Consistent estimation of dynamic and multi-layer block models

Qiuyi Han   QIUYIHAN@FAS.HARVARD.EDU
Department of Statistics, Harvard University, Cambridge, MA, USA

Kevin S. Xu   KEVINXU@OUTLOOK.COM
Technicolor Research, Los Altos, CA, USA

Edoardo M. Airoldi   AIROLDI@FAS.HARVARD.EDU
Department of Statistics, Harvard University, Cambridge, MA, USA

Abstract

Significant progress has been made recently on theoretical analysis of estimators for the stochastic block model (SBM). In this paper, we consider the *multi-graph* SBM, which serves as a foundation for many application settings including dynamic and multi-layer networks. We explore the asymptotic properties of two estimators for the multi-graph SBM, namely spectral clustering and the maximum-likelihood estimate (MLE), as the number of layers of the multi-graph increases. We derive sufficient conditions for *consistency* of both estimators and propose a variational approximation to the MLE that is computationally feasible for large networks. We verify the sufficient conditions via simulation and demonstrate that they are practical. In addition, we apply the model to two real data sets: a dynamic social network and a multi-layer social network with several types of relations.

1. Introduction

Modeling relational data arising from networks including social, biological, and information networks has received much attention recently. Various probabilistic models for networks have been proposed, including stochastic block models and their mixed-membership variants (Airoldi et al., 2008; Goldenberg et al., 2010). However, in many settings, we not only have a single network, but a collection of networks over a common set of nodes, which is often referred to as a *multi-graph*. Multi-graphs arise in several types of settings including dynamic networks with

time-evolving edges, such as time-stamped social networks of interactions between people, and multi-layer networks, where edges are measured in multiple ways such as phone calls, text messages, e-mails, face-to-face contacts, etc.

A significant challenge with multi-graphs is to extract common information across the *layers* of the multi-graph in a concise representation, yet be flexible enough to allow differences across layers. Motivated by the above examples, we consider the *multi-graph stochastic block model* first proposed by Holland et al. (1983), which divides nodes into classes that define blocks in the multi-graph. The key assumption is that nodes share the same block structure over the multiple layers, but the class connection probabilities may vary across layers. We believe this model is a flexible and principled way of analyzing multi-graphs and provides a strong foundation for many applications. The special case of a single layer, often referred to simply as the stochastic block model (SBM), has been studied extensively in recent years (Bickel and Chen, 2009; Rohe et al., 2011; Choi et al., 2012; Celisse et al., 2012; Jin, 2012; Bickel et al., 2013; Amini et al., 2013). However, the more general multi-graph case has not been studied as much.

In this paper, we explore the asymptotic properties of several estimators for the multi-graph SBM by letting the number of network layers grow while keeping the number of nodes *fixed*. We prove that a spectral clustering estimate of the class memberships is consistent for a special case of the model (Section 4.1). Next we derive sufficient conditions under which the maximum-likelihood estimate (MLE) of the class memberships is consistent in the general case (Section 4.2). Finally we propose a variational approximation to the MLE that is computationally tractable and is applicable to many multi-graph settings including dynamic and multi-layer networks (Section 4.2.1). We apply the spectral and variational approximation methods to several simulated and real data sets, including both a dynamic and a multi-relational social network (Section 5).

Our main contribution is the consistency analysis for the MLE, which ensures the tractability of the model and paves the way for more sophisticated models and inference techniques. To the best of our knowledge, we provide the *first* theoretical results for the multi-graph SBM for a growing number of layers.

2. Related work

Probabilistic models for networks have been studied for several decades; many commonly used models are discussed in the survey by Goldenberg et al. (2010). More recent work includes non-parametric network models using graphons (Airoldi et al., 2013; Wolfe and Olhede, 2013; Gao et al., 2014). Most previous models assume that a single network, rather than a multi-graph, is observed.

Two settings where multi-graphs arise include dynamic and multi-layer networks. Dynamic network models typically assume that a sequence of network snapshots is observed at discrete time steps. Previous work on dynamic network models has built upon models for a single network augmented with Markovian dynamics. Ahmed and Xing (2009); Hanneke et al. (2010) built upon exponential random graph models. Ishiguro et al. (2010); Yang et al. (2011); Ho et al. (2011); Xu and Hero (2014); Xu (2015) built on stochastic block models. Sarkar and Moore (2005); Sarkar et al. (2007); Durante and Dunson (2014) used latent space models. Foulds et al. (2011); Heaukulani and Ghahramani (2013); Kim and Leskovec (2013) used latent feature models.

Multi-layer networks consider multiple types of connections simultaneously. For example, Facebook users interact by using "likes", comments, messages, and other means. Multi-layer networks go by many other names like multi-relational, multi-dimensional, multi-view, and multiplex networks. The analysis of multi-layer networks has a long history (Holland et al., 1983; Fienberg et al., 1985; Szell et al., 2010; Mucha et al., 2010; Magnani and Rossi, 2011; Oselio et al., 2014). However there has not been much work on probabilistic modeling of such networks, aside from the multi-view latent space model proposed by Salter-Townshend and McCormick (2013), which couples the latent spaces of the multiple layers.

