




Modeling and Change Detection for Count-Weighted Multilayer Networks

Hang Dong, Nan Chen & Kaibo Wang

To cite this article: Hang Dong, Nan Chen & Kaibo Wang (2019): Modeling and Change Detection for Count-Weighted Multilayer Networks, Technometrics, DOI: [10.1080/00401706.2019.1625812](https://doi.org/10.1080/00401706.2019.1625812)

To link to this article: <https://doi.org/10.1080/00401706.2019.1625812>

 View supplementary material 

 Accepted author version posted online: 31 May 2019.
Published online: 05 Jul 2019.

 Submit your article to this journal 

 Article views: 99

 View related articles 

 View Crossmark data 



Modeling and Change Detection for Count-Weighted Multilayer Networks

Hang Dong^a, Nan Chen^b, and Kaibo Wang^a

^aDepartment of Industrial Engineering, Tsinghua University, Beijing, China; ^bDepartment of Industrial Systems Engineering and Management, National University of Singapore, Singapore

ABSTRACT

In a typical network with a set of individuals, it is common to have multiple types of interactions between two individuals. In practice, these interactions are usually sparse and correlated, which is not sufficiently accounted for in the literature. This article proposes a multilayer weighted stochastic block model (MZIP-SBM) based on a multivariate zero-inflated Poisson (MZIP) distribution to characterize the sparse and correlated multilayer interactions of individuals. A variational-EM algorithm is developed to estimate the parameters in this model. We further propose a monitoring statistic based on the score test of MZIP-SBM model parameters for change detection in multilayer networks. The proposed model and monitoring scheme are validated using extensive simulation studies and the case study from Enron E-mail network.

ARTICLE HISTORY

Received December 2017
Accepted May 2019

KEYWORDS

Network monitoring; Score test; Statistical process monitoring; Variational-EM

1. Introduction

Network data are ubiquitous nowadays. Many physical, biological, and social systems are structured as networks to show pairwise relations or interactions among entities. These interactions in modern complex systems are often of multiple types. For example, in online social networks such as Twitter, users can have different types of interactions including “friendship,” “following,” and “retweeting.” A set of Internet users can communicate with each other through E-mail, messaging, and social media. In bibliographic networks, two authors may coauthor an article and/or cite each other’s work. Such multiple interactions among a set of individuals can be represented as multilayer networks where each layer represents one type of interaction (Kivelä et al. 2014). Note that the multilayer network systems are not random combinations of single network layers but are organized with significant correlations among the layers (Nicosia and Latora 2015). It is therefore important to characterize the correlations among these multiple layers in a network.

In real-world applications, the interactions are sometimes directional and include other quantitative information. For example, there are senders and recipients for phone calls, text messages, and E-mails indicating the contact direction from sender to recipient; the activities in online social networks such as “mention” and “following” are also directional. When these interactions are aggregated at some specific time scale, we get a directed weighted network with the edge weights representing the frequency or intensity of the interactions. For example, Figure 1 shows a two-layer network of E-mails among 13 employees of Enron Corporation. The first layer represents the ordinarily sent E-mails, and the second layer represents the carbon copy/blind carbon copy (cc/bcc) E-mails. Each edge represents the E-mails from the sender to the receiver using

cc/bcc, with the weight (expressed through edge width in the figure) representing the frequency of such E-mails. It is obvious that the interaction patterns of these two layers are similar but have layer-specific features as well. It is clear from this figure that the directed weighted version of a social network contains richer information than the undirected binary representation (Gao et al. 2017), and could help us better understand the underlying behavior of a network. Therefore, multilayer network models are required to capture the layer-specific information as well as the correlations among different layers for this kind of interaction data.

Moreover, individuals in a network tend to form densely connected subgroups called communities (Girvan and Newman 2002). The members of the same community show more interactions with each other than with individuals from any other communities. In Figure 1, these employees fall into two communities according to their identities at Enron Corporation. It is clearly shown that there are dense interactions between vertices of the same color, while the interactions between the two communities are sparser.

