Stochastic Block Transition Models for Dynamic Networks

Kevin S. Xu

Technicolor Research, 175 S. San Antonio Rd., Los Altos, CA 94022 USA
kevinxu@outlook.com

November 21, 2014

Abstract

There has been great interest in recent years in the development of statistical models for dynamic networks. This paper targets networks evolving in discrete time in which both nodes and edges can *appear* and *disappear* over time, such as dynamic networks of social interactions. We propose a *stochastic block transition model* (SBTM) for dynamic networks that is inspired by the well-known stochastic block model (SBM) for static networks and several recent dynamic extensions of the SBM. Unlike most existing dynamic models, it does not make a hidden Markov assumption on the edge-level dynamics, allowing the presence or absence of edges to directly influence future edge probabilities. We demonstrate that the proposed SBTM is significantly better at reproducing durations of edges in real social network data between edges while retaining the interpretability of the SBM.

1 Introduction

The analysis of data in the form of networks has been a topic of significant interest across many disciplines, aided by the development of statistical models for network data. Many models have been proposed for static networks, where the data consist of a single observation of the network (Goldenberg et al., 2009). On the other hand, modeling dynamic networks is still in its infancy; much research on dynamic network modeling has appeared only in the past several years. Statistical models for static networks typically utilize a latent variable representation for the network; such models have been extended to dynamic networks by allowing the latent variables, which we refer to as *states*, to evolve over time.

1

Most of the previously proposed dynamic network models assume a hidden Markov structure, where an observed snapshot of the network at any particular time is conditionally independent from all previous snapshots given the current network states. Such an approach greatly simplifies the model and allows for tractable inference; however, it may not be sufficiently flexible to replicate certain observations from real network data, namely durations of edges over time, which are extremely inaccurately reproduced by existing models.

In this paper we propose a *stochastic block transition model* (SBTM) for dynamic networks, inspired by the well-known stochastic block model (SBM) for static networks. The approach generalizes two recent dynamic extensions of SBMs that utilize the hidden Markov assumption (Yang et al., 2011; Xu and Hero III, 2014). Unlike these dynamic models, the presence of an edge between two nodes at any given time t directly influences the probability that such an edge would appear at time $t+1$ in the SBTM.

Under the SBTM, we show that the sample mean of a scaled version of the observed adjacency matrix at each time is asymptotically Gaussian. By taking advantage of this property, we develop an efficient approximate inference procedure using a combination of an extended Kalman filter and a local search algorithm. We investigate the accuracy of the inference procedure via simulation experiments. Finally, we fit the SBTM to a real dynamic social network and demonstrate its ability to more accurately replicate edge durations while retaining the interpretability of the SBM.

2 Related Work

There has been significant research dedicated to statistical modeling of dynamic networks, mostly in the past several years. Much of the earlier work is covered in the excellent survey by Goldenberg et al. (2009). Key contributions in this area include temporal extensions of

- Exponential random graph models (Guo et al., 2007).

- Stochastic block models (Xing et al., 2010; Ho et al., 2011; Ishiguro et al., 2010; Yang et al., 2011; Xu and Hero III, 2014).

- Continuous latent space models (Sarkar and Moore, 2005; Sarkar et al., 2007; Hoff, 2011; Lee and Priebe, 2011; Durante and Dunson, 2014).

- Latent feature models (Foulds et al., 2011; Heaukulani and Ghahramani, 2013; Kim and Leskovec, 2013).

Temporal extensions of stochastic block models are related to the proposed model. Xing et al. (2010) and Ho et al. (2011) proposed dynamic mixed-membership SBMs, for which a continuous class membership matrix evolves over time, and snapshots of the network are generated by first sampling a class membership for each node followed by sampling relations between nodes. Ishiguro et al. (2010) proposed a temporal extension of the infinite relation model, which is a nonparametric version of the SBM where the number of classes is estimated from the data. The model is highly flexible and is able to capture merges and splits of clusters over time. Yang et al. (2011) and Xu and Hero III (2014) proposed temporal extensions of the standard SBM; these models are closely related to the model proposed in this paper and are further discussed in Section 3.2.

Most dynamic network models assume a hidden Markov structure for tractability. Specifically, the network parameters or states often follow Markovian dynamics, and it is assumed that consecutive snapshots of the network are conditionally independent given the current states. While tractable, such an assumption may not be realistic in many settings, including dynamic networks of social interactions. For example, if two people happen to run into each other at some time, it may be more likely for them to interact again in the near future. Indeed, Viswanath et al. (2009) discovered that this was the case for wall posts between Facebook users. Over 80% of pairs of users continued to interact one month after the initial interaction, and over 60% continued after three months.

This type of observation suggests that the presence of future edges in a dynamic network should depend on whether or not an edge is currently present, and not only on the current network states. The model we propose in this paper satisfies this property. To the best of our knowledge, the only other dynamic network model satisfying this property is the latent feature propagation model proposed by Heaukulani and Ghahramani (2013).

3   Stochastic Block Models

3.1   Static Stochastic Block Models

A static network is represented by a graph over a set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$. The nodes and edges are represented by a square adjacency matrix W, where an entry $w_{ij} = 1$ denotes that an edge is present from node $i \in \mathcal{V}$ to node $j \in \mathcal{V} \setminus \{i\}$, and $w_{ij} = 0$ denotes that no such edge is present. Unless otherwise specified, we assume directed graphs, i.e. $w_{ij} \neq w_{ji}$ in general, with no self-edges, i.e. $w_{ii} = 0$. Let $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_k\}$ denote a partition of $\mathcal{V}$

into k classes. We use the notation $i \in a$ to denote that node i belongs to class a. We represent the partition by a class membership vector $\mathbf{c}$, where $c_i = a$ is equivalent to $i \in a$.

A *stochastic block model* (SBM) for a static network is defined as follows (adapted from Definition 3 in Holland et al. (1983)):

Definition 1. Let W denote a *random* adjacency matrix for a static network and $\mathbf{c}$ denote a class membership vector. W is generated according to a *stochastic block model* with respect to the membership vector $\mathbf{c}$ if and only if,

1. For any nodes $i \neq j$, the random variables w_{ij} are statistically independent.

2. For any nodes $i \neq j$ and $i' \neq j'$, if

   (a) i and i' are in the same class, i.e. $c_i = c_{i'}$,
   (b) j and j' are in the same class, i.e. $c_j = c_{j'}$,

   then the random variables w_{ij} and $w_{i'j'}$ are identically distributed.

Let $\Theta \in [0,1]^{k \times k}$ denote the matrix of probabilities of forming edges between classes, which we refer to as the *block probability matrix*. It follows from Definition 1 and the requirement that W be an adjacency matrix that $w_{ij} \sim \text{Bernoulli}(\theta_{ab})$, where $i \in a$ and $j \in b$.

SBMs are used in two settings: the *a priori* setting, where class memberships are either known or assumed, and the *a posteriori* setting, where class memberships are estimated. Recent interest has focused on the more difficult a posteriori setting, which we consider in this paper. Many methods have been proposed for a posteriori estimation of the static SBM, including Gibbs sampling (Nowicki and Snijders, 2001), label-switching (Karrer and Newman, 2011; Zhao et al., 2012), and spectral clustering (Rohe et al., 2011; Sussman et al., 2012). The label-switching methods use a heuristic for solving the combinatorial optimization problem of maximizing the likelihood over the set of possible class memberships, which is too large for an exhaustive search to be tractable. The spectral clustering methods utilize the singular vectors of the adjacency matrix W or a similar matrix to estimate the class memberships.