A third related setting involves modeling populations of networks, where each observation consists of a network snapshot drawn from a probability mass function over a network-valued sample space. Durante et al. (2014) proposed a nonparametric Bayesian model for this setting. This setting differs from the multi-graph setting that we consider in this paper because the network snapshots (layers) are drawn in an independent and identically distributed (iid) fashion, with no coupling between the snapshots.

The statistical properties of the inference algorithms in both dynamic and multi-layer network models have not typically been studied. Recently there has been a lot of progress on consistency analysis for single networks. Maximum-likelihood estimation, its variational approximation, and spectral clustering have all been proven to be consistent under the stochastic block model (Bickel and Chen, 2009; Rohe et al., 2011; Choi et al., 2012; Celisse et al., 2012; Zhao et al., 2012; Jin, 2012; Bickel et al., 2013; Lei and Rinaldo, 2014; Yang et al., 2014) as the number of nodes $N \to \infty$. Intuitively, for each new node added to the graph, we observe N realizations, hence larger N provides more information leading to consistent estimation of the model.

We extend the ideas used in single networks to multi-graphs. We note that the asymptotic regime is different in this case. For a single network, one typically lets $N \to \infty$, while for multi-graphs, we let $T \to \infty$ with N *fixed*. Intuitively it means we do not need to observe a very large network to get a correct understanding of the structure. Instead, we can gain the information through multiple samples, which may represent, for example, multiple observations over time or multiple relationships. In practice, it may be more realistic to allow N to grow along with T, particularly for the dynamic network setting. Allowing N to grow provides *more* information; thus our analysis with fixed N serves as a conservative analysis for different settings.

3. Multi-graph stochastic block model

We present an overview of the *multi-graph stochastic block model* first proposed by Holland et al. (1983). A single relation is represented by an adjacency matrix $G^t = (G_{ij}^t)$, $i, j = 1, \ldots, N$. We focus on symmetric binary relations with no self-edges. For a multi-graph, we observe an *adjacency array* $\vec{G} = \{G^1, G^2, \ldots, G^T\}$ sharing the same set of nodes. Subscripts denote the same node pairs for any t, while the superscript t indexes layers of the multi-graph. A layer may refer to time or type of relation depending on the application. If $\vec{G}$ is a random adjacency array for N nodes and T relations, then the probability distribution of $\vec{G}$ is called a *stochastic multi-graph*. Let the edge G_{ij}^t be a Bernoulli random variable with success probability Φ_{ij}^t. $\Phi^t = (\Phi_{ij}^t) \in [0, 1]_{N \times N}$ is the probability matrix of graph G^t. Let $\vec{\Phi} = \{\Phi^1, \Phi^2 \ldots \Phi^T\}$ be the *probability array*. We assume the independence of edges within and across layers conditioned on the probability array. That is, the adjacency array is generated according to

$$G_{ij}^t | \vec{\Phi} \overset{\text{ind}}{\sim} \text{Bern}(\Phi_{ij}^t).$$

The multi-graph stochastic block model is a special case of a stochastic multi-graph. In the multi-graph SBM, networks are generated in the following manner. First

each node is assigned to a class with probability $\pi = \{\pi_1, \ldots, \pi_K\}$ where π_k is the probability for a node to be assigned to class k. Then, given that nodes i and j are in classes k and l, respectively, an edge between i and j in network layer t is generated with probability P_{kl}^t. In other words, nodes in the same classes in the same layer have the same connection probability governed by $\vec{P} = \{P^1, P^2, \ldots, P^T\} \in [0, 1]_{K \times K}$, the *class connection probability array*. Let $c_i \in \{1, \ldots, K\}$ denote the class label of node i. Then $\Phi_{ij}^t = P_{c_i c_j}^t$.

The nodes have class labels $\vec{c}$ *shared* by all of the layers of the multi-graph, but in each layer the class connection probabilities P_{kl}^t may be different. As we consider undirected networks, P^t is a symmetric matrix with $K(K + 1)/2$ free parameters. One can see that the (single network) SBM is a special case of the multi-graph SBM with $T = 1$.

Though simple, this multi-graph model has not been formally studied in the a posteriori setting where class labels are estimated. It serves as a basis for many settings including dynamic networks and networks with multiple relations. More importantly, it can be theoretically analyzed and can provide insight on more complex models.

4. Consistent estimation for the multi-graph stochastic block model

Holland et al. (1983) only discussed estimation of the multi-graph SBM with blocks specified a priori. The sample proportion of each layer t is the maximum-likelihood estimate (MLE) of the class probability matrix P^t. However, in most applications, the block structure is unknown. Hence our main goal is to accurately estimate the class memberships. We extend several inference techniques used for the single network SBM to the multi-layer case.

It is not immediately straightforward how we can utilize inference techniques designed for the single network SBM. One may imagine inferring $\vec{c}$ independently from each network and averaging across them, e.g. by majority voting. That is, each node is assigned the class label that occurs most often. We find in simulations that this ad-hoc method often does not work well.

We propose spectral clustering on the mean graph as a motivating method for a special case of the model. Then we discuss maximum-likelihood estimation, a natural way to combine the information contained in the different layers, for the general case. Maximum-likelihood estimation is intractable for large networks so we also consider a variational approximation to the MLE.

Our main focus is on the consistency properties of these methods. We consider a fixed number of nodes N but let

the number of graph layers $T \to \infty$. In reality, although we do not have infinite layers, we often encounter situations with a large number of layers, such as dynamic networks over long periods of time.