To characterize these kinds of interactions in a multilayer network, we propose a new stochastic block model (SBM) based on multivariate zero-inflated Poisson distribution called MZIP-SBM for multilayer directed weighted networks. The SBM framework is used to characterize the community structure of such a network. We assume that different layers share the same community structure, that is, the community label of a specific individual is the same across all layers, but the interaction intensity and density could be different for different layers. The concepts of interaction intensity and density are characterized in different parameters in the proposed model, enabling us to take care of sparse networks with only a few vertex pairs interacting frequently, while the rest remain inactive. This

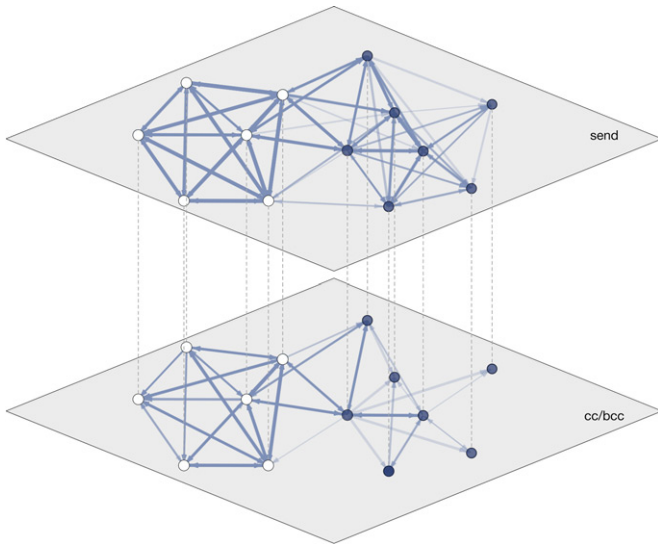


Figure 1. A two-layer directed network of Enron E-mails among 13 employees. The two layers represent directly sent E-mails and cc/bcc E-mails, respectively. Community labels of them are represented by different colors of vertices.

sparsity phenomenon is frequently observed in real networks with a number of nodes: the communication counts between all pairs of nodes are zero for a large proportion of pairs, and are much larger than one for almost all the rest ones, which motivates us to use a zero-inflated distribution to model these interaction counts. Furthermore, the proposed model could capture more characteristics of a multilayer network by utilizing correlations among different layers as well as the weight and direction information on the edges. A variational-EM algorithm is proposed to estimate the parameters in this model. We also propose a monitoring scheme based on a score test of the proposed model. Simulation experiments and a case study on the Enron E-mail corpus show that this model could characterize the sparsity feature and correlations among the layers of multilayer weighted networks well.

The remainder of the article is organized as follows. Section 2 reviews the literature related to statistical network modeling, extensions to multilayer networks, and network monitoring. In Section 3, we introduce the MZIP-SBM model for count-weighted multilayer networks. Section 4 describes the variational-EM algorithm for parameter estimation in the proposed model. Section 5 introduces a monitoring scheme according to the proposed model based on score test statistics. In Section 6, both simulation experiments and a case study on the Enron E-mail network are conducted to validate the performance of the proposed MZIP-SBM model and the consequent monitoring scheme. Finally, we conclude this work and propose several future research directions in Section 7.

2. Literature Review

There have been rich works in both statistical network modeling and anomaly detection on networks. In this section, we review the above two fields of research that are related to this work.

2.1. Statistical Network Modeling

Network modeling is a blooming research area with contributions from different fields, such as physics, statistics, and computer science. We will briefly review statistical modeling works of simple (single-layer) and multilayer networks; see Goldenberg et al. (2010) and Kivela et al. (2014) for a more detailed survey for single-layer and multilayer networks, respectively.