3.2 Dynamic Stochastic Block Models

In this paper, we consider dynamic networks evolving in discrete time. Both nodes and edges could *appear* or *disappear* over time. Let $(\mathcal{V}^t, \mathcal{E}^t)$ denote a graph snapshot, where the superscript t denotes the time index. Let $\mathcal{M}^t$ denote a mapping from $\mathcal{V}^t$, the set of observed nodes at time t, to the set $\{1, \ldots, |\mathcal{V}^t|\}$. Using the appropriate mapping $\mathcal{M}^t$, one can represent a dynamic network using a time-indexed sequence of adjacency matrices $W^{(T)} = \{W^1, \ldots, W^T\}$, and correspondence between rows and columns of different matrices can be established by inverting the mapping. In the remainder of this paper, we drop explicit reference to the mappings and assume that a node $i \in \mathcal{V}^{t-1} \cap \mathcal{V}^t$ is represented by row and column i in both W^{t-1} and W^t.

We define a *dynamic stochastic block model* (DSBM) for a time-evolving network in the following manner:

Definition 2. Let $W^{(T)}$ denote a *random sequence* of T adjacency matrices over the set of nodes $\mathcal{V}^{(T)} = \cup_{t=1}^T \mathcal{V}^t$, and let $\mathbf{c}^{(T)}$ denote a sequence of class membership vectors for these nodes. $W^{(T)}$ is generated according to a *dynamic stochastic block model* with respect to $\mathbf{c}^{(T)}$ if and only if for each time t, W^t is generated according to a static SBM with respect to $\mathbf{c}^t$.

This definition of a dynamic SBM encompasses dynamic extensions of SBMs previously proposed in the literature (Yang et al., 2011; Xu and Hero III, 2014), which model the sequence $W^{(T)}$ as observations from a hidden Markov-type model, where W^t is conditionally independent of all past adjacency matrices $W^{(t-1)}$ given the current SBM parameters Φ^t. We refer to these hidden Markov SBMs as HM-SBMs.

Yang et al. (2011) proposed an HM-SBM that posits a Markov model on the class membership vectors $\mathbf{c}^t$ parameterized by a transition matrix that specifies the probability that any node in class a at time t switches to class b at time $t + 1$ for all a, b, t. The authors found that approximate inference using variation expectation-maximization (EM) resulted in poor fits due to getting trapped in local maxima and proposed instead to use a combination of Gibbs sampling and simulated annealing, which they refer to as probabilistic simulated annealing (PSA).

Xu and Hero III (2014) proposed an HM-SBM that places a state-space model on the block probability matrices Θ^t. The temporal evolution of these probabilities is governed by a linear dynamic system on the logits of the probabilities $\Psi^t = \log(\Theta^t / (1 - \Theta^t))$, where the logarithms are applied to each entry of the matrix. The authors performed approximate inference by

using an extended Kalman filter augmented with a local search procedure, which was shown to perform competitively with the PSA procedure of Yang et al. (2011) in terms of accuracy but is about an order of magnitude faster.

4   Stochastic Block Transition Models

4.1   Motivation

One of the main disadvantages of using a hidden Markov-type approach for dynamic SBMs relates to the assumption that edges at time t are conditionally independent from edges at previous times given the SBM parameters (states) at time t. Hence the probability distribution of edge durations is given by

$$\Pr(\text{duration} = d) = \left(1 - \theta_{ab}^{t-1}\right)\theta_{ab}^t \cdots \theta_{ab}^{t+d-1}\left(1 - \theta_{ab}^{t+d}\right),$$

where the edge was first present at time t, and the nodes belong to classes a and b for the entire duration of the edge. Note that the edge durations are tied directly to the probabilities of forming edges at a given time θ_{ab}^t, which controls the densities of the blocks. Specifically, the presence or lack of an edge between two nodes at any particular time *does not directly influence* the presence or lack of such an edge at a future time, which is undesirable in certain settings, as noted in Section 2.

A model where the edge durations are de-coupled from the block densities would allow for edges with long durations even in blocks with low densities. We propose such a model, namely the *stochastic block transition model* (SBTM). The main idea is as follows: for any pair of nodes $i \in a$ and $j \in b$ at both times $t-1$ and t such that $w_{ij}^{t-1} = 1$, i.e. there is an edge from i to j at time $t-1$, w_{ij}^t are independent and identically distributed (iid). The same is true for $w_{ij}^{t-1} = 0$. In short, all edges in a block at time $t-1$ are equally likely to re-appear at time $t-1$, and non-edges in a block at time $t-1$ are equally likely to appear at time t, hence the name stochastic block *transition* model.

4.2   Model Definition

To formalize the stochastic block transition model, we also need to specify how to handle nodes that were not present at a previous time step as well as nodes that change classes over time. Let i and j denote nodes in classes

a and b, respectively, at both times $t - 1$ and t, and define

$$\pi_{ab}^{t|0} = \Pr(w_{ij}^t = 1 | w_{ij}^{t-1} = 0) \tag{1}$$

$$\pi_{ab}^{t|1} = \Pr(w_{ij}^t = 1 | w_{ij}^{t-1} = 1). \tag{2}$$

Unlike in the hidden Markov dynamic SBM, where edges are formed iid with probabilities according to the block probability matrix Θ^t, in the SBTM, edges are formed according to two block transition matrices: $\Pi^{t|0} = [\pi_{ab}^{t|0}]$, denoting the probability of *forming new edges* within blocks, and $\Pi^{t|1} = [\pi_{ab}^{t|1}]$, denoting the probability of *existing edges recurring* within blocks. Let a' and b' denote the class memberships of nodes i and j at time $t - 1$. If a node was not present at time $t - 1$, take the class membership to be 0.

We formally define a stochastic block transition model as follows:

Definition 3. Let $W^{(T)}$ and $\mathbf{c}^{(T)}$ denote the same quantities as in Definition 2. $W^{(T)}$ is generated according to a *stochastic block transition model* with respect to $\mathbf{c}^{(T)}$ if and only if,

1. The initial adjacency matrix W^1 is generated according to a static SBM with respect to $\mathbf{c}^1$.

2. At any given time t, for any nodes $i \neq j$, the random variables w_{ij}^t are statistically independent.

3. At time $t \geq 2$, for any nodes $i \neq j$ such that $c_i^t = a$ and $c_j^t = b$ and for any $u \in \{0, 1\}$,

$$\Pr(w_{ij}^t = 1 | w_{ij}^{t-1} = u) = \xi_{ij}^t \pi_{ab}^{t|u}. \tag{3}$$

The matrix of scaling factors $\Xi^t = [\xi_{ij}^t]$ is used to scale the transition probabilities $\pi_{ab}^{t|0}$ and $\pi_{ab}^{t|1}$ to account for new nodes entering the network as well as existing nodes changing classes over time.