4.1. Consistency of spectral clustering

Spectral clustering is a popular choice for estimating the block structure of the SBM because it scales to large networks and has shown to be consistent as $N \to \infty$ (Sussman et al., 2012). The method is based on singular value decomposition and K-means clustering on the singular vectors.

A natural way to extend spectral clustering from single networks to multi-graphs is to apply spectral clustering on the mean graph $\bar{G} = \frac{1}{T} \sum_{t=1}^{T} G^t$. This method is intuitively appealing as it matches with the assumption of a single set of class labels shared by all of the layers. We show that under some stationarity and ergodicity conditions, it indeed provides a consistent estimate of the class assignments. Specifically we consider the case where the class connection probabilities P^t vary across layers but have the same mean M. The following theorem shows the consistency of spectral clustering on the mean graph $\bar{G}$ if the mean M is identifiable.

Theorem 1. *Assume $\vec{P}$ follows a stationary ergodic process such that $E(P_{kl}^t) = \mu_{kl}$ and $Var(P_{kl}^t) = \epsilon_{kl}^2$ for all t. Assume $M = [\mu_{kl}]$ is identifiable, i.e. M has no identical rows. Let $\bar{G} = \frac{1}{T} \sum_{t=1}^{T} G^t$. Spectral clustering of $\bar{G}$ gives accurate labels as $T \to \infty$. That is, let $U_{N \times K}$ be the first K right singular vectors in the singular value decomposition of $\bar{G}$. K-means clustering on the rows of $U_{N \times K}$ outputs class estimates $\hat{c}_1, \ldots, \hat{c}_N$. Up to permutation, $\hat{c} = c$, a.s. as $T \to \infty$.*

We provide a sketch of the proof; details can be found in a longer version posted on arXiv (Han et al., 2015). Since we have independent errors in the probability matrix and also independent errors in the Bernoulli observations, averaging cancels the error so that $\bar{G} \to CMC'$. Here C is a rank K matrix incorporating the class assignment vectors. Using an inequality from Oliveira (2009), we bound the distance between the singular vectors of $\bar{G}$ and CMC'. Therefore, spectral clustering on $\bar{G}$ clusters the nodes into K different classes.

Remark 1. To determine the number of classes is a difficult model selection problem even for a single network. We will not discuss this problem in detail. We assume K is fixed and known in this paper.

Remark 2. The diagonal of G^t is always 0 because no self-edges are allowed; however the diagonal of CMC' is not necessarily 0. This may not cause a problem as $N \to \infty$. But for finite N, it may cause error in estimating the singular vectors. If this is the case, we may utilize the singular value decomposition that minimizes the off-diagonal mean

square error

$$\arg \min_{U,S} \sum_{i<j} (\bar{G}_{ij} - U_i S U_j')^2,$$

which can be computed by iterative singular value decomposition (Scheinerman and Tucker, 2010).

The condition in Theorem 1 requiring $\vec{P}$ to be stationary with identifiable mean M is restrictive. Spectral clustering on the mean graph is not effective in many cases. Consider the case for which

$$P^t \in \left\{ \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}, \begin{pmatrix} 0.3 & 0.7 \\ 0.7 & 0.3 \end{pmatrix} \right\}$$

where both outcomes are equally likely for all t. Then

$$M = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

is not identifiable. Spectral clustering on the mean graph fails to correctly estimate the class assignments as $T \to \infty$. But there is information contained in this multi-graph. We can use maximum likelihood estimation to estimate the class assignments correctly in this case.

4.2. Consistency of maximum likelihood estimate

Now we focus on the general case where we do not place any structure on $\vec{P}$. A natural way to estimate the class assignment is to use the maximum-likelihood estimate (MLE). We show that for a large enough fixed N, the MLE will estimate the class memberships correctly as $T \to \infty$.

First we define some notation. For any class assignment z, let $n_k(z) = \#\{i : z_i = k\}$ be the number of nodes in class k. Let $m(z) = \min_k n_k(z)$ denote the minimum number of nodes in any class under labels z. Let

$$n_{kl}(z) = \begin{cases} n_k n_l, & k \neq l \\ n_k(n_k - 1)/2, & k = l. \end{cases}$$

be the number of pairs of nodes in each block. We drop the dependency on z whenever it is unambiguous. We also drop the superscript t when we talk about a single layer of the network.

Now define some notation related to the MLE. The complete log-likelihood for parameters (z, P) is

$$l(z, P) = \sum_{i<j} \left(G_{ij} \log(P_{z_i z_j}) + (1 - G_{ij}) \log(1 - P_{z_i z_j}) \right).$$

Here P is a parameter not to be confused with the true class connection probability matrix. In particular, we are interested in the case $P_{kl} = \bar{P}_{kl}(z)$ where

$$\bar{P}_{kl}(z) = \frac{1}{n_{kl}(z)} \sum_{\substack{i:z_i=k \\ j:j \neq i, z_j=l}} P_{c_i c_j}.$$

Here $\bar{P}$ is the average of the true P under block assignment z. To ease notation, let

$$\sigma(p) = p \log(p) + (1 - p) \log(1 - p).$$

Denote the expectation of log-likelihood of $(z, \bar{P}(z))$ as

$$h(z) = E[l(z, \bar{P}(z))] = \sum_{k \leq l} n_{kl}(z) \, \sigma(\bar{P}(z)).$$

