Mixed Membership Blockmodels for Dynamic Networks with Feedback

Yoon-Sik Cho  
*Information Sciences Institute, University of Southern California, Marina del Rey, CA 90292, USA*

Greg Ver Steeg  
*Information Sciences Institute, University of Southern California, Marina del Rey, CA 90292, USA*

Aram Galstyan  
*Information Sciences Institute, University of Southern California, Marina del Rey, CA 90292, USA*

CONTENTS

24.1 Introduction ................................................................. 528  
24.1.1 Selection and Influence in Networks ............................ 528  
24.1.2 Mixed Membership Stochastic Blockmodels .................. 529  
24.1.3 Related Work .......................................................... 530  
24.2 Co-evolving Mixed Membership Blockmodel .................... 531  
24.2.1 Inference .............................................................. 532  
24.2.2 Variational E-step ................................................... 533  
24.2.3 Variational M-step ................................................... 534  
24.3 Results .................................................................................................. 535  
24.3.1 Experiments on Synthetic Data ....................................... 535  
24.3.2 Comparison with dMMSB ................................................. 536  
24.3.3 U. S. Senate Co-Sponsorship Network ......................... 536  
24.3.4 Interpreting Results ....................................................... 537  
24.3.5 Polarization Dynamics ................................................... 538  
24.4 Discussion ......................................................................................... 539  
Appendix ................................................................................................... 541  
References .............................................................................................. 543

Real-world networks are inherently complex dynamical systems, where both node attributes and network topology change in time. These changes often affect each other, providing complex feedback mechanisms between node and link dynamics. Here we propose a dynamic mixed membership model of networks that explicitly take into account such feedback. In the proposed model, the probability of observing a link between two nodes depends on their current group membership vectors, while those membership vectors themselves evolve in the presence of a link between the nodes. Thus, the network is shaped by the interaction of stochastic processes describing the nodes, while the processes themselves are influenced by the changing network structure. We derive an efficient variational procedure for inference, and validate the model using both synthetic and real-world data.
24.1 Introduction

Networks are a useful paradigm for representing various social, biological, and technological systems. Modeling the structure and formation of networks is made more difficult when the nodes in the network and the topology of the network change over time. The growth of the internet and social media, in particular, has provided researchers with huge amounts of data that make such studies both feasible and highly desirable.

A standard approach to network modeling assumes a generative model for links based on node attributes. That is, the nodes or objects modeled are assumed to have some (possibly latent) attributes, e.g., group membership, and these latent properties determine the formation of links between nodes. A version of this approach which has achieved great success is the mixed membership stochastic blockmodel (MMSB) (Airoldi et al., 2008). MMSBs recognize that nodes often have multiple attributes (mixed membership) that may come into play when determining whether two nodes should be linked. Thus, MMSBs are a special case of a more general class of latent space models, which assume that nodes’ attributes are described in some abstract space, and the formation of links between nodes depends on the distance between their attributes in that space (Hoff et al., 2002; Krioukov et al., 2009). In MMSBs each actor is characterized by a probability distribution over his attributes, so the corresponding latent space is a simplex (Airoldi, 2007; Blei and Fienberg, 2007).

A common limitation of these approaches is that the attributes of nodes are assumed to be unchanging over time. If the nodes represent people, for instance, we know that attributes like interests, location, or job may change over time and this may affect a person’s connections to the network. In this case, it is necessary to model the dynamics of the nodes’ hidden attributes as well. Despite recent progress in modeling time-varying networks (Fu et al., 2009; Ho et al., 2011; Kolar et al., 2010; Kolar and Xing, 2011; Xing et al., 2010), there are still some open problems. In particular, the existing models so far have neglected the possibility that the change in a node’s attributes at one time step may depend on the network structure at previous time steps. The network structure, on the other hand, depends on the nodes’ attributes, thus resulting in a feedback loop between node dynamics and network evolution.

A concrete example of this phenomena occurs in social networks. For instance, it is known that new friendship links are often formed as a result of selection effects like homophily: actors often befriend people with similar interests (Snijders et al., 2006). In turn, social actors introduce their friends to new ideas and interests in a process known as social influence or diffusion. Together, these dynamics cause both the nodes and the network structure to evolve simultaneously.

Our contribution is to combine a model of node dynamics that depends on network topology with an MMSB-inspired generative model for link formation that depends on changing node attributes. We use this model to describe the co-evolution of selection and influence for real-world dynamic network data. The rest of the chapter is structured as follows. We begin with a high-level description of dynamic networks and how we can adapt MMSBs to describe them, followed by a discussion of related work. In Section 24.2, we describe the details of our co-evolving mixed membership stochastic blockmodel (CMMSB), including a discussion of how to efficiently infer model parameters. In Section 24.3, we apply a CMMSB to a synthetic dataset and a real-world dataset consisting of the bill co-sponsorship network among U.S. senators. A discussion of results follows in Section 24.4. We provide detailed calculations in the Appendix.

24.1.1 Selection and Influence in Networks

Suppose we have \( N \) nodes and we observe a network structure among them at discrete time steps, \( t = 0, 1, \ldots, T \). If there exists a directed link from node \( p \) to node \( q \) at time \( t \), we say \( Y_t(p, q) = 1 \),
otherwise 0. There are many examples of real-world data that fit this format including friendship ties in a social network and gene regulatory networks.

We suppose that the nodes themselves are described by some hidden attribute that changes over time, i.e., node $p$ is described at time $t$ by $\mu^t_p$. For a social network, this vector could represent interests, group membership, or behavioral traits, while in a gene regulatory network this could indicate response to stages of a cell cycle. Then by selection we mean that the probability of a link between two nodes depends on their attribute vector:

$$\text{Prob}(Y_t(p,q) = 1) = g(\mu^t_p, \mu^t_q).$$

(24.1)

One of the most famous forms of selection is homophily, or assortative mixing, which states that nodes tend to interact with other nodes that have similar attributes. We stress, however, that different selection mechanisms are possible as well, i.e., disassortative mixing patterns such as buyer-seller relationship, etc.