We propose to choose the scaling factors ξ_{ij}^t to satisfy the following properties:

1. If nodes $i \in a$ and $j \in b$ at both times $t - 1$ and t, then $\xi_{ij}^t = 1$.

2. The scaled transition probability is a valid probability, i.e. $0 \leq \xi_{ij}^t \pi_{ab}^{t|u} \leq 1$ for all $i \neq j$ such that $c_i^t = a$, $c_j^t = b$, and $u \in \{0, 1\}$.

3. The marginal distribution of the adjacency matrix W^t should follow a static SBM.

Property 1 follows from the definition of the transition probabilities (1) and (2). Property 2 ensures that the SBTM is a valid model. Finally, property 3 provides the connection to the static SBM.

4.3 Derivation of Scaling Factors

We now derive an expression for the scaling factor that satisfies each of the three properties. Consider two nodes $i' \in a$ and $j' \in b$ at both times $t - 1$ and t. From property 3, the marginal probability $\Pr(w_{i'j'}^t = 1)$ depends only on a and b, so we denote it by θ_{ab}^t, which can be expressed as

$$
\begin{aligned}
\theta_{ab}^t &= \Pr(w_{i'j'}^t = 1) \\
&= \Pr(w_{i'j'}^t = 1 | w_{i'j'}^{t-1} = 0) \Pr(w_{i'j'}^{t-1} = 0) \\
&\qquad + \Pr(w_{i'j'}^t = 1 | w_{i'j'}^{t-1} = 1) \Pr(w_{i'j'}^{t-1} = 1) \\
&= \pi_{ab}^{t|0}(1 - \theta_{ab}^{t-1}) + \pi_{ab}^{t|1} \theta_{ab}^{t-1},
\end{aligned}
\tag{4}
$$

where (4) follows from property 1.

Now consider the general case, where nodes $i \in a'$ and $j \in b'$ at time $t - 1$ and $i \in a$ and $j \in b$ at time t. We begin with the case where $a' = 0$ or $b' = 0$, indicating that either node i or j, respectively, was not present at time $t - 1$. For this case, $w_{ij}^{t-1} = 0$ so

$$
\Pr(w_{ij}^t = 1) = \Pr(w_{ij}^t = 1 | w_{ij}^{t-1} = 0) = \xi_{ij}^t \pi_{ab}^{t|0}
$$

Property 1 does not apply. In order for property 3 to hold, $\Pr(w_{ij}^t = 1)$ must be equal to θ_{ab}^t. Thus $\xi_{ij}^t = \theta_{ab}^t / \pi_{ab}^{t|0}$. Note that this also satisfies property 2 because θ_{ab}^t is a valid probability.

Next consider the case where $a', b' \neq 0$, i.e. both nodes were present at the previous time. Then

$$
\begin{aligned}
\Pr(w_{ij}^t &= 1) \\
&= \Pr(w_{ij}^t = 1 | w_{ij}^{t-1} = 0) \Pr(w_{ij}^{t-1} = 0) \\
&\qquad + \Pr(w_{ij}^t = 1 | w_{ij}^{t-1} = 1) \Pr(w_{ij}^{t-1} = 1) \quad (5) \\
&= \xi_{ij}^{t|0} \pi_{ab}^{t|0}(1 - \theta_{a'b'}^{t-1}) + \xi_{ij}^{t|1} \pi_{ab}^{t|1} \theta_{a'b'}^{t-1}, \quad (6)
\end{aligned}
$$

where (6) follows from substituting (3) into (5) and by letting the scaling factor

$$
\xi_{ij}^t = \begin{cases} \xi_{ij}^{t|0}, & \text{if } w_{ij}^{t-1} = 0 \\ \xi_{ij}^{t|1}, & \text{if } w_{ij}^{t-1} = 1 \end{cases}.
\tag{7}
$$

According to property 3, $\Pr(w_{ij}^t = 1) = \theta_{ab}^t$. Hence one must choose the scaling factor ξ_{ij}^t such that this is the case. The seemingly obvious solution is to compare coefficients of $\pi_{ab}^{t|0}$ and $\pi_{ab}^{t|1}$ between (4) and (6) but this solution does not satisfy property 2. Instead, we first identify a range of choices for the scaling factor ξ_{ij}^t that satisfy properties 2 and 3, then we select a particular choice that satisfies property 1.

Property 2 implies the following inequalities:

$$0 \leq \xi_{ij}^{t|0} \leq 1/\pi_{ab}^{t|0} \tag{8}$$

$$0 \leq \xi_{ij}^{t|1} \leq 1/\pi_{ab}^{t|1}. \tag{9}$$

Meanwhile, property 3 implies that

$$\theta_{ab}^t = \xi_{ij}^{t|0} \pi_{ab}^{t|0}(1 - \theta_{a'b'}^{t-1}) + \xi_{ij}^{t|1} \pi_{ab}^{t|1} \theta_{a'b'}^{t-1}. \tag{10}$$

Re-arranging (10) and substituting into (9), one obtains

$$\frac{\theta_{ab}^t - \theta_{a'b'}^{t-1}}{\pi_{ab}^{t|0}(1 - \theta_{a'b'}^{t-1})} \leq \xi_{ij}^{t|0} \leq \frac{\theta_{ab}^t}{\pi_{ab}^{t|0}(1 - \theta_{a'b'}^{t-1})}. \tag{11}$$

Combining (8), (10), and (11), one arrives at the necessary conditions on $\pi_{ab}^{t|0}$ in order to satisfy properties 2 and 3:

$$\alpha(a', b') \leq \xi_{ij}^{t|0} \leq \beta(a', b'), \tag{12}$$

where the upper and lower bounds are functions of a' and b', the classes for i and j, respectively, at time $t-1$ and are given by

$$\alpha(a', b') = \max\left(0, \frac{\theta_{ab}^t - \theta_{a'b'}^{t-1}}{\pi_{ab}^{t|0}(1 - \theta_{a'b'}^{t-1})}\right) \tag{13}$$

$$\beta(a', b') = \min\left(\frac{1}{\pi_{ab}^{t|0}}, \frac{\theta_{ab}^t}{\pi_{ab}^{t|0}(1 - \theta_{a'b'}^{t-1})}\right) \tag{14}$$

From (12)–(14), it follows that

$$\xi_{ij}^{t|0} = \alpha(a', b') + \frac{\beta(a', b') - \alpha(a', b')}{\gamma(a', b')} \tag{15}$$

is a valid solution for any $\gamma(a', b') \geq 1$.

In order to satisfy property 1 as well, $\xi_{i'j'}^{t|0}$ must be equal to 1 for any nodes $i' \in a$ and $j' \in b$ at time $t-1$. This can be accomplished by choosing

$$\gamma(a', b') = \frac{\beta(a,b) - \alpha(a,b)}{1 - \alpha(a,b)}. \tag{16}$$

Notice that the arguments in $\alpha(\cdot)$ and $\beta(\cdot)$ are the current classes a and b, regardless of the previous classes.

The assignment for $\xi_{ij}^{t|0}$ is thus obtained by substituting (16) into (15). This value can then be substituted into (10) to obtain the assignment for $\xi_{ij}^{t|1}$.