Now, as we do not observe the true P, the natural step is to estimate it with the empirical mean for any given z. So let

$$o_{kl}(z) = \sum_{\substack{z_i=k \\ z_j=l}} \begin{cases} G_{ij}, & k \neq l \\ G_{ij}/2, & k = l. \end{cases}$$

be the observed number of edges in block (k, l). Then the profile log-likelihood (Bickel and Chen, 2009) is defined as

$$f(z) = \sum_{k \leq l} n_{kl}(z) \, \sigma\left(\frac{o_{kl}(z)}{n_{kl}(z)} \right).$$

Let the expectation of f be

$$g(z) = E(f(z)) = \sum_{k \leq l} n_{kl}(z) \, E\left[\sigma\left(\frac{o_{kl}(z)}{n_{kl}(z)} \right) \right].$$

Now we are ready to state the consistency of the MLE for the multi-graph SBM. If all elements of P^t are bounded away from 0 and 1 and their column differences are at least some distance apart, then when we have a sufficient number of nodes in each block, the true label c uniquely maximizes the sum of profile log-likelihoods over the layers.

Theorem 2. *Let*

$$C_0 = \inf_{t,k,l} (P_{kl}^t, 1 - P_{kl}^t)$$

$$\delta = \inf_{t,k,l} \max_m \left[\sigma(P_{km}^t) + \sigma(P_{lm}^t) - 2\sigma\left(\frac{P_{km}^t + P_{lm}^t}{2} \right) \right].$$

Assuming $C_0 > 0$ and $\delta > 0$, if $m(c) = \min_k n_k(c)$ is sufficiently large, then

$$\hat{c} = \arg\max_z \sum_t f^t(z) \to c, \text{ a.s. as } T \to \infty.$$

The idea is that $\sum_t f^t(z)$ is the sum of independent profile log-likelihoods. We need N to be sufficiently large so that the expectation of the profile log-likelihood at each layer is maximized at the true labels c. Then as $T \to \infty$, we have convergence to expectation for $\sum_t f^t(z)$. We formalize the ideas by establishing the following lemmas.

Lemma 1 (From Choi et al. (2012)). *For any label assignment z, let $r(z)$ count the number of nodes whose true class*

assignments under c are not in the majority within their respective class assignment under z. Let

$$\delta = \min_{k,l} \max_{m} \sigma(P_{km}) + \sigma(P_{lm}) - 2\sigma\left(\frac{P_{km} + P_{lm}}{2}\right).$$

Then the expectation of the log-likelihood h is maximized by $h(c)$, and

$$h(c) - h(z) \geq \frac{r(z)}{2}\delta \min_{k} n_k(c).$$

In particular, for all $z \neq c$,

$$h(c) - h(z) \geq \frac{1}{2}\delta \min_{k} n_k(c).$$

Lemma 1 shows the expectation of the log-likelihood is maximized at the true parameters, and the difference of the true parameters and any other candidate is at least some distance apart which depends on the column difference of the probability matrix. However, as we work with the profile log-likelihood, we establish Lemmas 2 and 3 to bound the difference between the expectation of the profile log-likelihood and the complete log-likelihood.

Lemma 2. *Let $x \sim \frac{1}{N}Bin(N,p)$. For $p \in (0,1)$,*

$$E(\sigma(x)) \to \sigma(p) + \frac{1}{2N} + O\left(\frac{1}{N^2}\right), \text{ as } N \to \infty.$$

Lemma 3. *Assume $C_0 \leq P_{kl} \leq 1 - C_0$, $C_0 > 0$. For any $\delta_0 > 0$ and any z, if $\min_k n_k(z)$ is large enough, then the difference between the expectation of the profile log-likelihood $g(z)$ and the expectation of the complete log-likelihood $h(z)$ is bounded in the following manner:*

$$\left| g(z) - h(z) - \frac{K(K+1)}{4} \right| \leq \delta_0.$$

Lemma 2 utilizes a Taylor series expansion. For simplicity, we use big $O(\cdot)$ notation instead of specifying an actual bound. Readers can refer to the longer version posted on arXiv for the bound and the constants in the bound (Han et al., 2015). Lemma 3 uses Lemma 2 and shows that with a sufficiently large number of nodes, the difference of the expectation of the profile log-likelihood and complete log-likelihood is $K(K+1)/4$ and a negligible term δ_0. Combining the lemmas and using the concentration inequality, we can show that Theorem 2 provides sufficient conditions for the consistency of the multi-graph SBM. The proofs of Lemmas 1–3 and Theorem 2 can be found also in the longer version posted on arXiv (Han et al., 2015).

Remark 3. The main difference between the $N \to \infty$ case considered in most previous work and the $T \to \infty$ case that we consider is that, for $N \to \infty$, a direct bound is put on f and l. For $T \to \infty$, we need only to bound the *expectation*

of f and l. This is newly studied here. In other words, for some particular class connection probability matrix P, the number of nodes required in a single network to have an accurate estimate is much larger than what is needed in a multi-graph with a growing number of layers.

