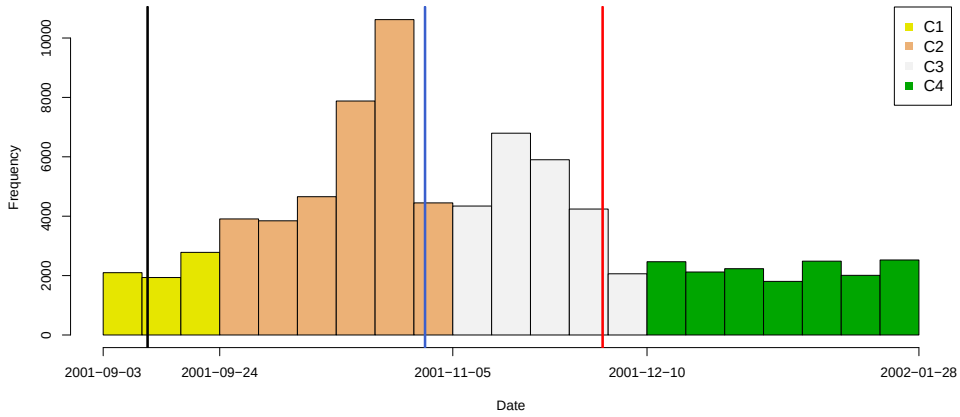


Figure 4: Time clustering results with dSTBM on the Enron data set (Sept. 2001 - Jan. 2002). The black vertical line marks the day September, 11, 2001, the blue vertical line marks the day October, 31st, 2001 (investigation opened by the SEC), the red vertical line marks the day December, 2nd, 2001 (Enron's bankruptcy).

Nodes clustering. The main clustering results are summarized in Figure 5. Four graphs are associated with the time clusters detected by the model. Each node in a graph corresponds to a cluster of nodes and node sizes are proportional to group membership probabilities ρ. The edge colors indicate the most discussed topics in the corresponding (group) interactions (see also Figure 3). The larger the arrow is, the more frequent the respective interactions are. Some remarks can be made by looking at this figure.

1. Consider Group 4 (pink), consisting of 32 agents (mainly vice presidents, CEOs and managers). The topic used by this group for internal communications changes on each time segment: topic 9 in time clusters 1 and 4, topic 7 in time clusters 2, topic 8 in time cluster 3.

2. It is interesting to observe that Topic 7 appears as a main topic in the network during

the time cluster $\mathcal{C}_2$, starting on September, 24th, 2001, exactly two weeks after the 9/11 attacks.

3. Topic 5 is only used for communications *between* clusters during the time cluster $\mathcal{C}_2$. Topic 5 (as well as Topic 7) is no longer a main topic during the other time clusters.

4. Group 6 (yellow), 18 persons, has a similar composition of Group 4. It is concerned with Topic 1 during the first three time clusters and switches to Topic 4 after the company bankruptcy, during the fourth segment.

5. Group 5 (red), 17 employees, looks like a real persistent community both in terms of interactivity pattern and used topic. This group focuses during the whole time horizon on Topic 3.

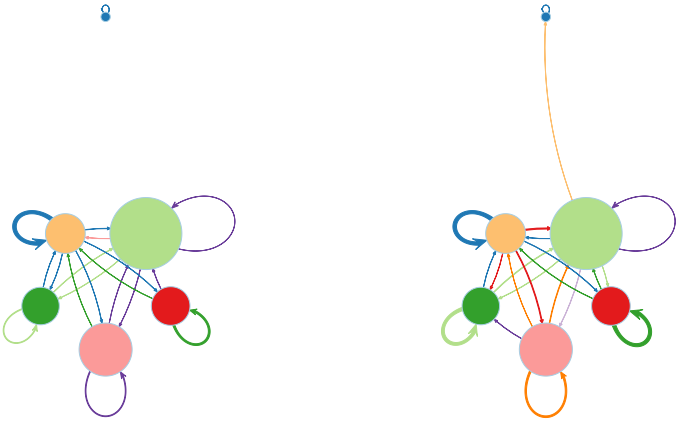
Finally, Figure 6 shows four graph snapshots associated with the Enron dataset. Each snapshot is obtained by aggregating the interactions over the corresponding time cluster. Nodes of the same color are assigned to the same cluster by the C-VEM algorithm and edges of the same color are associated with the same majority topic on the considered time cluster.