The next step is to explicitly model the dynamics in the latent space. For instance, $\mu$ may drift over time, or perhaps it responds to either one-time or recurring external events. As discussed in the introduction, we are particularly interested in modeling the influence of a node’s neighbors on his/her dynamics. Toward this end, we allow a feedback mechanism where an interaction at one time step affects the position of the node at the next time step. That is, we want to model dynamics of the form

$$\mu^{t+1}_p = f(\mu^t_p, \mu^t_{S^t_p}),$$

(24.2)

where $S^t_p$ denotes neighbors of node $p$ at time $t$. For instance, to model positive social influence one should select a function $f$ such that the distance between nodes contracts after the interaction. It is possible to have more general (e.g., repulsive) interactions as well, depending on the concrete scenario.

Together, Equations (24.1) and (24.2) provide a very high-level description of our approach. We would like to emphasize that while distance-based interactions (such as given by Equation (24.1)) are at the core of most prior work, introducing a feedback mechanism via the influence model as in Equation 24.2 is one of the main ideas distinguishing our approach from a previous attempt to formulate dynamic MMSBs in Fu et al. (2009).

Once we have specified a model for node dynamics, the task of fixing a model for link formation remains. Ideally, a generative model for link formation based on the node dynamics should capture our intuitions about real link formation while admitting some uncertainty and allowing efficient inference. For these reasons, we chose to adapt MMSBs, which we describe in the next section.

### 24.1.2 Mixed Membership Stochastic Blockmodels

In this paper we will use a latent space representation of the nodes based on MMSBs (Airoldi et al., 2008). In this section, we will purposely adhere to a high-level description of MMSBs and their dynamic extensions, whereas we will discuss a detailed implementation in Section 24.2. Starting with a static MMSB, we see that each node has a normalized mixed membership vector $\pi_p \in \mathbb{R}^K$, which describes the probability for node $p$ to take one of $K$ roles. The role that a node takes in a particular interaction is sampled according to the membership vector, and the probability of a link between $p, q$ then depends on the roles they take and the role compatibility matrix, $B$. The generative process is as follows:

- $\pi_p \sim \text{Prior distribution}$
- $z_{p\rightarrow q} \sim \text{Multinomial}(\pi_p)$
- $z_{q\rightarrow p} \sim \text{Multinomial}(\pi_q)$
- $Y(p, q) \sim \text{Bernoulli}(z_{p\rightarrow q}^\top B z_{p\rightarrow q})$. 


The most naive dynamic extension is to simply add a $t$ index to all the variables in the previous expression. This amounts to learning $T$ independent, static MMSBs and fails to take into account any of our knowledge of the underlying node dynamics. An extension considered in Fu et al. (2009) is to say that the prior distribution for the $\pi^t$ should evolve over time. However, each mixed membership vector is still sampled from the same distribution at each time, so the effect is to model only aggregate dynamics.

In contrast, and as discussed in the previous section, we would prefer that the mixed membership vector of nodes evolved individually but under mutual influence. The particular form of influence we will study is

$$\mu_{p}^{t+1} = (1 - \beta_p)\mu_{p}^{t} + \beta_p\mu_{avg}^{t} + \text{noise term},$$

(24.3)

where $\mu_{avg}^{t} = \frac{1}{|S_p^t|} \sum_{q \in S_p^t} w_{p \rightarrow q} \mu_{q}^{t}$ is the weighted average of node $q$’s neighbors’ log-membership vectors. Thus, the membership vector of node $q$ at time $t+1$ is a weighted average of his membership vector at time $t$ as well as the membership vectors of the nodes he has interacted with at time $t$. This feature of our model has the desired effect of incorporating feedback between network structure and individual node dynamics. The relative importance of the neighbors is captured by the parameter $0 < \beta_p < 1$; larger $\beta_p$ means that node $p$ is more susceptible to influence from his neighbors.

Before proceeding further, we note that exact inference is not feasible even for static MMSBs, so adding dynamics to a model makes the inference problem much harder. Here we use a variational EM approach that allows us to do efficient approximate inference (Beal and Ghahramani, 2003; Xing et al., 2003).

### 24.1.3 Related Work

The problem of properly characterizing selection and influence has been a subject of extensive studies in sociology. For instance, Steglich et al. (2010) suggested a continuous time agent-based model of network co-evolution. In this model, each agent is characterized by a certain utility function that depends on the agent’s individual attributes as well as his/her local neighborhood in the network. The agents evolve as continuous-time Markovian processes which, at randomly chosen time points, select an action to maximize their utility. Despite its intuitive appeal, a serious shortcoming of this model is that it cannot handle missing data well, thus most of the attributes have to be fully observable. This was addressed in Fan and Shelton (2009), where a continuous dynamic Bayesian approach was developed. Continuous-time models have certain advantages when the network observations are infrequent and well-separated in time. In situations where more fine-grained data is available, however, discrete-time models are more suitable (Hanneke et al., 2010).

The model represented here is based on MMSBs (Airoldi et al., 2008). MMSBs are an extension of stochastic blockmodels that have been studied extensively both in social sciences and in computer science (Holland et al., 1983; Goldenberg et al., 2010). In a stochastic blockmodel each node is assigned to a block (or a role), and the pattern of interactions between different nodes depends only on their block assignment. Many situations, however, are better described by multi-faceted interactions, where nodes can bear multiple latent roles that influence their relationships to others. MMSBs account for such “mixed” interactions by allowing each node to have a probability distribution over roles and by making the interactions role-dependent (Airoldi et al., 2008). A different approach to mixed membership community detection has been developed in physics (Ball et al., 2011; Ahn et al., 2010). In particular, Ahn et al. (2010) suggested a definition of communities in terms of links rather than nodes.

Previously, several works have considered a dynamic extension of the MMSB which we will henceforth refer to as dMMSB (Fu et al., 2009; Ho et al., 2011; Xing et al., 2010). In contrast to dMMSB, where the dynamics were imposed externally, our model assumes that the membership evolution is driven by the interactions between the nodes through a parametrized influence mechanism. At the same time, the patterns of those interactions themselves change due to the evolution of
the node memberships. An advantage of the present model over dMMSB is that the latter models the aggregate dynamics, e.g., the mean of the logistic normal distribution from which the membership vectors are sampled. CMMSB, however, models each node’s trajectory separately, thus providing better flexibility for describing system dynamics. Of course, more flexibility comes at a higher computational cost, as CMMSBs track the trajectories of all nodes individually. This additional cost, however, can be well justified in scenarios when the system as a whole is almost static (e.g., no shift in the mean membership vector), but different subsystems experience dynamic changes. One such scenario that deals with political polarization in the U.S. Senate is presented in our experimental results section.