4.2.1. VARIATIONAL APPROXIMATION

The MLE is computationally infeasible for large networks because the number of candidate class assignments grows exponentially with the number of nodes. To overcome the computational burden, variational approximation, which replaces the joint distribution with independent marginal distributions, can be used to approximate the MLE. Daudin et al. (2008) has a detailed discussion of variational approximation in the SBM. We adapt it to the multi-graph SBM, resulting in the following update equations:

$$b_{ik} \propto \pi_k \prod_{j \neq i} \prod_t \prod_l \left[(P_{kl}^t)^{g_{ij}^t}(1 - P_{kl}^t)^{1-g_{ij}^t} \right]^{b_{jl}}$$

$$\pi_k \propto \sum_i b_{ik}$$

$$P_{kl}^t = \frac{\sum_{i \neq j} b_{ik} b_{jl} g_{ij}^t}{\sum_{i \neq j} b_{ik} b_{jl}},$$

where the b_{ik}'s denote the variational parameters. The derivation is straightforward; we provide details in the longer version posted on arXiv (Han et al., 2015).

Variational approximation has been shown to be consistent in the SBM (Celisse et al., 2012; Bickel et al., 2013). We conjecture that the performance of variational approximation is also good in the multi-graph SBM. Unless otherwise specified we use variational approximation to replace the MLE in all experiments.

5. Experiments

5.1. Numerical illustration

We begin with a toy example where we investigate empirically how many nodes are needed for the profile MLE to correctly recover the classes as $T \to \infty$. Due to the computational intractability of computing the exact profile MLE, we consider very a small network with $N = 16$ nodes and $K = 2$ classes where each class has 8 nodes. Consider two multi-graph SBMs with the following probability matrices:

$$\text{Case 1: } P^t \equiv \left(\begin{array}{cc} 0.55 & 0.45 \\ 0.45 & 0.55 \end{array} \right)$$

$$\text{Case 2: } P^t \equiv \left(\begin{array}{cc} 0.51 & 0.49 \\ 0.49 & 0.51 \end{array} \right)$$

The δ (defined in Theorem 2) corresponding to the row difference of P^t is much smaller in case 2. Empirically
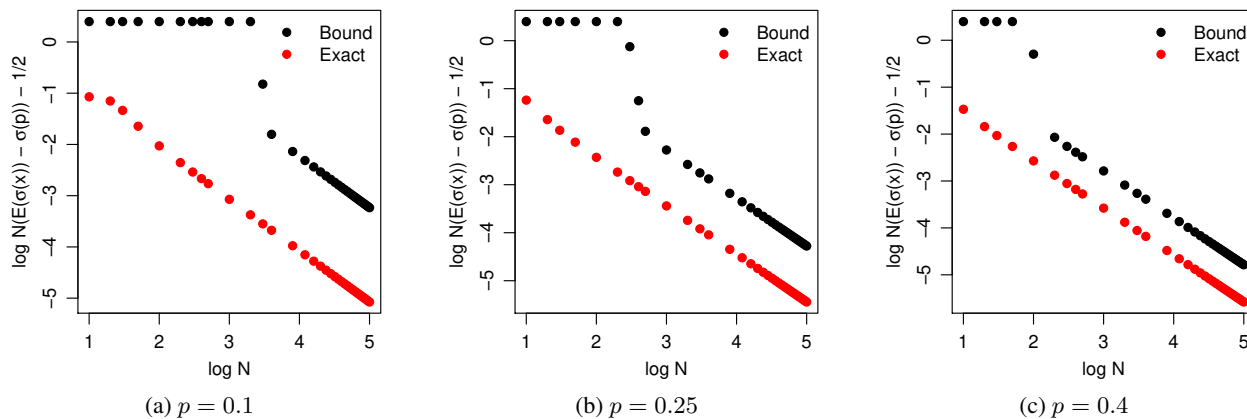
Figure 1. Comparison of bound in Lemma 2 to exact values of $N(E(\sigma(x)) - \sigma(p)) - 1/2$ for varying N and p. The tightness of the bound affects the minimum number of nodes required to guarantee consistency in Theorem 2.

Table 1. Minimum number of nodes N required for consistency of the profile MLE with $K = 2$ classes under different values for parameters C_0 and δ from Theorem 2.

δ \ C_0	0.3	0.25	0.2	0.15	0.1	0.05
0.165	42	50	64	88	124	184
0.091	44	52	66	92	142	234
0.040	46	56	70	94	148	314
0.010	66	68	74	100	156	330

the profile MLE succeeds to get the true labels in case 1 while it fails in case 2. Further analysis shows that in order to have consistency given the class connection probability matrix P^t in case 2, the total number of nodes should be at least 40. This toy example demonstrates that conditions on the probability matrices and network size are necessary for consistency. Theorem 2 provides *sufficient* conditions.

Next we investigate the tightness of the conditions in Theorem 2. The tightness of Lemma 1 was studied by Choi et al. (2012). We check the tightness of Lemma 2. For different p, we can calculate the exact value of $N(E(\sigma(x)) - \sigma(p)) - 1/2$ and compare it to the bound from Lemma 2. Figure 1 shows that the bound is loose for small N, but has almost the same asymptotic decay as the exact calculation. For small N, the remainder in Taylor expansion causes deviation. Also the bounds are looser for p closer to 0 or 1 but still informative in most cases.