6 Conclusion

We proposed in this paper the dynamic stochastic topic block model (dSTBM), a new probabilistic model for the clustering of both nodes and edges of a textual dynamic network. Moreover, relying on an external time partition, our methodology allows one to uncover time clusters during which the network is stationary both in terms interaction frequency (between groups of nodes) and discussed topics. The inference procedure relies on a classification VEM approach and an ICL model selection criterion is derived in order to estimate

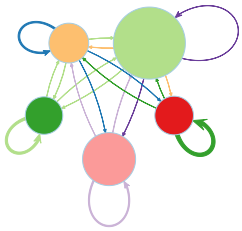
the number of node groups, time clusters and discussed topics. Numerical experiments on simulated data allowed us to highlight the main features of the proposed methodology, which proves to generalize several existing approaches. Finally, the application of dSTBM to the Enron communication network leaded to likely results.

Future researches could focus on a "clever" way to set a time partition, either including this partition between the model parameters or adopting a data driven choice (as done by Matias et al., 2015, for a dynamic SBM-like model). Alternatively, the dSTBM model could be extended to deal with overlapping clusters, allowing individuals to belong to multiple groups. In this context, a starting point could be the mixed memberships SBM (MMSBM, Airoldi et al., 2008).

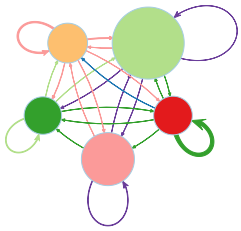


(a) Time cluster $\mathcal{C}_1$.

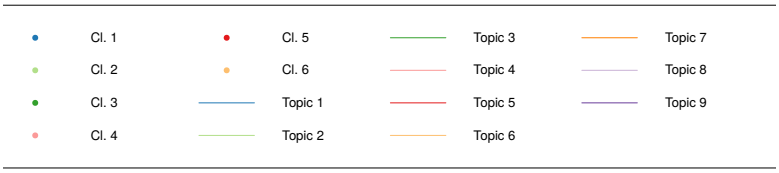
(b) Time cluster $\mathcal{C}_2$.



(c) Time cluster $\mathcal{C}_3$.

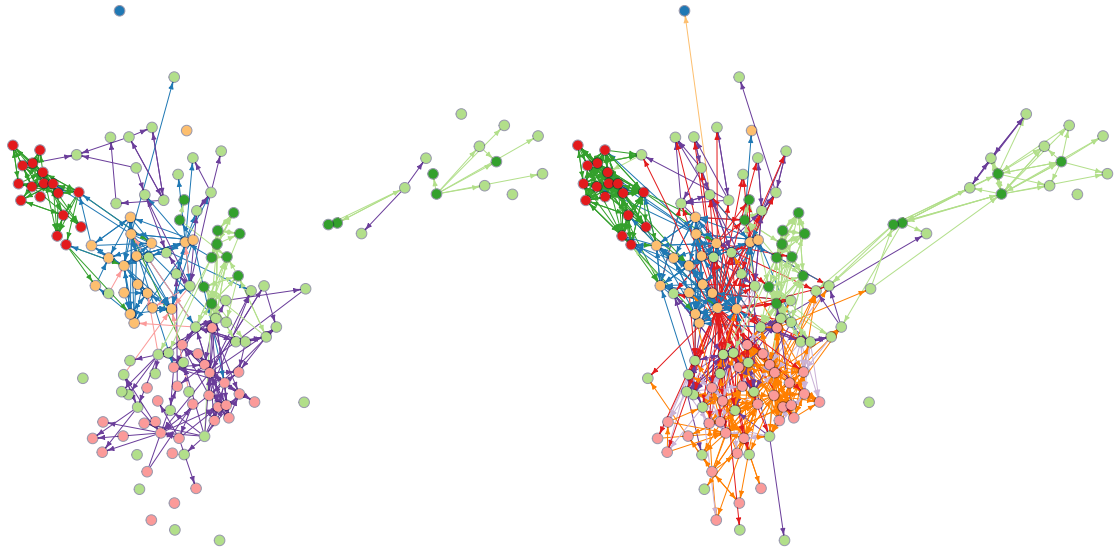


(d) Time cluster $\mathcal{C}_4$.