24.2 Co-evolving Mixed Membership Blockmodel

Consider a set of $N$ nodes, each of which can have $K$ different roles, and let $\pi_p^t$ be the mixed membership vector of node $p$ at time $t$. Let $Y_t$ be the network formed by those nodes at time $t$: $Y_t(p,q) = 1$ if the nodes $p$ and $q$ are connected at time $t$, and $Y_t(p,q) = 0$ otherwise. Further, let $Y_{0:T} = \{Y_0, Y_1, \ldots, Y_T\}$ be a time sequence of such networks. The generative process that induces this sequence is described below.

- For each node $p$ at time $t = 0$, employ a logistic normal distribution\(^1\) to sample an initial membership vector,

$$\pi_{p,k}^0 = \exp(\mu_{p,k}^0 - C(\mu_p^0)), \quad \mu_p^0 \sim \mathcal{N}(\alpha^0, A),$$

where $C(\mu) = \log(\sum_k \exp(\mu_k))$ is a normalization constant, and $\alpha^0, A$ are the prior mean and covariance matrix.

- For each node $p$ at time $t > 0$, the mean of each normal distribution is updated due to influence from the neighbors at its previous step:

$$\alpha_p^t = (1 - \beta_p)\mu_{p}^{t-1} + \beta_p \mu_{S_p}^{t-1},$$

where $\mu_{S_p}^t$ is the average of the weighted membership vector $\mu$-s of the nodes which node $p$ is connected to at time $t$:

$$\mu_{S_p}^t = \frac{1}{|S_p|} \sum_{q \in S_p} w_{p,q} \mu_q^t,$$

$\beta_p$ describes how easily the node $p$ is influenced by its neighbors, while the weights, $w$, allow for different degrees of influence from different neighbors. The membership vector at time $t$ is

$$\pi_{p,k}^t = \exp(\mu_{p,k}^t - C(\mu_p^t)), \quad \mu_p^t \sim \mathcal{N}(\alpha_p^t, \Sigma_\mu),$$

where the covariance $\Sigma_\mu$ accounts for noise in the evolution process.

- For each pair of nodes $p, q$ at time $t$, sample role indicator vectors from multinomial distributions:

$$z_{p \rightarrow q}^t \sim \text{Multinomial}(\pi_p^t), \quad z_{p \leftarrow q}^t \sim \text{Multinomial}(\pi_q^t).$$

Here $z_{p \rightarrow q}$ is a unit indicator vector of dimension $K$, so that $z_{p \rightarrow q,k} = 1$ means node $p$ undertakes role $k$ while interacting with $q$.

---

\(^1\)We found that the logistic normal form of the membership vector suggested in Fu et al. (2009) leads to more tractable equations compared to the Dirichlet distribution used for static MMSBs.
Sample a link between $p$ and $q$ as a Bernoulli trial:

$$Y_t(p, q) \sim \text{Bernoulli}((1 - \rho)z_{p+q}^{t\top}B^t_{p+q}z_{p+q}^t),$$

where $B$ is a $K \times K$ role compatibility matrix, so that $B^t_{p+q}$ describes the likelihood of interaction between two nodes in roles $r$ and $s$ at time $t$. When $B^t$ is diagonal, the only possible interactions are among the nodes in the same role. Here $\rho$ is a parameter that accounts for the sparsity of the network (Airoldi et al., 2008).

Thus, the coupling between dynamics of different nodes is introduced by allowing the role vector of a node to be influenced by the role vectors of its neighbors. To benefit from computational simplicity, we updated $\pi$ by changing its associated $\mu$. This update of $\mu$ is a linear combination of its current state and the values of its neighbors’ current states. The influence is measured by a node-specific parameter $\beta_p$, and $w_{p+q}$, which need to be estimated from the data. $\beta_p$ describes how easily the node $p$ is influenced by its neighbors: $\beta_p = 0$ means it is not influenced at all, whereas $\beta_p = 1$ means the behavior is solely determined by the neighbors. On the other hand, $w_{p+q}$ reflects the weight of the specific influence that node $q$ exerts on node $p$, so that larger values correspond to more influence.

### 24.2.1 Inference

Under the CEMMSB, the joint probability of the data $Y_{0:T}$ and the latent variables $\{\mu^t_{1:N}, \mathbf{z}^t_{p+q} : p, q \in N, \mathbf{z}^t_{p+q} : p, q \in N\}$ can be written in the following factored form. To simplify the notation, we define $\mathbf{z}^t_{p+q}$ as a pair of $\mathbf{z}^t_{p+q}$ and $\mathbf{z}^t_{p+q}$. Also, denote the sets of latent group indicators $\{\mathbf{z}^t_{p+q} : p, q \in N\}$ as $\mathbf{Z}^t_{p+q}$ and $\mathbf{Z}^t_{p+q}$.

$$p(Y_{0:T}, \mathbf{Z}^0_{0:T}, \mathbf{Z}^0_{0:T}, \mathbf{Z}^0_{0:T} | \alpha, A, B, \beta_p, w_{p+q}, \Sigma) =$$

$$\prod_{t \in N} P(Y_t(p, q) | \mathbf{z}^t_{p+q}, B^t) P(\mathbf{z}^t_{p+q} | \mu_p^t, \mu_q^t)$$

$$\times \prod_{p} P(\mu_p^0 | \alpha^0, A) \prod_{t \neq 0} P(\mu_p^t | \mu^{t-1}_p, \mu^{t-1}_q, \Sigma, \beta_p).$$

In Equation (24.4), the term describing the dynamics of the membership vector is defined as follows:

$$P(\mu^t_p | \mu^{t-1}_p, \mu^{t-1}_q, \Sigma, Y_t, \beta_p) = f_G(\mu^t_p - f_b(\mu^{t-1}_p, \mu^{t-1}_q, \Sigma), \mu)$$

$$f_G(\mathbf{x}, \Sigma) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} e^{-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x}}$$

$$f_b(\mu^{t-1}_p, \mu^{t-1}_q) = (1 - \beta_p)\mu^{t-1}_p + \beta_p \mu^{t-1}_q.$$

As we already mentioned, performing exact inference with this model is not feasible. Thus, one needs to resort to approximate techniques. Here we use a variational EM approach (Beal and Ghahramani, 2003; Xing et al., 2003). The main idea behind variational methods is to posit a simpler distribution $q(X)$ over the latent variables with free (variational) parameters, and then fit those parameters so that the distribution is close to the true posterior in KL divergence.

$$D_{KL}(q||p) = \int_X q(X) \log \frac{q(X)}{p(X, Y)} dX.$$ (24.6)

Here we introduce the following factorized variational distribution:
where \( q_1 \) is the normal distribution, and \( q_2 \) is the multinomial distribution, and \( \gamma^t_p, \Sigma_p, \phi^t_{p\rightarrow q}, \phi^t_{p\leftarrow q} \) are the variational parameters. Intuitively, \( \phi^t_{p\rightarrow q,g} \) is the probability of node \( p \) undertaking the role \( g \) in an interaction with node \( q \) at time \( t \), and \( \phi^t_{p\leftarrow q,h} \) is defined similarly.