For the special case of $K = 2$ classes, we can calculate all of the constants in the sufficient conditions in Theorem 2 for different values of C_0 and δ by enumerating cases. Details are provided in the longer version posted on arXiv (Han et al., 2015). Table 1 shows the smallest number of nodes N that is sufficient for consistency of the profile

MLE to hold for different values of C_0 and δ. Note that the minimum N is in the tens or hundreds, suggesting that the bounds in Theorem 2 are not overly loose and are indeed of practical significance.

5.2. Comparison with majority voting

As previously mentioned, majority voting is another way to utilize inference methods for a single network on multigraphs. We consider two majority vote methods as baselines for comparison, one that utilizes spectral clustering on each layer, and one that applies a variational approximation to each layer. When using majority voting between different layers of the network, the estimated class labels for each layer must first be aligned or matched. We utilize the Hungarian algorithm (Kuhn, 1955) to compute the maximum agreement matching between the estimated labels at layer t with the majority vote up to layer $t - 1$.

We conduct simulations to compare our proposed methods of spectral clustering on the mean graph and profile maximum-likelihood estimation with the majority vote baselines. We consider a well-studied scenario where we have 128 nodes initialized randomly into 4 classes (Newman and Girvan, 2004). For each layer, the within-class connection probability is 0.0968, and the between-class connection probability is 0.0521. Under such connection probabilities, the classes are below the detectability limit (Decelle et al., 2011) for a single layer, so the class estimation accuracy from a single layer is very low. We increase the number of layers and observe how the accuracy changes.

Figure 2 shows the accuracy of the two proposed methods compared to the two majority voting methods averaged over 100 replications. Both the profile MLE and spectral clustering on the mean graph have the anticipated in-
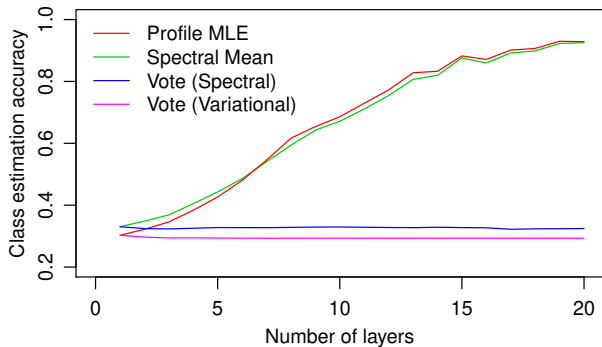
Figure 2. Simulation experiment comparing the proposed methods of profile MLE and spectral clustering on the mean graph with two majority vote baselines. The proposed methods increase in accuracy as the number of layers increases, but the two heuristic methods based on majority vote do not.

creasing accuracy over time. But the accuracies of the two heuristic majority vote methods do not improve. Though one may expect the errors in majority vote to be canceled out over time, these results show that, without careful averaging of errors, we cannot gain from the multiple layers. We find that this is due to choosing connection probabilities below the detectability limit; if we make the estimation problem easier by increasing the within-class probability above the detectability limit, then the majority vote methods do improve with increasing layers, albeit much slower than the methods we propose in this paper.

5.3. MIT Reality Mining data

Next we apply our model on the MIT Reality Mining data set (Eagle and Pentland, 2006). This data set comprises 93 students and staff at MIT in the 2004-2005 school year during which time their cell phone activities were recorded. We construct dynamic networks based on physical proximity, which was measured using scans for nearby Bluetooth devices at 5-minute intervals. We exclude data near the beginning and end of the experiment where participation was low. We discretize time into 1-week intervals, similar to Mutlu and Aviyente (2012); Xu et al. (2014), resulting in 39 time steps between August 2004 and May 2005.

We treat the affiliation of participants as ground-truth class labels and test our proposed methods. Two communities are found: one of 26 Sloan business school students, and one of 67 staff working in the same building. Since degree heterogeneity may cause problems in detecting communities using the SBM (Karrer and Newman, 2011), we reduce its impact by connecting each participant to the 5 other participants who spent the most time in physical proximity during each week. Figure 3 shows the empirical block connection probabilities within and between the two classes,
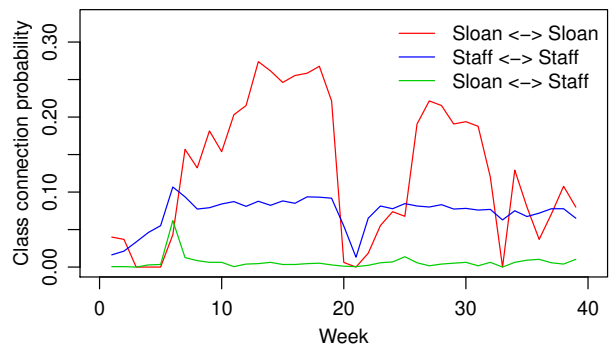

Figure 3. Estimates of class connection probabilities in the Reality Mining data set. The probabilities vary significantly over time, particularly for edges between Sloan students.

Table 2. Class estimation accuracy in the Reality Mining data set given data up to week listed in the first column. Best performer in each row is listed in bold. Both the proposed spectral clustering on the mean graph and profile maximum-likelihood estimation approaches improve over time, but majority vote does not.