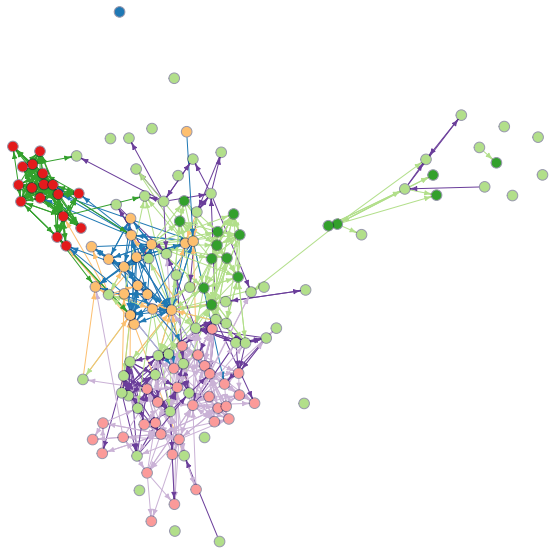
(e) Legend.

Figure 5: Summary of the interaction intensities (Λ, edge widths), group proportions (ρ, node size) and main topic for group interactions (edge colors) during each time cluster.

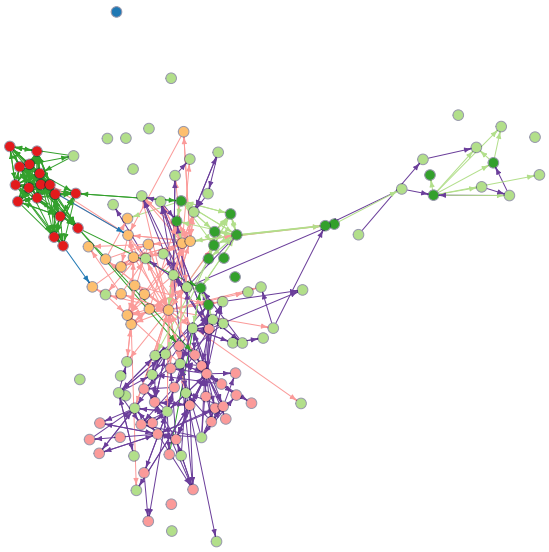


(a) Time cluster $\mathcal{C}_1$.

(b) Time cluster $\mathcal{C}_2$.



(c) Time cluster $\mathcal{C}_3$.



(d) Time cluster $\mathcal{C}_4$.

Figure 6: Clustering results with dSTBM on the Enron data set (Sept. 2001 - Jan. 2002). Each graph corresponds to a time cluster.

A Proofs

A.1 Proof of Proposition 1

Proof. The VEM update step for the distribution $R(Z_n^{iju})$, for all i, j, u and n, is given by

$$\begin{aligned}
\log R(Z_n^{iju}) &= \mathbf{E}_{R(Z^{\setminus i,j,u,n},\theta)}[\log p(W|Z,D,\beta) + \log p(Z|D,Y,X,\theta)] + C \\
&= \sum_{k=1}^{K} Z_{nk}^{iju} \sum_{v=1}^{V} W_{nv}^{iju} \log \beta_{kv} + \sum_{q,r}^{Q} \sum_{l=1}^{L} Y_{iq} Y_{jr} X_{ul} \sum_{k=1}^{K} Z_{nk}^{iju} \mathbf{E}_{\theta_{qrl}}[\log \theta_{qrl}] + C \\
&= \sum_{k=1}^{K} Z_{nk}^{iju} \left(\sum_{v=1}^{V} W_{nv}^{iju} \log \beta_{kv} + \sum_{q,r}^{Q} \sum_{l=1}^{L} Y_{iq} Y_{jr} X_{ul} \left(\psi(\gamma_{qrlk}) - \psi(\sum_{k=1}^{K} \gamma_{qrlk}) \right) \right) + C,
\end{aligned} \tag{18}$$