For this choice of the variational distribution, we rewrite Equation (24.6) as follows:

\[
D_{KL}(q||p) = \sum_t \sum_p \left( E_q[\log \prod_p q_1(\mu^t_p, \gamma^t_p, \Sigma_p)] + E_q[\log \prod_{p,q} q_2(z^t_{p\rightarrow q}|\phi^t_{p\rightarrow q})] \right) - E_q[\sum_p \log \prod_{q \neq 0} P(\mu^t_p|\mu^t_p, \Sigma_p)] - E_q[\log \prod_p P(\mu^t_0|\alpha^0, A)].
\]

In the third line of the above equation, we need to compute the expected value of \( \log [\sum_k \exp(\mu_k)] \) under the variational distribution, which is problematic. Toward this end, we introduce \( N \) additional variational parameters \( \xi \), and replace the expectation of the log by its upper bound induced from the first-order Taylor expansion (Blei and Lafferty, 2007):

\[
\log [\sum_k \exp(\mu_k)] \leq \log \xi - 1 + \frac{1}{\xi} \sum_k \exp(\mu_k).
\]

The variational EM algorithm works by iterating between the E-step of calculating the expectation value using the variational distribution, and the M-step of updating the model (hyper)parameters so that the data likelihood is locally maximized. The pseudo-code is shown in Algorithm 1, and the details of the calculations are discussed below.

### 24.2.2 Variational E-step

In the variational E-step, we minimize the KL distance over the variational parameters. Taking the derivative of KL divergence with respect to each variational parameter and setting it to zero, we obtain a set of equations that can be solved via iterative or other numerical techniques. For instance, the variational parameters \( \phi^t_{p\rightarrow q,g}, \phi^t_{p\leftarrow q,h} \) corresponding to a pair of nodes \( (p, q) \) at time \( t \), can be found via the following iterative scheme:

\[
\phi^t_{p\rightarrow q,g} \propto \exp(\gamma^t_{p,q}) \prod_h (B(g, h) Y_t(p, q) (1 - B(h, g))^{1 - Y_t(p,q)}) \phi^t_{p\rightarrow q,h}. \tag{24.10}
\]

\[
\phi^t_{p\leftarrow q,h} \propto \exp(\gamma^t_{q,h}) \prod_g (B(g, h) Y_t(p, q) (1 - B(g, h))^{1 - Y_t(p,q)}) \phi^t_{p\rightarrow q,h}. \tag{24.11}
\]

In the above equations, \( \phi^t_{p\rightarrow q,g} \) and \( \phi^t_{p\leftarrow q,h} \) are normalized after each update. Note also that Equations (24.10) and (24.11) are coupled with each other as well as with the parameters \( \gamma^t_{p,q,g}, \gamma^t_{q,h} \).

Sets of variational parameters, \( \{\gamma\}^t \) and \( \{\sigma\}^t \), are initialized at the beginning of variational EM.
Input: data \( Y_t(p,q) \), size \( N, T, K \)

Initialize all \( \{ \gamma \}_t, \{ \sigma \}_t \)

Start with an initial guess for the model parameters.

repeat
  repeat
    for \( t = 0 \) to \( T \) do
      repeat
        Initialize \( \phi_{t,p\rightarrow q}, \phi_{t,p\rightarrow q} \) to \( \frac{1}{K} \) for all \( g, h \)
        repeat
          Update all \( \{ \phi \}_t \)
        until convergence of \( \{ \phi \}_t \)
        Find \( \{ \gamma \}_t, \{ \sigma \}_t \)
        Update all \( \{ \zeta \}_t \)
      until convergence in time \( t \)
    end for
  until convergence across all time steps

Update hyperparameters.

until convergence in hyperparameters

Algorithm 1: Variational EM.

For \( \{ \gamma \}_t \), we sample it from normal distribution \( N(\alpha^0, A) \), and for \( \{ \sigma \}_t \) we initialize it to the same value over all nodes across the whole time steps. Once the \( \{ \phi \}_t \) are converged to optimal points, we then update \( \{ \gamma \}_t \) and \( \{ \sigma \}_t \) using the update equations. Both of the variational parameters do not have closed forms of solution, and the details are given in the KL-Distance section of the Appendix. Here we simply note that their general form is:

\[
\gamma_{t,p} = f_\gamma(\gamma_{t-1,p}, \gamma_{t+1,p}, \gamma_{t,q}, \phi_{t,p\rightarrow q}, \phi_{t,q\rightarrow p}, \zeta_{t,p}, \Sigma_{t,p}).
\] (24.12)

Thus, the parameter \( \gamma_{t,p} \) depends on its immediate past and future values, \( \gamma_{t-1,p} \) and \( \gamma_{t+1,p} \), as well as the parameters of its neighbors.

For the variational parameters of a covariance matrix \( \Sigma_{t,p} \), which is assumed to be a diagonal matrix with components \( (\sigma_{t,p,1}^2, (\sigma_{t,p,2}^2, ... (\sigma_{t,p,k}^2) \), the general form of the optimal point is:

\[
\sigma_{t,p,k} = f_\sigma(\gamma_{t,p,k}, \zeta_{t,p,k}).
\] (24.13)

Finally, for the variational parameters \( \zeta \) we have

\[
\zeta_{t,p} = \sum_i \exp(\gamma_{t,i} + \frac{(\sigma_{t,i}^2)}{2}).
\] (24.14)

Note that the above equations can be solved via a simple iterative update as before. To expedite convergence, however, we combine the iterations with the Newton-Raphson method, where we solve for individual parameters while keeping the others fixed, and then repeat this process until all the parameters have converged.