WEEK	MAJ. VOTE	SPECTRAL MEAN	PROFILE MLE
10	**0.76**	0.62	0.57
15	0.82	0.94	**0.95**
20	0.83	0.95	**0.98**
25	0.78	0.95	**0.99**
30	0.80	0.97	**0.99**
35	0.80	0.97	**0.99**
END	0.77	0.97	**0.99**

estimated by the profile MLE. The class connection probabilities vary significantly over time, which validates the importance of the varying class connection probability assumption in our model. Notice that the two communities become well-separated around week 8. The class estimation accuracies for the different methods are shown in Table 2. Since the community structure only becomes clear at around week 8, the spectral and profile MLE methods are initially worse than majority voting but quickly improve and are superior over the remainder of the data trace. By combining information across time, the proposed methods successfully reveal the community structure while majority voting continues to improperly estimate the classes of about 20% of the people.

5.4. AU-CS multi-layer network data

We look at another example from a multi-layer network comprising five kinds of self-reported on-line and off-line relationships between the employees of a research department: Facebook, leisure, work, co-authorship, and lunch (AU-CS ML, 2014). We assume the class structure to be invariant across the different types of relations and apply our
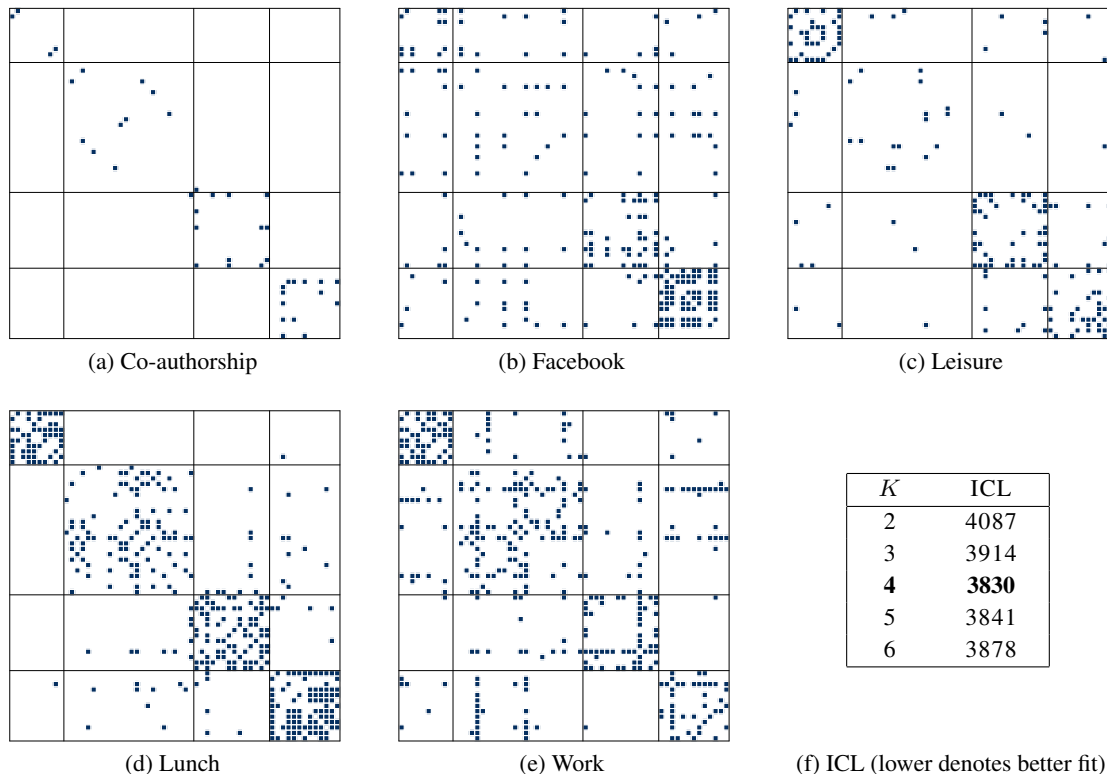
(a) Co-authorship

(b) Facebook

(c) Leisure

(d) Lunch

(e) Work

K	ICL
2	4087
3	3914
4	**3830**
5	3841
6	3878

(f) ICL (lower denotes better fit)

Figure 4. The estimated community structures in the AU-CS multi-layer networks overlaid onto the adjacency matrices of different relations. The dots denote connections (edges), and the grids correspond to SBM blocks.

model. For model selection, we extend the Integrated Classification Likelihood (ICL) criterion proposed by Daudin et al. (2008) for the single network SBM to multi-graphs to select the number of blocks K. Specifically we maximize

$$-2Q(\vec{G}) + (K-1)\log N + \frac{TK(K+1)}{2}\log\frac{N(N-1)}{2},$$

where $Q(\vec{G})$ is the variational approximation to the complete log-likelihood. We initialize the variational approximation with different randomizations as well as the spectral clustering solution. The term is maximized at $K = 4$.

Figure 4 shows the estimated 4 classes overlaid onto the adjacency matrix of each relation. Although we have no ground truth for this data set, we detect well-separated communities in all relations aside from co-authorship, which is an extremely sparse layer. Notice once again the difference in empirical connection probabilities over the multiple layers of the multi-graph.

For this data set, we do not have ground truth labels to evaluate the class estimation accuracy. We note, however, that the ICL obtained by our variational approximation algorithm is much better than the ICLs obtained by fitting an SBM on the mean graph and by majority vote, both of which are over 4000.