where the expectation is taken with respect to the distribution $R(Z, \theta)$ conditional on Z_n^{iju} to be fixed, C includes all the terms not depending on Z_n^{iju} and $\psi(\cdot)$ denotes the digamma function. The functional form of a multinomial distribution can be recognised

$$R(Z_n^{iju}) = \mathcal{M}\left(Z_n^{iju}; 1, \phi_n^{iju} = \{\phi_{n1}^{iju}, \ldots, \phi_{nK}^{iju}\}\right),$$

where

$$\phi_{nk}^{iju} \propto \left(\prod_{v=1}^{V} \beta_{kv}^{W_{nv}^{iju}} \right) \prod_{q,r}^{Q} \prod_{l=1}^{L} \exp\left(\psi(\gamma_{qrlk}) - \psi(\sum_{k=1}^{K} \gamma_{qrlk}) \right)^{Y_{iq} Y_{jr} X_{ul}}.$$

□

A.2 Proof of Proposition 2

Proof. The VEM update step for distribution the distribution $R(\theta)$ is given by

$$\begin{aligned}
\log R(\theta) &= \mathbf{E}_{R(Z)}[\log p(Z|D,Y,X,\theta)] + C \\
&= \sum_{j \neq i}^{M} \sum_{u=1}^{U} \sum_{n=1}^{N_{iju}} \sum_{q,r}^{Q} \sum_{l=1}^{L} Y_{iq} Y_{jr} X_{ul} \sum_{k=1}^{K} \mathbf{E}_{R(Z)}[Z_{nk}^{iju}] \log \theta_{qrlk} + \sum_{q,r}^{Q} \sum_{l=1}^{L} \sum_{k=1}^{K} (\alpha_k - 1) \log \theta_{qrlk} + C \\
&= \sum_{q,r}^{Q} \sum_{l=1}^{L} \sum_{k=1}^{K} \left(\alpha_k + \sum_{j \neq i}^{M} \sum_{u=1}^{U} \sum_{n=1}^{N_{iju}} Y_{iq} Y_{jr} X_{ul} \phi_{nk}^{iju} - 1 \right) \log \theta_{qrlk} + C,
\end{aligned} \tag{19}$$

where C contains the terms not depending on θ. The functional form of a Dirichlet distribution can be recognized

$$R(\theta) = \prod_{q,r}^{Q} \prod_{l=1}^{L} \mathrm{Dir}(\theta_{qrl}; \gamma_{qrl} = \{\gamma_{qrl1}, \ldots, \gamma_{qrlK}\}),$$

with

$$\gamma_{qrlk} = \alpha_k + \sum_{j \neq i}^{M} \sum_{u=1}^{U} \sum_{n=1}^{N_{iju}} Y_{iq} Y_{jr} X_{ul} \phi_{nk}^{iju}.$$

□

A.3 Derivation of the lower bound

The functional $\tilde{\mathcal{L}}(R(\cdot); D, Y, X, W, \beta)$ in (12) given in Proposition 2 and Proposition 3, is given by