24.2.3 Variational M-step

The M-step in the EM algorithm computes the parameters by maximizing the expected log-likelihood found in the E-step. The model parameters in our case are: \( B^t \), the role compatibility matrix, the covariance matrix \( \Sigma_{t,p} \), \( \beta_p \) for each node, \( w_{t,p\rightarrow q} \) for each pair, \( \alpha \), and \( A \) from the prior.

If we assume that the time variation of the block compatibility matrix is small compared to the
Mixed Membership Blockmodels for Dynamic Networks with Feedback

535

evolution of the node attributes, we can neglect the time dependence in $B$ and use its average across
time, which yields:

$$\hat{B}(g, h) = \frac{\sum_{p,q,t} Y_t(p, q) \cdot \phi^t_{p\rightarrow q,g} \phi^t_{p\rightarrow q,h}}{\sum_{p,q,t} \phi^t_{p\rightarrow q,g} \phi^t_{p\rightarrow q,h}}.$$  

(24.15)

Likewise, for the update of diagonal components of the noise covariance matrix $\Sigma_\mu$,

$$\langle \gamma_k \rangle^2 = \frac{1}{N(T-1)} E_q \left[ \sum_{p,t} (\mu^t_{p,k} - (1 - \beta) \mu^{t-1}_{p,k} - \beta \mu^{S_{t-1},k})^2 \right].$$  

(24.16)

Similar equations are obtained for $\beta_p$ and $w^t_{p\rightarrow q}$. The update equation of $\beta_p$ and $w^t_{p\rightarrow q}$ is a function
of $\gamma$ and $\sigma$, which are related to the transition for specific node $p$.

$$\beta_p = \frac{\sum_{t>0} \sum_k (\gamma^t_{p,k} - \gamma^{t-1}_{p,k} - \gamma^{t-1}_{p,k} - \gamma^{t-1}_{p,k} \gamma^{S_{t-1},k} + \gamma^{t-1}_{p,k} \gamma^{S_{t-1},k})}{\sum_{t>0} \sum_k (\gamma^t_{p,k} - \gamma^{t-1}_{p,k} - 2 \gamma^{t-1}_{p,k} \gamma^{S_{t-1},k} + \sum_{t>0} \sum_k (\gamma^{S_{t-1},k} + \sigma^{S_{t-1},k})},$$

where $\gamma_{S_{p}}$ and $\Sigma_{S_{p}}$ are the mean and covariance of a set of nodes which node $p$ is connected to at
time $t$.

The priors of the model can be expressed in closed form as below:

$$\alpha^0 = \frac{1}{N} \sum_{p} \gamma^0_{p,k}.$$  

(24.17)

$$a_k = \sqrt{\frac{1}{N} \sum_{p} \left( \gamma^0_{p,k} \right)^2 + \left( \sigma^0_{p,k} \right)^2 - 2 a_k^0 \gamma^0_{p,k} + (a_k^0)^2}. $$

(24.18)

24.3 Results

24.3.1 Experiments on Synthetic Data

We tested our model by generating a sequence of networks according to the process described
above, for 50 nodes, and $K = 3$ latent roles across $T = 8$ time steps. We used a covariance matrix
of $A = 3I$, and mean $\alpha^0$ having homogeneous values for the prior, so that initially nodes had a well-defined role (i.e., the membership vector would have peaked around a single role). More precisely,
the majority of nodes had around 90% of membership probability mass centered at a specific role,
and on average a third of those nodes had 90% on role $k$. For the role compatibility matrix, we gave
high weight at the diagonal.

Starting from some initial parameter estimates, we performed variational EM and obtained re-
estimated parameters which were very close to the original values (ground truth). With those learned
parameters, we inferred the hidden trajectory of agents as given by their mixed membership vector
for each time step. The results are shown in Figure 24.1, where, for three nodes, we plot the projection
of trajectories onto the simplex. One can see that for all three nodes, the inferred trajectories
are very close to the actual ones.
24.3.2 Comparison with dMMSB

As a further verification of our results, we compare the performance of our inference method to the dynamic mixed membership stochastic blockmodel (dMMSB) (Fu et al., 2009). We use synthetic data generated in a manner similar to the previous section. This time, though, for simplicity we keep $K = 2$ and set all the $\beta$s to some constant for all the nodes: $\beta = 0.1$ in one trial and $\beta = 0.2$ in the other. In this case, we compare performance by evaluating the distance in $L_2$ norm between actual and inferred mixed membership vectors for each method. At each time step, we calculate the average over all nodes of the $L_2$ distance from the actual membership vector.

As we show in Figures 24.2(a) and 24.2(b), CMMSBs capture the dynamics better than the dMMSBs. This is due to the fact that our model tracks all of the nodes individually (internal dynamics), while dMMSBs regard the dynamism as an evolution of the environment (external dynamics). Here, we have only included results for relatively small and homogeneous dynamics. In fact, we noticed that our method tends to fare even better as we increase the degree of dynamics or the heterogeneity of dynamics across nodes (node-varying values of $\beta$). We believe heterogeneous dynamics are more prevalent in real systems, and so we expect our method to outperform dMMSB even more than is indicated by Figure 24.2(b).

24.3.3 U. S. Senate Co-Sponsorship Network

We have also performed some preliminary experiments for testing our model against real-world data. In particular, we used senate co-sponsorship networks from the 97th to the 104th Senate, by considering each senate as a separate time point in the dynamics. There were 43 senators who remained part of the senate during this period. For any pair of senators $(p, q)$ in a given senate, we generated a directed link $p \rightarrow q$ if $p$ co-sponsored at least three bills that $q$ originally sponsored. The threshold of three bills was chosen to avoid having too dense of a network. With this data, we wanted to test (a) to what extent senators tend to follow others who share their political views (i.e., conservative vs. liberal) and (b) whether some senators change their political creed more easily than others.
FIGURE 24.2
Inference error for dMMSB and CMMSB for synthetic data generated with $K = 2$ and $\beta = 0.1$ for all the nodes (a), and when $\beta = 0.2$ for all the nodes (b).