6. Discussion

In this paper, we investigated the multi-graph stochastic block model applied to dynamic and multi-layer networks with invariant class structure. Both spectral clustering on the mean graph and maximum-likelihood estimation are proved to be consistent for a fixed number of nodes when we have an increasing number of network layers, provided certain sufficient conditions are satisfied.

There are several interesting avenues for extensions of our analysis. First we can add a layer of probabilistic modeling on the probability matrices if we have additional information. Since dynamic networks usually vary smoothly over time, we can put a state-space model on the adjacency array (Xu and Hero, 2014). We can also use a hierarchical model on the probability matrices to couple them for analyzing multi-layer networks. Since our sufficient conditions do not consider such additional structure, an interesting area of future work would be to derive sufficient conditions that utilize the structure on the probability matrices, which would likely produce tighter bounds. It would also be interesting to draw connections to recent work on consistent estimation for populations of networks (Durante et al., 2014), for which no coupling between samples (layers) exists.

Acknowledgments

References

Ahmed, A. and Xing, E. P. (2009). Recovering time-varying networks of dependencies in social and biological studies. *Proceedings of the National Academy of Sciences*, 106(29):11878–11883.

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

Airoldi, E. M., Costa, T. B., and Chan, S. H. (2013). Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In *Advances in Neural Information Processing Systems 26*, pages 692–700.

Amini, A. A., Chen, A., Bickel, P. J., and Levina, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics*, 41(4):2097–2122.

AU-CS ML (2014). AU-CS MultiLayer network. http://sigsna.net/impact/datasets/.

Bickel, P., Choi, D., Chang, X., and Zhang, H. (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics*, 41(4):1922–1943.

Bickel, P. J. and Chen, A. (2009). A nonparametric view of network models and Newman-Girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073.

Celisse, A., Daudin, J.-J., and Pierre, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electronic Journal of Statistics*, 6:1847–1899.

Choi, D. S., Wolfe, P. J., and Airoldi, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika*, 99:273–284.

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

Decelle, A., Krzakala, F., Moore, C., and Zdeborová, L. (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106.

Durante, D. and Dunson, D. (2014). Bayesian logistic gaussian process models for dynamic networks. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, pages 194–201.

Durante, D., Dunson, D. B., and Vogelstein, J. T. (2014). Nonparametric bayes modeling of populations of networks. *arXiv preprint arXiv:1406.7851*.

Eagle, N. and Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268.

Fienberg, S. E., Meyer, M. M., and Wasserman, S. S. (1985). Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association*, 80(389):51–67.

Foulds, J. R., DuBois, C., Asuncion, A. U., Butts, C. T., and Smyth, P. (2011). 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.

Gao, C., Lu, Y., and Zhou, H. H. (2014). Rate-optimal graphon estimation. *arXiv preprint arXiv:1410.5837*.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends® in Machine Learning*, 2(2):129–233.

Han, Q., Xu, K. S., and Airoldi, E. M. (2015). Consistent estimation of dynamic and multi-layer block models. arXiv preprint arXiv:1410.8597.

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

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

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

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

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

Jin, J. (2012). Fast network community detection by SCORE. *arXiv preprint arXiv:1211.5803*.

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

Kim, M. and Leskovec, J. (2013). Nonparametric multi-group membership model for dynamic networks. In *Advances in Neural Information Processing Systems 26*, pages 1385–1393.

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97.

Lei, J. and Rinaldo, A. (2014). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237.

Magnani, M. and Rossi, L. (2011). The ML-model for multi-layer social networks. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, pages 5–12.

Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., and Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878.

Mutlu, A. Y. and Aviyente, S. (2012). Dynamic network summarization using convex optimization. In *Proceedings of the IEEE Workshop on Statistical Signal Processing*, pages 117–120.

Newman, M. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113.

Oliveira, R. I. (2009). Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. *arXiv preprint arXiv:0911.0600*.

Oselio, B., Kulesza, A., and Hero, A. O. (2014). Multi-layer graph analytics for dynamic social networks. *IEEE Journal of Selected Topics in Signal Processing*, 8.

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

Salter-Townshend, M. and McCormick, T. H. (2013). Latent space models for multiview network data. Technical Report 622, University of Washington.

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

Sarkar, P., Siddiqi, S. M., and Gordon, G. J. (2007). A latent space approach to dynamic embedding of co-occurrence data. In *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics*, pages 420–427.

Scheinerman, E. R. and Tucker, K. (2010). Modeling graphs using dot product representations. *Computational Statistics*, 25(1):1–16.

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

Szell, M., Lambiotte, R., and Thurner, S. (2010). Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31):13636–13641.

Wolfe, P. J. and Olhede, S. C. (2013). Nonparametric graphon estimation. *arXiv preprint arXiv:1309.5936*.

Xu, K. S. (2015). Stochastic block transition models for dynamic networks. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, pages 1079–1087.

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

Xu, K. S., Kliger, M., and Hero, A. O. (2014). Adaptive evolutionary clustering. *Data Mining and Knowledge Discovery*, 28(2):304–336.

Yang, J. J., Han, Q., and Airoldi, E. M. (2014). Nonparametric estimation and testing of exchangeable graph models. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, pages 1060–1067.

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

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