$$\begin{aligned}
\tilde{\mathcal{L}}(R(\cdot);D, Y, X, W, \beta) = & \sum_{j \neq i}^{M} \sum_{u=1}^{U} \sum_{n=1}^{N_{iju}} \sum_{k=1}^{K} \sum_{v=1}^{V} W_{nv}^{iju} \phi_{nk}^{iju} \log(\beta_{kv}) \\
& + \sum_{j \neq i}^{M} \sum_{u=1}^{U} \sum_{n=1}^{N_{iju}} \sum_{k=1}^{K} \phi_{nk}^{iju} \left(\sum_{q,r}^{Q} \sum_{l}^{L} Y_{iq} Y_{jr} X_{ul} \left(\psi(\gamma_{qrlk}) - \psi(\sum_{k=1}^{K} \gamma_{qrlk}) \right) \right) \\
& + \sum_{q,r}^{Q} \sum_{l}^{L} \left(\log \Gamma(\sum_{k=1}^{K} \alpha_k) - \sum_{k=1}^{K} \log \Gamma(\alpha_k) + \sum_{k=1}^{K} (\alpha_k - 1) \left(\psi(\gamma_{qrlk}) - \psi(\sum_{k=1}^{K} \gamma_{qrlk}) \right) \right) \\
& - \sum_{j \neq i}^{M} \sum_{u=1}^{U} \sum_{n=1}^{N_{iju}} \sum_{k=1}^{K} \phi_{nk}^{iju} \log(\phi_{nk}^{iju}) \\
& - \sum_{q,r}^{Q} \sum_{l}^{L} \left(\log \Gamma(\sum_{k=1}^{K} \gamma_{qrlk}) - \sum_{k=1}^{K} \log \Gamma(\gamma_{qrlk}) + \sum_{k=1}^{K} (\gamma_{qrlk} - 1) \left(\psi(\gamma_{qrlk}) - \psi(\sum_{k=1}^{K} \gamma_{qrlk}) \right) \right).
\end{aligned}$$

A.4 Proof of Proposition 3

Proof. The maximization of the functional in (12) with respect to β is considered at first. By isolating the terms depending on β and introducing K Lagrange multipliers accounting for the constraints $\sum_{v=1}^{V} \beta_{kv} = 1$, $\forall k$, we obtain the following objective function

$$f(\beta) := \sum_{j \neq i}^{M} \sum_{u=1}^{U} \sum_{n=1}^{N_{iju}} \sum_{k=1}^{K} \sum_{v=1}^{V} \phi_{nk}^{iju} \log \beta_{kv} + \sum_{k=1}^{K} \lambda_k \left(\sum_{k=1}^{K} \beta_{kv} - 1 \right),$$

whose gradient can be easily computed and set equal to zero to find the β_{kv} in (13).

In a similar fashion, when optimizing with respect to ρ, the following objective function is introduced

$$f(\rho) := \sum_{i=1}^{M}\sum_{q=1}^{Q} Y_{iq}\log\rho_q + \lambda\left(\sum_{q=1}^{Q}\rho_q - 1\right), \tag{20}$$

and its first derivative with respect to ρ_q is set equal to zero to obtain the stationary point in equation (15). The optimization with respect to δ is analogous and (14) is a consequence of the likelihood in (4). □

A.5 Proof of Proposition 4

Proof. A factorizing prior distribution being attached to the model parameters, $(\Lambda, \rho, \delta, \beta)$, the integrated complete-data log-likelihood $\log p(W, D, Y, X|Q, L, K)$ can easily be written as

$$\begin{aligned}\log p(W, D, Y, X|Q, L, K) = &\log\int_{\beta} p(W|D, Y, X, \beta, Q, L, K)p(\beta|K)d\beta\\ &+ \log\int_{\Lambda} p(D|Y, X, \Lambda, Q, L)p(\Lambda|Q, L)d\Lambda\\ &+ \log\int_{\rho} p(Y|\rho, Q)p(\rho|Q)d\rho\\ &+ \log\int_{\delta} p(X|\delta, L)p(\delta|L)d\delta,\end{aligned} \tag{21}$$

where the dependency on (Q, L, K) is made explicit and the pair (Z, θ) is integrated out as in Section 3.1. Following the derivation of the ICL criterion (Biernacki et al., 2000) we rely on a BIC-like approximation of the second term on the right hand side of the above equation to obtain

$$\log\int_{\Lambda} p(D|Y, X, \Lambda, Q, L)p(\Lambda|Q, L)d\Lambda \approx \max_{\Lambda}\log p(D|Y, X, \Lambda, Q, L) - \frac{Q^2 L}{2}\log(MU(M-1)).$$

Similarly the last two terms can be approximated as

$$\log\int_{\rho} p(Y|\rho, Q)p(\rho|Q)d\rho \approx \max_{\rho}\log p(Y|\rho, Q) - \frac{Q-1}{2}\log(M)$$

and

$$\log\int_{\delta} p(X|\delta, L)p(\delta|L)d\delta \approx \max_{\delta}\log p(X|\delta, L) - \frac{L-1}{2}\log(U).$$