The number of roles $K = 2$ was chosen to reflect the mostly bi-polar nature of the U.S. Senate. The susceptibility of senator $p$ to influence is measured by the corresponding parameter $\beta_p$, which is learned using the EM algorithm. High $\beta$ means that a senator tends to change his/her role more easily. Likewise, the power of influence of senator $q$ on senator $p$ is measured by the parameter $w_{p\leftarrow q}$, where $w_{p\leftarrow q_1} > w_{p\leftarrow q_2}$ means senator $q_1$ is more influential on senator $p$ than senator $q_2$. Here the direction of the arrow reflects the direction of the influence which is opposite to the direction of the link. To initialize the EM procedure, we assigned the same $\beta$ and $w$ to all the senators, and start with a matrix which is weighted at the diagonal for $B$.

Another method for validation is to compare the degree of influence. Our model handles and learns the degree of influence in the update equation. Sorting out influential senators is an area of active research. Recently, KNOWLEGIS has been ranking U.S. senators based on various criteria, including influence, since 2005. Since our data was extracted from the 97th Senate to the 104th Senate, direct comparison of the rankings was impossible. Another study (Maisel, 2010) ranked the 10 most influential senators in both parties who have been elected since 1955. We compared our top five influential senators, and were able to find three senators (Senator Robert Byrd, Senator Strom Thurmond, and Senator Bob Dole) on the list.

24.3.4 Interpreting Results
The role compatibility matrix learned from the variational EM has high values on the diagonal confirming our intuition that interaction is indeed more likely between senators that share the same role. Furthermore, the learned values of $\beta$ showed that senators varied in their “susceptibility.” In particular, Senator Arlen Spector was found to be the most influenceable one, while Sen. Dole was found to be one of the most inert ones. Note that while there are no direct ways of estimating the “dynamism” of senators, our results seem to agree with our intuition about both senators (e.g., Sen. Spector switched parties in 2009 while Sen. Dole became his party’s candidate for President in 1996).

To get some independent verification, we compared our results to the yearly ratings that the ACU (American Conservative Union) and ADA (Americans for Democratic Action) assign to senators.\(^3\)

\(^3\)Accessible at http://www.conservative.org/, http://www.adaction.org/.
ACU/ADA rated every senator based on selected votes which they believe display a clear ideological distinction, so that high scores in the ACU mean that they are truly conservative, while lower scores in the ACU suggest they are liberal, and for the ADA vice versa. To compare the ratings with our predictions (given by the membership vector) we scaled the former to get scores in the range \([0, 1]\).

Figure 24.3 shows the relationship between these scores and our mixed membership vector score, confirming our interpretation of the two roles in our model as corresponding to liberal/conservative. Although these values cannot be used for quantitative agreement, we found that at least qualitatively, the inferred trajectories agree reasonably well with the ACU/ADA ratings. This agreement is rather remarkable since the ACU/ADA scores are based on selected votes rather than co-sponsorship network as in our data.

![Correlation between ACU/ADA scores and inferred probabilities.](image)

FIGURE 24.3
Correlation between ACU/ADA scores and inferred probabilities.

Of course, we are most interested in correctly identifying the dynamics for each senator. We compare our inferred trajectory of the most dynamic senator, and the inert senator to the scores of the ACU and ADA. In Figure 24.4 the scores of the ADA have been flipped, so that we can compare all of the scores in the same measurement. However, since ACU/ADA scores are rated for every senator each year, the dynamics of inference and the dynamics of ACU/ADA scores cannot be compared one to one. Not all senators showed high correlation of the trend like Sen. Specter and Sen. Dole.

24.3.5 Polarization Dynamics

The yearly ACU/ADA scores give a good comparison of the relative political position of senators scored in each year. However, they are not very appropriate for comparison between years, a point illustrated by the fact that the score is based on voting records for different bills in each year. Therefore, for validation of the dynamics we turn to another scoring system highly regarded by political scientists and used to observe historical trends, the DW-NOMINATE score. For the time period of our study, McCarty et al. (2006) shows that the political polarization of the senate was increasing. In particular, they show that the gap between the average DW-NOMINATE score of Republicans and Democrats is monotonically increasing, as we show in Figure 24.5. In fact, the polarization for the entire senate was stronger every year. This is due to the unbalanced seats in the entire senate. In other words, our data had 22 Republicans and 21 Democrats, while for the entire senate, majority outnumbered minority by around 10 seats. For comparison, for each time step we took the average of our inferred score for the 14 most and least conservative senators. As we show
in Figure 24.5, our inferred result agrees qualitatively with the results of McCarty et al. (2006), showing an increase in polarization for every senate in the studied time-window. Since the DW-NOMINATE score uses its own metric, and our polarization is measured by the difference between upper average and lower average probability, we should not expect to get quantitative agreement. We would like to highlight, however, that the direction of the trend is correctly predicted for each of the eight terms.

24.4 Discussion

We have presented the CMMSB for modeling inter-coupled node and link dynamics in networks. We used a variational EM approach for learning and inference with CMMSB, and were able to
reproduce the hidden dynamics for synthetically generated data, both qualitatively and quantitatively. We also tested our model using the U.S. Senate bill co-sponsorship data, and obtained reasonable results in our experiments. In particular, CMMSBs were able to detect increasing polarization in the senate as reported by other sources that analyze individual voting records of the senators.

Our results with the U.S. Senate dataset suggest that our dynamical model can actually capture some nuances of individual dynamics. While we lack a ground truth for the true position of senators, third party analyses qualitatively support the findings of our model. Of course, many factors are not explicitly modeled in our approach, but we hope that by including individual dynamical terms we capture these effects implicitly. For instance, external events like upcoming re-election campaigns surely affect senator’s actions. While the true chain of events may rely on these events, if all relevant external events are not or cannot be included in our model, then capturing dynamics through shifts in observed relationships is a good proxy.

The approach to modeling influence described in Section 24.2 is only one of several possibilities. Although we learned a static parameter \( \beta \) for each node, describing how easily influenced they are, we also pointed out the possibility of adding a weight that varies for each pair: that is, a node may be more influenced by one person than another. Additionally, someone’s influence may change over time. Finally, we chose a simple linear influence mechanism. In principle, someone may be more influential along one axis than another. For instance, a node may be influenced by a friend’s musical taste, but not by his politics.

As future work, we intend to test our model against different real-world data, such as communication networks or co-authorship networks of publications. We also plan to extend CMMSBs in several ways. A significant bottleneck of the current model is that it explicitly considers links between all the pairs of nodes, resulting in a quadratic complexity in the network size. Most real-world networks, however, are sparse, which is not accounted for in the current approach. Introducing sparsity into the model would greatly enhance its efficiency. We note that this is also a drawback for static MMSBs, but progress has already been made towards reducing this complexity (Mørup et al., 2011).