Proposition 1. *The scaling factor assignment given by* (10), (15), *and* (16) *satisfies the three properties specified in Section 4.2.*

Proof. Begin with property 1. Let $i' \in a$ and $j' \in b$ at both times $t-1$ and t. From (15) and (16),

$$\begin{aligned}
\xi_{i'j'}^{t|0} &= \alpha(a,b) + (\beta(a,b) - \alpha(a,b))/\gamma(a,b) \\
&= \alpha(a,b) + (1 - \alpha(a,b)) \\
&= 1.
\end{aligned} \tag{17}$$

Substituting (17) and (10) into (11),

$$\xi_{i'j'}^{t|1} = \frac{\pi_{ab}^{t|0}(\theta_{ab}^{t-1} - 1)}{\pi_{ab}^{t|1}\theta_{ab}^{t-1}} + \frac{\pi_{ab}^{t|0}(1 - \theta_{ab}^{t-1}) + \pi_{ab}^{t|1}\theta_{ab}^{t-1}}{\pi_{ab}^{t|1}\theta_{ab}^{t-1}} = 1.$$

Thus property 1 is satisfied.

From the derivation of the scaling factor assignment, it was shown that properties 2 and 3 are satisfied provided $\gamma(a', b') \geq 1$ for all (a', b'). From (16), this is true if and only if $\beta(a,b) \geq 1$ for all (a,b). From (14), $\beta(a,b) \geq 1/\pi_{ab}^{t|0} \geq 1$ because $\pi_{ab}^{t|0}$ is a probability and hence must be between 0 and 1, and

$$\beta(a,b) \geq \frac{\theta_{ab}^t}{\pi_{ab}^{t|0}(1 - \theta_{ab}^{t-1})} = 1 + \frac{\pi_{ab}^{t|1}\theta_{ab}^{t-1}}{\pi_{ab}^{t|0}(1 - \theta_{ab}^{t-1})} \geq 1,$$

where the equality follows from (4), and the final inequality results from $\pi_{ab}^{t|0}$, $\pi_{ab}^{t|1}$, and θ_{ab}^{t-1} all being probabilities and hence between 0 and 1. Thus properties 2 and 3 are also satisfied. $\square$

Proposition 2. *An SBTM with respect to* $\mathbf{c}^{(T)}$ *satisfying such an assumption is a dynamic SBM; that is, any sequence* $W^{(T)}$ *generated by the SBTM also satisfies the requirements of a dynamic SBM.*

Proposition 2 holds trivially from property 3, which is satisfied due to Proposition 1. Both the SBTM and HM-SBM are dynamic SBMs; the main difference between the two is that, under the SBTM, the presence or lack of an edge between two nodes at a particular time *does* affect the presence or lack of such an edge at a future time as indicated by (3).

4.4 State Dynamics

The SBTM, as defined in Definition 3, does not specify the model governing the dynamics of the sequence of adjacency matrices $W^{(T)}$ aside from the dependence of W^t on W^{t-1} specified in requirement 3. To complete the model, we use a linear dynamic system on the logits of the probabilities, similar to Xu and Hero III (2014). Unlike Xu and Hero III (2014), however, the states of the system would be the logits of the block transition matrices $\Pi^{t|0}$ and $\Pi^{t|1}$.

Let $\mathbf{x}$ denote the vectorized equivalent of a matrix X, obtained by stacking columns on top of one another, so that $\boldsymbol{\pi}^{t|0}$ and $\boldsymbol{\pi}^{t|1}$ are the vectorized equivalents of $\Pi^{t|0}$ and $\Pi^{t|1}$, respectively. The states of the system can then be expressed as a vector

$$\boldsymbol{\psi}^t = \begin{bmatrix} \log(\boldsymbol{\pi}^{t|0}/(1-\boldsymbol{\pi}^{t|0})) \\ \log(\boldsymbol{\pi}^{t|1}/(1-\boldsymbol{\pi}^{t|1})) \end{bmatrix}, \tag{18}$$

resulting in the dynamic linear system

$$\boldsymbol{\psi}^t = F^t \boldsymbol{\psi}^{t-1} + \mathbf{v}^t, \tag{19}$$

where F^t is the state transition model applied to the previous state, and $\mathbf{v}^t$ is a random vector of zero-mean Gaussian entries, commonly referred to as process noise, with covariance matrix Γ^t. Note that (19) is the same dynamic system equation as in Xu and Hero III (2014), only with a different definition (18) for the state vector.

5 Model Inference

5.1 Asymptotic Distribution of Observations

The inference procedure for the dynamic SBM of Xu and Hero III (2014) utilized a Central Limit Theorem (CLT) approximation for the block densities, which are scaled sums of independent, identically distributed Bernoulli random variables w_{ij}^t. Such an approach cannot be used for the SBTM because blocks no longer consist of identically distributed variables w_{ij}^t due to

the dependency between W^t and W^{t-1}. Furthermore, the presence of the scaling factors ξ_{ij}^t in the transition probabilities (3) ensure that w_{ij}^t are not identically distributed even after conditioning on w_{ij}^{t-1}.

We show, however, that the sample mean of a scaled version of the adjacencies, is asymptotically Gaussian. For $a, b \in \{1, \ldots, k\}$ and $u \in \{0, 1\}$, let

$$\mathcal{B}_{ab}^{t|u} = \{(i,j) : i \neq j, c_i^t = a, c_j^t = b, w_{ij}^{t-1} = u\}.$$

Note that $\mathcal{B}_{ab}^{t|0}$ denotes the set of non-edges in block (a, b) at time $t-1$, which is also the set of *possible new edges* at time t, and $\mathcal{B}_{ab}^{t|1}$ denotes the set of edges in block (a, b) at time $t-1$, which is also the set of *possible recurring edges* at time t. Let

$$m_{ab}^{t|u} = \sum_{(i,j) \in \mathcal{B}_{ab}^{t|u}} \frac{w_{ij}^t}{\xi_{ij}^t},$$

$$n_{ab}^{t|u} = |\mathcal{B}_{ab}^{t|u}|.$$

$m_{ab}^{t|0}$ and $m_{ab}^{t|1}$ denote the scaled number of new and existing edges, respectively, within block (a, b) at time t, while $n_{ab}^{t|0}$ and $n_{ab}^{t|1}$ denote the number of possible new and existing edges, respectively. The following theorem shows that the sample mean of the scaled adjacencies within $\mathcal{B}_{ab}^{t|u}$ is asymptotically Gaussian as the block size increases.

Theorem 1. *The sample mean of the scaled adjacencies*

$$\frac{m_{ab}^{t|u}}{n_{ab}^{t|u}} = \frac{1}{n_{ab}^t} \sum_{(i,j) \in \mathcal{B}_{ab}^{t|u}} \frac{w_{ij}^t}{\xi_{ij}^t} \to \mathcal{N}\left(\pi_{ab}^{t|u}, \left(\frac{s_{ab}^{t|u}}{n_{ab}^{t|u}}\right)^2\right)$$

in distribution as $n_{ab}^{t|u} \to \infty$, where

$$\left(s_{ab}^{t|u}\right)^2 = \sum_{(i,j) \in \mathcal{B}_{ab}^{t|u}} \mathrm{Var}_u\left(\frac{w_{ij}^t}{\xi_{ij}^t}\right)$$

$$= \pi_{ab}^{t|u} \sum_{(i,j) \in \mathcal{B}_{ab}^{t|u}} \frac{1}{\xi_{ij}^t} - n_{ab}^{t|u}\left(\pi_{ab}^{t|u}\right)^2, \qquad (20)$$

and $\mathrm{Var}_u(\cdot)$ denotes the condition variances $\mathrm{Var}(\cdot|w_{ij}^{t-1} = u)$.