Notice that the last three approximations lead to the ICL criterion for the dSBM model

$$\begin{aligned}ICL_{dSBM} := &\max_{\Lambda}\log p(D|Y, X, \Lambda, Q, L) - \frac{Q^2 L}{2}\log(MU(M-1))\\ &+ \max_{\rho}\log p(Y|\rho, Q) - \frac{Q-1}{2}\log(M)\\ &+ \max_{\delta}\log p(X|\delta, L) - \frac{L-1}{2}\log(U).\end{aligned}$$

The exact version of this criterion is maximized relying on a greedy search approach in Corneli et al. (2016b).

Consider now the first term on the right hand side of (A.5). Recalling that the documents W can be organized as $W = (\tilde{W}_{qrl})_{q,r,l}$ such that all words in $\tilde{W}_{qrl}$ follow the same mixture distribution over topics, we adopt the BIC-like approximation obtained in Bouveyron et al. (2016) corrected by the number of documents in dSTBM

$$\log \int_{\beta} p(W|D, Y, X, \beta, Q, L, K) p(\beta|K) d\beta \approx \max_{\beta} \log p(W|D, Y, X, \beta, Q, L, K) - \frac{K(V-1)}{2} \log(Q^2 L).$$

Since the first term on the right hand side of the above approximation is not tractable, it is replaced by its variational approximation $\tilde{\mathcal{L}}(R(\cdot); D, Y, X, W, \beta)$, defined in (12), and the proposition is proven. □

References

Airoldi, E., Blei, D., Fienberg, S., and Xing, E. (2008). Mixed membership stochastic blockmodels. *The Journal of Machine Learning Research*, 9:1981–2014.

Aitkin, M. (1991). Posterior bayes factors (disc: p128-142). *Journal of the Royal Statistical Society, Series B: Methodological*, 53:111–128.

Biernacki, C., Celeux, G., and Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(7):719–725.

Blei, D. M. and Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pages 113–120. ACM.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.

Blondel, V. D., loup Guillaume, J., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks.

Bouveyron, C., Latouche, P., and Zreik, R. (2016). The stochastic topic block model for the clustering of vertices in networks with textual edges. *Statistics and Computing*.

Celeux, G. and Govaert, G. (1991). A classification EM algorithm for clustering and two stochastic versions. Research Report RR-1364, INRIA. Projet CLOREC.

Côme, E. and Latouche, P. (2015). Model selection and clustering in stochastic block models based on the exact integrated complete data likelihood. *Statistical Modelling*, 15(6):564–589.

Corneli, M., Latouche, P., and Rossi, F. (2015). Modelling time evolving interactions in networks through a non stationary extension of stochastic block models. In Pei, J., Silvestri, F., and Tang, J., editors, *International Conference on Advances in Social Networks Analysis and Mining ASONAM 2015*, pages 1590–1591, Paris, France. IEEE/ACM, ACM.

Corneli, M., Latouche, P., and Rossi, F. (2016a). Block modelling in dynamic networks with non-homogeneous poisson processes and exact ICL. *Social Network Analysis and Mining*, 6(1):1–14.

Corneli, M., Latouche, P., and Rossi, F. (2016b). Exact ICL maximization in a non-stationary temporal extension of the stochastic block model for dynamic networks. *Neurocomputing*, 192:81 – 91.

Corneli, M., Latouche, P., and Rossi, F. (2017). Multiple change points detection and clustering in dynamic networks. *Statistics and Computing*, In press.

Daudin, J.-J., Picard, F., and Robin, S. (2008). A mixture model for random graphs. *Statistics and Computing*, 18(2):173–183.

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

Friel, N., Rastelli, R., Wyse, J., and Raftery, A. E. (2016). Interlocking directorates in irish companies using a latent space model for bipartite networks. *Proceedings of the National Academy of Sciences*, 113(24):6629–6634.

Grün, B. and Hornik, K. (2011). topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13):1–30.