An additional drawback of MMSBs (and stochastic blockmodels in general) is the inability to properly deal with degree heterogeneity. Indeed, MMSBs (or related latent space models) might assign nodes to the same group based merely on the frequency of their interactions with the other nodes. Possible remedies are found in the degree-correct blockmodel recently proposed in Karrer and Newman (2011) or in exponential random graph models that separately model node and group variability (Reichardt et al., 2011). The problem reveals a fundamental ambiguity about network modeling. A priori, we have no reason to believe that node connectivity is a less important dimension for clustering nodes than homophily for some hidden attribute. Our intuition leads us to expect otherwise for human networks, but this intuition must be explicitly modeled. In the co-sponsorship network studied here, most senators are well-connected and so the network structure is better explained by political views than node connectivity. However, large variability in node connectivity has been observed in many social networks where this effect will have to be explicitly modeled.

**Acknowledgments**

This research was supported in part by the National Science Foundation under grant No. 0916534, U.S. ARO MURI grant No. W911NF0610094, and US AFOSR MURI grant No. FA9550-10-1-0569.
Appendix

Alternative View of EM Algorithm

We start with the log-likelihood function where $Y$ is the data, $X$ is the set of latent variables, and $\Theta$ is the set of model parameters:

$$
\log p(Y|\Theta) = \log \int p(Y,X|\Theta) dX
$$

$$
= \log \int q(X) \frac{p(Y,X|\Theta)}{q(X)} dX 
\geq \int q(X) \log \frac{p(Y,X|\Theta)}{q(X)} dX \quad \text{(Jensen’s Inequality)}.
$$

We define the lower bound as free energy:

$$
F(q, \Theta) = \int q(X) \log p(X|Y, \Theta) p(Y|\Theta) q(X) dX \quad (24.20)
$$

$$
= \log P(Y|\Theta) - D_{KL}(q(X)||p(Y,X)).
$$

The goal is to maximize the lower bound (free energy) by updating $q$ and $\Theta$. In E-step, we minimize the KL-distance of two distributions, and in M-step, we maximize the free energy under fixed $q$ distribution obtained in E-step.

KL-Distance

Here we present the KL-distance between $q(X)$, and $p(Y,X)$:

$$
D_{KL}(q||p) = 
\sum t \sum p \left(-\frac{1}{2} E_q[(\mu_p^t - \gamma_p^t)^T (\Sigma_p^t)^{-1} (\mu_p^t - \gamma_p^t)] - \log(2\pi)^{k/2} - \log(|\Sigma_p^t|^{1/2})\right)
$$

$$
+ \sum t \sum p \sum g \phi_p^{t \rightarrow q,g} \log \phi_p^{t \rightarrow q,g} + \sum t \sum p \sum h \phi_p^{t \rightarrow q,h} \log \phi_p^{t \rightarrow q,h}
$$

$$
- \sum t \sum p \sum g \phi_p^{t \rightarrow q,g} \phi_p^{t \rightarrow q,g} f(Y_t(p,q), B(g,h))
$$

$$
- \sum t \sum p \sum g \phi_p^{t \rightarrow q,g} \log \zeta_p^t + 1 - \frac{1}{\zeta_p^t} \sum_k \exp(\gamma_{p,k}^t + \frac{\sigma_{p,k}^2}{2})
$$

$$
- \sum t \sum p \sum h \phi_p^{t \rightarrow q,h} \log \zeta_q^t + 1 - \frac{1}{\zeta_q^t} \sum_k \exp(\gamma_{q,k}^t + \frac{\sigma_{q,k}^2}{2})
$$

$$
- \sum t \sum p \frac{1}{2} E_q \left[(\mu_p^t - f_b(\mu_p^{t-1}, \mu_{S_p}^{t-1}))^T \Sigma_p^{-1} \cdot (\mu_p^t - f_b(\mu_p^{t-1}, \mu_{S_p}^{t-1})) - (k/2) \log(2\pi) - \log(|\Sigma_p|^{1/2})\right]
$$

$$
- \sum p \left(-\frac{1}{2} E_q[(\mu_p^0 - \alpha^0)^T A^{-1} (\mu_p^0 - \alpha^0)] - \log(2\pi)^{k/2} - \log(|A|^{1/2})\right),
$$
where \( \exp(\gamma_{p,k} + \frac{\sigma_{p,k}^2}{2}) \) comes from the moment-generating function of the normal distribution, \( M_X(t) := E[e^{tX}] \), with \( t=1 \). The first line simplifies to \( -\sum_{t,p,k} \log \sigma_{p,k}^t \), where, once again, we have taken the covariance matrix to be diagonal.

**Variational E-step**

In the variational E-step, we minimize the KL distance over the variational parameters. Variational parameters \( \{\gamma_p\}_t \) and \( \{\sigma_{p,k}\}_t \) need to be solved analytically. We use the Newton-Raphson method as an optimization algorithm for tightening the bound with respect to those variational parameters.

First, we minimize the divergence with respect to \( \gamma_p^t \). Since the other variational parameters \( \Sigma^t_p \) are assumed to be a diagonal matrix, we treat the multivariate normal distribution as a combination of independent normal distribution and update the mean and variance for each coordinate. We use the Newton-Raphson method for each coordinate where the derivative is:

\[
dD_{KL}(q||p)/d\gamma^t_{p,k} =
\]

\[
= -\sum_{q} \phi^t_{p\rightarrow q,k} + \sum_{q} \phi^t_{p\rightarrow q,k} \frac{1}{\gamma^t_{p,k}} \exp(\gamma^t_{p,k} + \frac{\sigma^2_{p,k}}{2})
\]

\[
- \sum_{q} \phi^t_{q\rightarrow p,k} + \sum_{h} \phi^t_{q\rightarrow p,h} \frac{1}{\gamma^t_{p,k}} \exp(\gamma^t_{p,k} + \frac{\sigma^2_{p,k}}{2})
\]

\[
+ \left[ \frac{\gamma^t_{p,k}}{|\eta_k|^2} + (1 - \beta_p)^2 \frac{\gamma^t_{p,k}}{|\eta_k|^2} \right] + \sum_{q \in S^t_p} \beta_q \frac{\gamma^t_{S^t_q,k}/|S^t_q|}{|\eta_k|^2}
\]

\[
- (1 - \beta_p)^{\gamma^t_{p,k} - 1} - (1 - \beta_p)^{\gamma^t_{p,k} + 1} + \beta_p (1 - \beta_p) \frac{\gamma^t_{S^t_q,k}/|S^t_q|}{|\eta_k|^2} - \beta_p \frac{\gamma^t_{S^t_{q-1},k}/|S^t_{q-1}|}{|\eta_k|^2}
\]

\[
+ (1 - \beta_p) \sum_{q \in S^t_p} \beta_q \frac{\gamma^t_{S^t_k,k}/|S^t_k|}{|\eta_k|^2} - \sum_{q \in S^t_p} \beta_q \frac{\gamma^t_{S^t_{q+1},k}/|S^t_{q+1}|}{|\eta_k|^2}.
\]