12

Proof. The scaled adjacencies w_{ij}^t/ξ_{ij}^t are independent, but not identically distributed, so the classical CLT no longer applies. However, the Lyapunov CLT can be applied provided Lyapunov's condition is satisfied (Billingsley, 1995). In this setting, Lyapunov's condition specifies that for some $\delta > 0$,

$$\lim_{n_{ab}^{t|u} \to \infty} \frac{1}{\left(s_{ab}^{t|u}\right)^{2+\delta}} \sum_{(i,j) \in \mathcal{B}_{ab}^{t|u}} E_u \left[\left| \frac{w_{ij}^t}{\xi_{ij}^t} - \pi_{ab}^{t|u} \right|^{2+\delta} \right] = 0,$$

where $E_u[\cdot]$ denotes the conditional expectation $E[\cdot|w_{ij}^{t-1} = u]$.

We demonstrate that Lyapunov's condition is satisfied for $\delta = 2$. First note that, although there are an infinite number of terms in the summation (in the limit), there are a finite number of unique terms. Specifically $w_{ij}^t \in \{0,1\}$, and ξ_{ij}^t depends only on i, j through their current and previous class memberships a, b, a', and b', which are all in $\{0, 1, \ldots, k\}$. Hence

$$\frac{1}{\left(s_{ab}^{t|u}\right)^4} \sum_{(i,j) \in \mathcal{B}_{ab}^{t|u}} E_u \left[\left(\frac{w_{ij}^t}{\xi_{ij}^t} - \pi_{ab}^{t|u} \right)^4 \right] \tag{21}$$

$$\leq \frac{n_{ab}^{t|u}}{\left(s_{ab}^{t|u}\right)^4} \max_{(i,j) \in \mathcal{B}_{ab}^{t|u}} E_u \left[\left(\frac{w_{ij}^t}{\xi_{ij}^t} - \pi_{ab}^{t|u} \right)^4 \right]$$

$$= \frac{1}{O\left(n_{ab}^{t|u}\right)},$$

where the last equality follows from (20). Thus (21) approaches 0 as $n_{ab}^{t|u} \to \infty$, and Lyapunov's condition is satisfied. The Lyapunov CLT states that

$$\frac{1}{s_{ab}^{t|u}} \sum_{(i,j) \in \mathcal{B}_{ab}^{t|u}} \left(\frac{w_{ij}^t}{\xi_{ij}^t} - \pi_{ab}^{t|u} \right) \xrightarrow{d} \mathcal{N}(0,1)$$

where $\xrightarrow{d}$ denotes convergence in distribution. By rearranging terms one obtains the desired result. $\qquad\square$

5.2   State-space Model Formulation

Theorem 1 shows that the sample means $m_{ab}^{t|u}/n_{ab}^{t|u}$ are asymptotically Gaussian. Assume that they are indeed Gaussian. Stack these entries to form

13

the observation vector

$$
\begin{aligned}
\mathbf{y}^t &= \left[\frac{m_{11}^{t|0}}{n_{11}^{t|0}} \cdots \frac{m_{kk}^{t|0}}{n_{kk}^{t|0}} \frac{m_{11}^{t|1}}{n_{11}^{t|1}} \cdots \frac{m_{kk}^{t|1}}{n_{kk}^{t|1}} \right]^T \\
&= h\left(\boldsymbol{\psi}^t\right) + \mathbf{z}^t,
\end{aligned}
\tag{22}
$$

where the function $h : \mathbb{R}^{2k^2} \to \mathbb{R}^{2k^2}$ is defined by

$$
h_i(\mathbf{x}) = 1/(1 + e^{-x_i}),
\tag{23}
$$

i.e. the logistic sigmoid applied to each entry of $\mathbf{x}$, $\boldsymbol{\psi}^t$ was defined in (18), and $\mathbf{z}^t \sim \mathcal{N}(\mathbf{0}, \Sigma^t)$, where Σ^t is a diagonal matrix with entries given by $\left(s_{ab}^{t|u}/n_{ab}^{t|u}\right)^2$.

The equations (19) and (22) form a non-linear dynamic system with zero-mean Gaussian observation and process noise terms $\mathbf{z}^t$ and $\mathbf{v}^t$, respectively, with the only non-linearity due to the logistic function $h(\cdot)$. Assume that the initial state is also Gaussian distributed, i.e. $\boldsymbol{\psi}^1 \sim \mathcal{N}\left(\boldsymbol{\mu}^1, \Gamma^1\right)$, and that $\{\boldsymbol{\psi}^1, \mathbf{v}^2, \dots, \mathbf{v}^t, \mathbf{z}^2, \dots, \mathbf{z}^t\}$ are mutually independent. If (22) was linear, then the optimal estimate for $\boldsymbol{\psi}^t$ (given observations $\mathbf{y}^{(T)}$) in terms of minimum mean-squared error and maximum a posteriori probability (MAP) would be given by the Kalman filter. Due to the non-linearity, we apply the extended Kalman filter (EKF), which linearizes the dynamics about the predicted state and results in a *near-optimal* estimate (in the MAP sense) when the estimation errors are small enough to make the linearization accurate. The EKF was used for inference in systems of the form of (19) and (22) in Xu and Hero III (2014).

5.3 Inference Procedure

Once the vector of sample means $\mathbf{y}^t$ is obtained, a near-optimal estimate of the state vector $\boldsymbol{\psi}^t$ can be obtained using the EKF. Recall that $\boldsymbol{\psi}^t$ contains the logits of the probabilities of forming new edges $\boldsymbol{\pi}^{t|0}$ and the probabilities of existing edges re-occurring $\boldsymbol{\pi}^{t|1}$. In order to compute the sample means $\mathbf{y}^t$, however, one needs to first estimate the following quantities:

1. The unknown hyperparameters $(\boldsymbol{\psi}^0, \Gamma^0, \Sigma^t, \Gamma^t)$ of the state-space model (19) and (22).

2. The vector of class memberships $\mathbf{c}^t$.

3. The matrix of scaling factors Ξ^t.

Algorithm 1 SBTM inference procedure

At time step 1:

1: Initialize estimated class assignment using spectral clustering on W^1
2: Compute ML estimates $\hat{\mathbf{c}}^1$ and $\hat{\Theta}^1$ by local search
3: Compute predicted state vector $\hat{\boldsymbol{\psi}^{2|1}}$ at time step 2 using EKF predict phase

At time step $t > 1$:

1: Initialize estimated class assignment $\hat{\mathbf{c}}^t \leftarrow \hat{\mathbf{c}}^{t-1}$
2: **repeat** {Local search (hill climbing) algorithm}
3:   **for all** neighboring class assignments **do**
4:     Compute plug-in estimate $\hat{\Xi}^t$ of scaling matrix using $\hat{\Theta}^{t-1}$, EKF predicted state $\hat{\boldsymbol{\psi}}^{t|t-1}$, and current class assignment
5:     Compute plug-in estimate $\hat{\mathbf{y}}^t$ of sample means using $\hat{\Xi}^t$, W^t, and current class assignment
6:     Compute estimate $\hat{\boldsymbol{\psi}}^{t|t}$ of state vector using EKF update phase
7: **until** reached local maximum of posterior density
8: Compute predicted state vector $\boldsymbol{\psi}^{\hat{t+1}|t}$ at time step $t+1$ using EKF predict phase

---

Methods for estimating items 1 and 2 are discussed in Xu and Hero III (2014). Item 1 can be addressed using standard methods for state-space models, typically alternating between state and hyperparameter estimation (Nelson, 2000). Item 2 is handled by alternating between a local search (hill climbing) algorithm to estimate class memberships and the EKF to estimate edge transition probabilities.