Guigourès, R., Boullé, M., and Rossi, F. (2012). A triclustering approach for time evolving graphs. In *Co-clustering and Applications, IEEE 12th International Conference on Data Mining Workshops (ICDMW 2012)*, pages 115–122, Brussels, Belgium.

Guigourès, R., Boullé, M., and Rossi, F. (2015). Discovering patterns in time-varying graphs: a triclustering approach. *Advances in Data Analysis and Classification*, pages 1–28.

Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354.

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

Hoff, P., Raftery, A., and Handcock, M. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.

Jernite, Y., Latouche, P., Bouveyron, C., Rivera, P., Jegou, L., and Lamassé, S. (2014). The random subgraph model for the analysis of an acclesiastical network in merovingian gaul. *Annals of Applied Statistics*, 8(1):55–74.

Krivitsky, P. N. and Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):29–46.

Latouche, P., Birmelé, E., and Ambroise, C. (2012). Variational bayesian inference and complexity control for stochastic block models. *Statistical Modelling*, 12(1):93–115.

Liu, Y., Niculescu-Mizil, A., and Gryc, W. (2009). Topic-link lda: Joint models of topic and author community. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 665–672, New York, NY, USA. ACM.

Matias, C. and Miele, V. (2016). Statistical clustering of temporal networks through a dynamic stochastic block model. *The Journal of the Royal Statistical Society: Series B, to appear*.

Matias, C., Rebafka, T., and Villers, F. (2015). Estimation and clustering in a semiparametric Poisson process stochastic block model for longitudinal networks. *ArXiv e-prints*.

McCallum, A., Corrada-Emmanuel, A., and Wang, X. (2005). The author-recipient-topic model for topic and role discovery in social networks. In *Workshop on Link Analysis, Counterterrorism and Security*.

Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113.

Nouedoui, L. and Latouche, P. (2013). Bayesian non parametric inference of discrete valued networks. In *21-th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013)*, pages 291–296, Bruges, Belgium.

Nowicki, K. and Snijders, T. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087.

Pathak, N., DeLong, C., Banerjee, A., and Erickson, K. (2008). Social topic models for community extraction.

Peel, L. and Clauset, A. (2014). Detecting change points in the large-scale structure of evolving networks. *CoRR*, abs/1403.0989.

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

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social networks*, 29(2):173–191.

Rosen-Zvi, M., Griffiths, T., Steyvers, M., and Smyth, P. (2004). The author-topic model for authors and documents. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, UAI '04, pages 487–494, Arlington, Virginia, United States. AUAI Press.

Sachan, M., Contractor, D., Faruquie, T. A., and Subramaniam, L. V. (2012). Using content and interactions for discovering communities in social networks. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 331–340, New York, NY, USA. ACM.

Sarkar, P. and Moore, A. W. (2005). Dynamic social network analysis using latent space models. *ACM SIGKDD Explorations Newsletter*, 7(2):31–40.

Sewell, D. K. and Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512):1646–1657.

Sewell, D. K. and Chen, Y. (2016). Latent space models for dynamic networks with weighted edges. *Social Networks*, 44:105–116.

Steyvers, M., Smyth, P., Rosen-Zvi, M., and Griffiths, T. (2004). Probabilistic author-topic models for information discovery. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 306–315, New York, NY, USA. ACM.

von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416.

Wang, Y. and Wong, G. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82:8–19.

Wu, C. F. J. (1983). On the convergence properties of the em algorithm. *Ann. Statist.*, 11(1):95–103.

Xu, K. S. and Hero III, A. O. (2013). Dynamic stochastic blockmodels: Statistical models for time-evolving networks. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 201–210. Springer.

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

Zhou, D., Manavoglu, E., Li, J., Giles, C. L., and Zha, H. (2006). Probabilistic models for discovering e-communities. In *Proceedings of the 15th International Conference on World Wide Web*, WWW '06, pages 173–182, New York, NY, USA. ACM.

Zreik, R., Latouche, P., and Bouveyron, C. (2016). The dynamic random subgraph model for the clustering of evolving networks. *Computational Statistics*, pages 1–33.