\( \gamma^t_{S^t_p} \equiv \sum_{q \in S^t_p} (w_{p\rightarrow q}) (\gamma^t_{p,k}) \) is the mean of set of neighbors of node \( p \) at time \( t \), and \( \sigma^2_{S^t_p,k} \equiv \sum_{q \in S^t_p} (w_{p\rightarrow q})^2 (\sigma_{q,k}^t)^2 \) are the variance of set of neighbors of node \( p \) at time \( t \). Mean and variance of neighbors can be easily computed since the components of neighbors are independent of each other and are Gaussian themselves. The derivative above is valid for \( \gamma^t_{p,k} \) when \( 0 < t < T \); the form is slightly different when \( t = 0 \) or \( t = T \).

Second, we minimize the divergence with respect to \((\sigma^t_{p,1})^2, (\sigma^t_{p,2})^2, \ldots (\sigma^t_{p,K})^2)\) using the Newton-Raphson method. The derivative with respect to \( \sigma^t_{p,k} \) is:

\[
dD_{KL}(q||p)/d\sigma^t_{p,k} =
\]

\[
2(N-1) \frac{\sigma^t_{p,k}}{\gamma^t_{p,k}} \exp(\gamma^t_{p,k} + \frac{\sigma^2_{p,k}}{2}) - \frac{1}{\sigma^t_{p,k}}
\]

\[
+ 1 + (1 - \beta_p)^2 + \sum_q Y(t,q,p) \beta^2_q \frac{w^2_{p\rightarrow q}}{|S^t_q|^2} \frac{\gamma^t_{S^t_q,k}/|S^t_q|}{|\eta^2_k|} \sigma^t_{p,k}.
\]
Mixed Membership Blockmodels for Dynamic Networks with Feedback

where \( \eta_k^2 \) is the diagonal component of the covariance matrix \( \Sigma_\mu \). When \( t = 0 \) or \( t = T \), the derivative slightly differs from the above equation.

**Variational M-step**

\[
p(Y|\Theta) \geq \int_X q(X) \frac{\log p(Y, X|\Theta)}{q(X)}. \tag{24.24}
\]

The M-step in the EM algorithm computes the hyperparameters by maximizing the lower bound under fixed \( q \) found in the E-step. The lower bound of the log-likelihood is from Jensen’s inequality (Equation 24.24), and the expectation is taken with respect to a variational distribution. Hence the general form of the update equation at the \( k \)th step is as below:

\[
\Theta_k = \arg \max_\Theta \int q_k(X) \log p(Y, X|\Theta) dX. \tag{24.25}
\]

Since the final form of most model parameters are quite intuitive, we only derive Equation (24.17) in this section. To obtain the update equation of \( \beta_p \), we start from differentiating the expected log-likelihood and setting it to zero:

\[
0 = \sum_{t>T} \sum_k \left( - (1 - \beta_p) (\gamma_{p,k}^t)^2 + \sigma_{p,k}^2 \right) + \sum_{t>0} \sum_k \left( \gamma_{p,k}^{t-1} - \gamma_{p,k}^t \right) \gamma_{p,k}^{t-1} \gamma_{p,k}^t + (1 - 2\beta_p) \gamma_{p,k}^{t-1} \gamma_{p,k}^t \right). \tag{24.26}
\]

Solving the equation above,

\[
\beta_p = \frac{\sum_{t>0} \sum_k (\gamma_{p,k}^{t-1} - \gamma_{p,k}^t) \gamma_{p,k}^{t-1} \gamma_{p,k}^t + \gamma_{p,k}^t \gamma_{p,k}^{t-1})}{\sum_{t>0} \sum_k (\gamma_{p,k}^{t-1} + \gamma_{p,k}^t - 2\gamma_{p,k}^{t-1} \gamma_{p,k}^t) + \sum_{t<T} \sum_k (\gamma_{p,k}^{t} \gamma_{p,k}^{t-1} \gamma_{p,k}^t \gamma_{p,k}^{t-1})}. \tag{24.27}
\]

For solving the optimal weight, we differentiate the lower bound with respect to \( w_{p \rightarrow q_i} \) and set it to zero:

\[
0 = \sum_k \left( \beta_p^2 \frac{w_{p \rightarrow q_i}}{|S_{p}^t|} (\gamma_{q_1,k}^t)^2 + \sigma_{q_1,k}^2 \right) - \frac{\beta_p}{|S_{p}^t|} \gamma_{q_1,k}^{t+1} \gamma_{p,k}^t + \frac{\beta_p}{|S_{p}^t|} \left( \sum_{q \in S_{p}^t, q \neq q_i} Y(p, q_1) w_{p \rightarrow q} \gamma_{q,k}^t \right). \tag{24.28}
\]

Finally, the update equation for weight becomes,

\[
w_{p \rightarrow q_i} = \frac{\sum_k \frac{\beta_p}{|S_{p}^t|} \gamma_{q_1,k}^{t+1} \gamma_{p,k}^t - (1 - \beta_p) \gamma_{q_1,k}^{t+1} \gamma_{p,k}^t - \frac{\beta_p^2}{|S_{p}^t|^2} \sum_{q \in S_{p}^t, q \neq q_i} Y(p, q_1) w_{p \rightarrow q} \gamma_{q,k}^t}{\sum_k \frac{\beta_p^2}{|S_{p}^t|^2} (\gamma_{q_1,k}^t)^2 + \sigma_{q_1,k}^2}}. \tag{24.29}
\]

**References**