The main difference between the inference procedures of the HM-SBM and the SBTM proposed in this paper involves item 3. The matrix of scaling factors Ξ^t is a function of the marginal edge probabilities at the current and previous times (Θ^t and Θ^{t-1}, respectively) as well as the current probabilities of new and existing edges ($\Pi^{t|0}$ and $\Pi^{t|1}$, respectively). Θ^t can be computed from the other three quantities from (4).

We propose to use plug-in estimates of Θ^{t-1}, $\Pi^{t|0}$, and $\Pi^{t|1}$ to estimate the scaling matrix Ξ^t. Recall from (18) that the state vector $\boldsymbol{\psi}^t$ consists of the logits of the new and existing edge probabilities. Hence we obtain

$$\left[\hat{\boldsymbol{\pi}}^{t|0} \ \ \hat{\boldsymbol{\pi}}^{t|1} \right]^T = h \left(\hat{\boldsymbol{\psi}}^{t|t-1} \right),$$

where $\hat{\boldsymbol{\psi}}^{t|t-1}$ is the EKF prediction of the state vector at time t. The recursion is initialized at $t = 2$ using the maximum-likelihood (ML) estimate $\hat{\Theta}^1$ obtained from W^1. The spectral clustering procedure of Sussman et al. (2012) can be used to initialize the class assignments. A sketch of the entire inference procedure is shown in Algorithm 1.

6 Experiments

6.1 Simulated Networks

In this experiment we generate synthetic networks in a manner similar to a simulation experiment in Yang et al. (2011) and Xu and Hero III (2014), except with the stochastic block transition model rather than the hidden Markov stochastic block model. The network consists of 128 nodes initially split into 4 classes of 32 nodes each. The edge probabilities for blocks at the initial time step are chosen to be $\theta_{aa}^1 = 0.2580$ and $\theta_{ab}^1 = 0.0834$ for $a, b = 1, 2, 3, 4; a \neq b$. The mean $\boldsymbol{\mu}^1$ is chosen such that $\pi_{aa}^{1|0} = 0.1$, $\pi_{ab}^{1|0} = 0.05, a \neq b$, $\pi_{aa}^{1|1} = 0.7$, and $\pi_{ab}^{1|1} = 0.45, a \neq b$. The covariance Γ^1 for the initial state is chosen to be a scaled identity matrix $0.04I$, respectively. The state vector $\boldsymbol{\psi}^t$ evolves according to a Gaussian random walk model on $\boldsymbol{\psi}^t$, i.e. $F^t = I$ in (18). Γ^t is constructed such that $\gamma_{ii}^t = 0.01$ and $\gamma_{ij}^t = 0.0025$ for $i \neq j$. 10 time steps are generated, and at each time step, 10% of the nodes are randomly selected to leave their class and are randomly assigned to one of the other three classes. For consistency with Yang et al. (2011) and Xu and Hero III (2014), we generate *undirected* graph snapshots in this experiment.

We check the validity of the asymptotic Gaussian distribution of the scaled sample means $\mathbf{y}^t$. In this simulation experiment, the true means and standard deviations for $\mathbf{y}^t$ are known and are used to standardize $\mathbf{y}^t$. Q-Q plots for the standardized $\mathbf{y}^t$ are shown in Figure 1. Figure 1a shows the distribution of $\mathbf{y}^t$ when both the true classes and true scaling factors (calculated using the true states) are used. Notice that the empirical distribution is close to the asymptotic Gaussian distribution, with slightly heavier tail. Experimentally we find that this deviation decreases as the block sizes increase, as one would expect from Theorem 1.

Figure 1b shows that the distribution of $\mathbf{y}^t$ is roughly the same when using estimated scaling factors along with the true classes, which is an encouraging result and suggests that the EKF-based inference procedure would likely work well in the a priori block model setting. Figure 1c shows that

(a) True classes and scaling

(b) True classes, estimated scaling
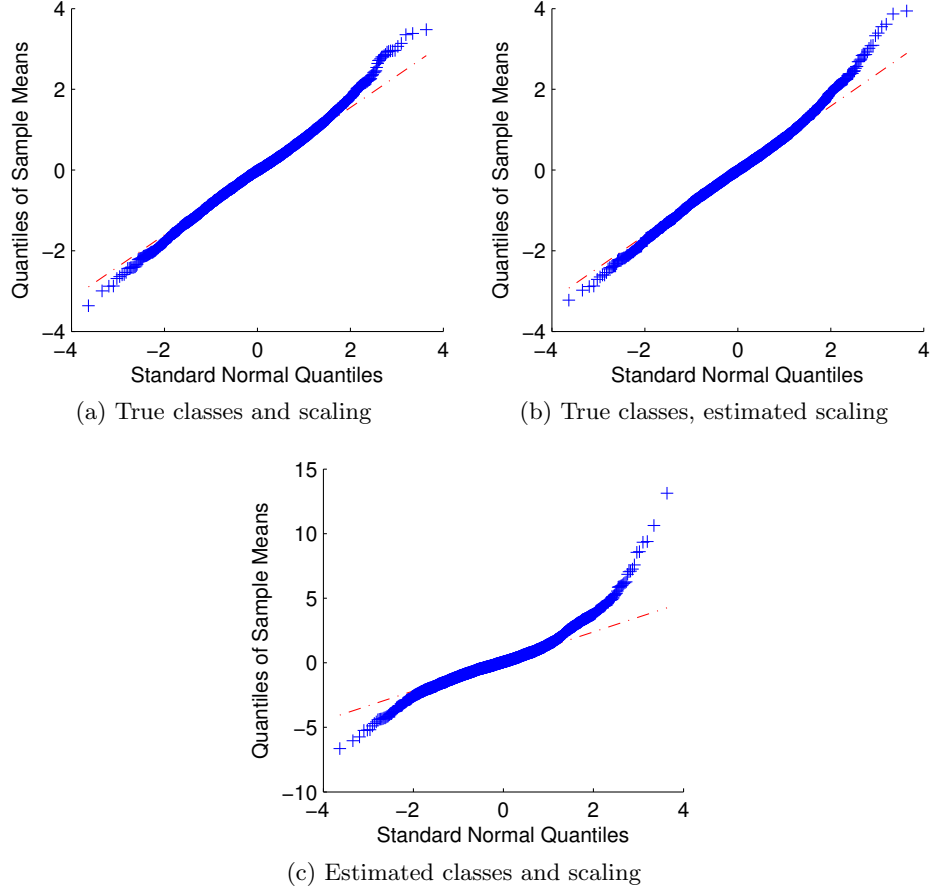
(c) Estimated classes and scaling

Figure 1: Q-Q plots of standardized sample means $\hat{\mathbf{y}}^t$ in the simulation experiment under three levels of estimation. With (a) true classes and scaling factors, $\hat{\mathbf{y}}^t$ is close to the asymptotic Gaussian distribution predicted by Theorem 1. Even with (b) estimated scaling factors, $\hat{\mathbf{y}}^t$ is still close to the asymptotic Gaussian distribution. When (c) class memberships are also estimated, $\hat{\mathbf{y}}^t$ is heavier tailed due to the errors in the estimated classes.
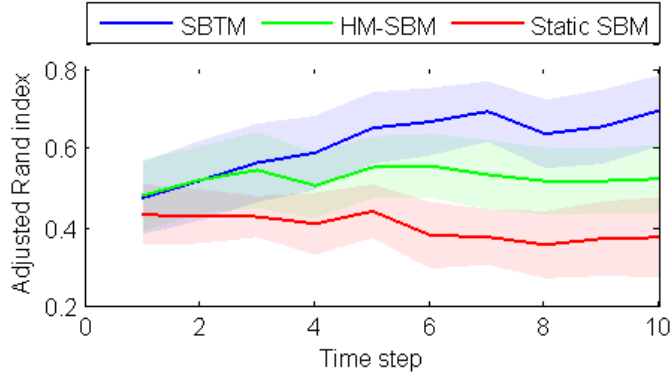
Figure 2: Adjusted Rand indices with 95% confidence bands for three inference algorithms on the simulated networks.

the distribution of $\mathbf{y}^t$ when using both estimated scaling factors and classes is significantly more heavy-tailed. Since this is not seen in Figure 1b, we conclude that it is due to errors in the class estimation, which causes the distribution of $\mathbf{y}^t$ to deviate from the asymptotically Gaussian distribution when using true classes. The heavier tails suggest that perhaps a more robust filter, such as a filter that assumes Student-t distributed observations, may provide more accurate estimates in the a posteriori setting.

Figure 2 shows a comparison of the class estimation accuracies of three different inference algorithms: the EKF-based algorithm for the SBTM proposed in this paper, the EKF algorithm for the HM-SBM (Xu and Hero III, 2014), and a static SBM fit using spectral clustering refined with a local search to reach a local maximum of the likelihood for each snapshot. As one might expect, the static SBM approach does not improve as more time snapshots are provided. The poorer performance of the HM-SBM approach compared to the proposed SBTM approach is also not a surprise since the dynamics on the marginal block probabilities no longer follow a dynamic linear system as assumed by Xu and Hero III (2014). The SBTM approach is more accurate than the other two; however it still makes enough mistakes to cause the heavier-tailed distribution of $\mathbf{y}^t$ as previously discussed.

6.2 Facebook Wall Posts

We now test the proposed SBTM inference algorithm on a real data set, namely a dynamic social network of Facebook wall posts (Viswanath et al., 2009). Similar to the analysis by Viswanath et al. (2009), we use 90-day time

steps from the start of the data trace in June 2006, with the final complete 90-day interval ending in November 2008, resulting in 9 total time steps. We filter out people who were active for less than 7 of the 9 times as well as those who posted to less than 20 other people's walls, leaving 522 remaining people.

We fit the SBTM to this dynamic network using Algorithm 1, beginning with a spectral clustering initialization at the first time step. From examination of the singular values of the first snapshot, we choose a fit with $k = 4$ classes. Three dense and one loosely-connected community are observed at all time steps, with only slight changes in community structure over time. Visualizations of the class structure overlaid onto the adjacency matrices at several time steps are included in the supplementary material. The initial snapshot contains only 381 active nodes, so most changes in class assignments are due to new nodes entering the network over time. The networks are very sparse, with the densest block having marginal edge probability of 0.05. We find that the probabilities of forming new edges do not vary significantly between blocks, ranging from about 0.05 to 0.07. The probabilities of existing edges re-occurring show greater variation between blocks, ranging from 0.25 to 0.55.

Adjacency matrices at three time steps are shown in Figure 3. The estimated class structure at the final time step is overlaid onto the adjacency matrices. Notice that all of the classes are actually communities, with denser diagonal blocks compared to off-diagonal blocks. Only slight changes in the class structure are observed over time.

A histogram of the edge durations observed in the network is shown in Figure 4a. Notice that, despite the low density of the blocks, more than 20% of the edges appear over multiple time steps. We generate 10 synthetic networks each from the HM-SBM and SBTM fits to the observed networks. The histogram of edge durations from synthetic networks generated from the HM-SBM is shown in Figure 4b. Due to the hidden Markov assumption, only the densities of the blocks are being replicated over time, and as such, the majority of edges are not repeated at the following time step. Compare this to the edge durations generated from the proposed SBTM, shown in Figure 4c. Notice that a significant fraction of edges are indeed repeated in these synthetic networks, much like in the observed networks. These edge durations cannot be replicated by hidden Markov models for the snapshots, demonstrating the importance of the SBTM.

Notice also that the edge durations from the synthetic networks are actually slightly longer than from the observed networks. This is an artifact that appears because not all nodes are active at all time steps in the observed
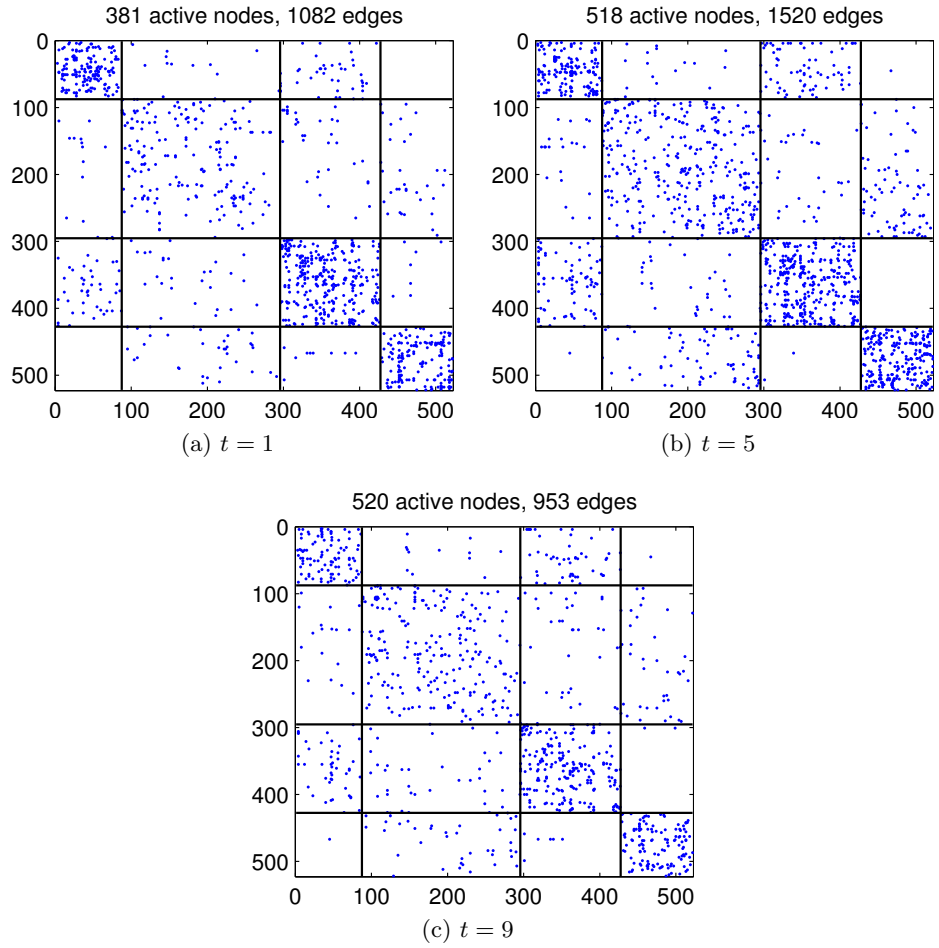
19

Figure 3: Adjacency matrices at three different time steps constructed from Facebook wall posts. The estimated classes at the final time $t = 9$ are overlaid onto the adjacency matrices. Three dense and one loosely-connected community are observed at all time steps, with only slight changes in community structure over time.

(a) Observed network

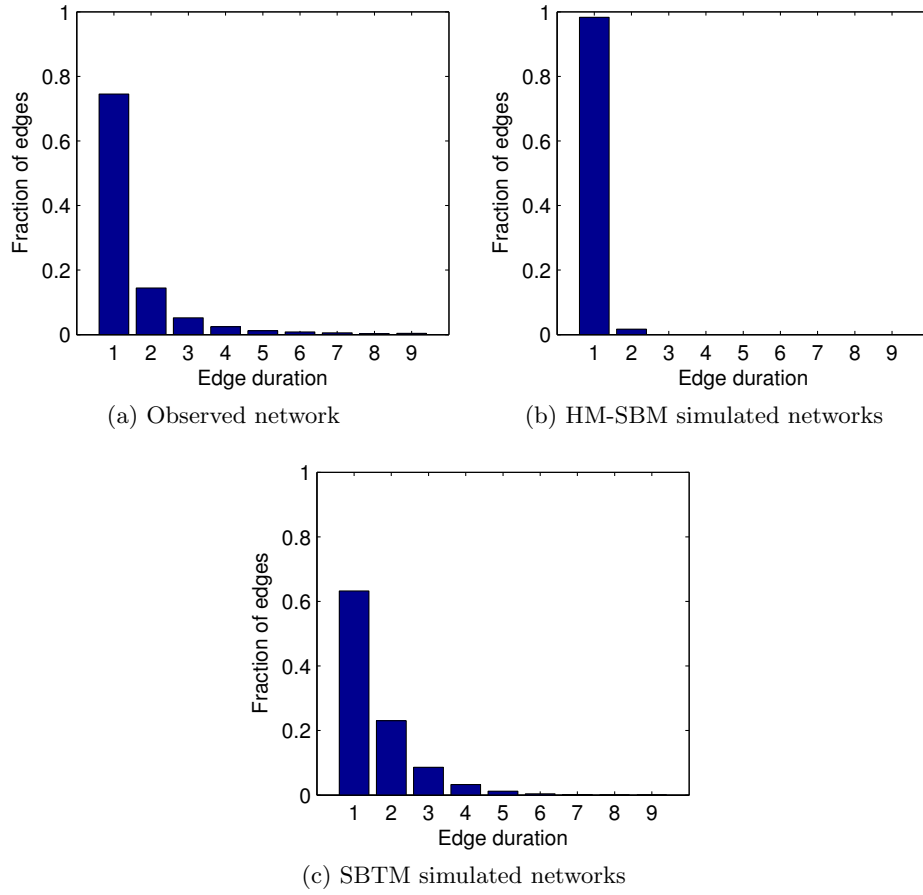(b) HM-SBM simulated networks

(c) SBTM simulated networks

Figure 4: Histograms of edge durations in (a) observed Facebook network, (b) simulated networks from HM-SBM fit to observed network, and (c) simulated networks from SBTM fit to observed network.

networks, causing edge durations to be shortened. One could perhaps replicate this effect by adding a layer to the dynamic model simulating nodes entering and leaving the network over time, which would be an interesting direction for future work.

The proposed SBTM can also be extended to have edges depend directly on whether edges were present further back than just the previous time step. Such an approach would likely improve forecasting ability; however, it also increases the number of states that need to be estimated, which creates further challenges.

References

P. Billingsley. *Probability and measure*. Wiley-Interscience, 3rd edition, 1995.

D. Durante and D. B. Dunson. Nonparametric Bayes dynamic modelling of relational data. *Biometrika*, In press, 2014.

J. R. Foulds, C. DuBois, A. U. Asuncion, C. T. Butts, and P. Smyth. A dynamic relational infinite feature model for longitudinal social networks. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, pages 287–295, 2011.

A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi. A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233, 2009.

F. Guo, S. Hanneke, W. Fu, and E. P. Xing. Recovering temporally rewiring networks: A model-based approach. In *Proceedings of the 24th International Conference on Machine Learning*, pages 321–328, 2007.

C. Heaukulani and Z. Ghahramani. Dynamic probabilistic models for latent feature propagation in social networks. In *Proceedings of the 30th International Conference on Machine Learning*, pages 275–283, 2013.

Q. Ho, L. Song, and E. P. Xing. Evolving cluster mixed-membership blockmodel for time-varying networks. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, pages 342–350, 2011.

P. D. Hoff. Hierarchical multilinear models for multiway data. *Computational Statistics and Data Analysis*, 55(1):530–543, 2011.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.

K. Ishiguro, T. Iwata, N. Ueda, and J. B. Tenenbaum. Dynamic infinite relational model for time-varying relational data analysis. In *Advances in Neural Information Processing Systems 23*, pages 919–927, 2010.

B. Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.

M. Kim and J. Leskovec. Nonparametric multi-group membership model for dynamic networks. In *Advances in Neural Information Processing Systems 25*, 2013.

N. H. Lee and C. E. Priebe. A latent process model for time series of attributed random graphs. *Statistical Inference for Stochastic Processes*, 14(3):231–253, 2011.

A. T. Nelson. *Nonlinear estimation and modeling of noisy time-series by dual Kalman filtering methods*. PhD thesis, Oregon Graduate Institute of Science and Technology, 2000.

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

K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.

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

P. Sarkar, S. M. Siddiq, and G. J. Gordon. A latent space approach to dynamic embedding of co-occurrence data. In *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics*, 2007.

D. L. Sussman, M. Tang, D. E. Fishkind, and C. E. Priebe. A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association*, 107(499):1119–1128, 2012.

B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the evolution of user interaction in Facebook. In *Proceedings of the 2nd ACM Workshop on Online Social Networks*, pages 37–42, New York, New York, USA, 2009.

E. P. Xing, W. Fu, and L. Song. A state-space mixed membership block-model for dynamic network tomography. *The Annals of Applied Statistics*, 4(2):535–566, 2010.

K. S. Xu and A. O. Hero III. Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):552–562, 2014.

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

Y. Zhao, E. Levina, and J. Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics*, 40(4):2266–2292, 2012.