2.1.1. Statistical Models for Single-Layer Networks

Statistical network modeling begins with investigating the generation mechanism of a simple network, which could be the result of aggregating the observations of interactions among a group of individuals through a specific time period. Binary edges are mostly used to represent whether there exists any observed interactions between two vertices. Erdős and Rényi (1960) introduced the earliest random graph to characterize the generative mechanism of a subset of binary networks. In a simple random graph, the probability of connection between any two vertices is set to independently follow a certain probability distribution. Based on the idea of modeling the probability distribution of connection, there are following works on modeling networks using different approaches to link the presence of edges with vertex-specific information. Recent statistical network models in literature generally fall into two classes: the SBM and the latent space model.

The SBM was formally brought up by Holland, Laskey, and Leinhardt (1983) to characterize the community structure in networks. Under the setting of a SBM, the edges in a social network are conditionally independent, given the latent community membership of each vertex. In such a type of model, all the members in the same community have the same probability distribution of connecting to each other, and thus these members in the same community are structurally equivalent (Holland, Laskey, and Leinhardt 1983) or stochastically equivalent. To consider node-specific variability in the interactions, Airoldi et al. (2008) further developed a mixed membership SBM that assigned mixed community memberships to nodes. Another modification was the work by Karrer and Newman (2011), which proposed the degree-corrected SBM to incorporate heterogeneous vertex degrees into SBMs. Aicher, Jacobs, and Clauset (2015) introduced the weighted SBM to generalize the SBM to networks with edge weights drawn from an exponential family distribution.

Another group of statistical network models is based on the latent space model. The latent space model was first proposed by Hoff, Raftery, and Handcock (2002). It embeds the individuals into a latent space, and the probabilities of forming edges are then developed according to a projection function that is influenced by the latent positions of the corresponding end vertices. Further tasks, such as link prediction, could be conducted based on this latent space model (Miller, Griffiths, and Jordan 2009). Unlike the underlying assumption for community structure in networks, the latent space model does not have explicit group-clustering interpretations.

2.1.2. Models for Multilayer Networks

Rich information about multiple types of connections or interactions calls for the research in multilayer modeling

of networks. The history of multilayer networks and more sophisticated features as well as terminologies are explained by Kivelä et al. (2014) in detail. Usually in the statistical network modeling field, a multilayer network represents a network where different types of edges make up different layers over the same set of vertices. Han, Xu, and Airolidi (2015) studied a multigraph SBM which treated the probability of forming binary edges in different layers independently. Paul and Chen (2016) proposed a restricted multilayer SBM to characterize multi-relational binary data, but these two works were not able to deal with weighted edges. De Bacco et al. (2017) extended the mixed membership SBM to multilayer networks, assuming the probability distribution of forming edges in each layer to be Poisson distributions with mean affected by an affinity matrix to describe the density of edges between each pair of groups for each layer.

However, these models for multilayer networks did not explicitly characterize the correlations among different layers, which is a major feature for multilayer networks. Furthermore, the existing models for weighted networks assumed the interaction counts follow a Poisson distribution, which did not consider the sparsity characteristics of networks in real world. Caring for this issue, our proposed model used zero-inflated Poisson distribution to characterize the sparse interactions in each layer and explicitly characterize the correlations among different layers in multilayer networks.

2.2. Network Monitoring

The main purpose of network monitoring lies in detecting sudden changes across the network. These abrupt changes are also referred to as anomalies in literature. Savage et al. (2014) surveyed anomaly detection works in online social networks, and classified this field according to whether the target network is static or dynamic, and whether the vertices are labelled with classes or attributes. Another review of statistical methods for monitoring social networks was given by Woodall et al. (2017), which also showed the relationships between network monitoring methods and monitoring methods in engineering statistics and public health surveillance.

One group of change detection methods for networks are based on the community structure (Jun and Shun-zheng 2009), including methods based on variants of SBMs (Wilson, Stevens, and Woodall 2016).

Another group of methods monitors representative network statistics to detect anomalies. For example, Priebe et al. (2005) used scan statistics on networks to detect anomalous events on E-mail networks. The cross-correlations between scanned network metrics in the moving window were utilized to detect network changes by Cheng and Dickinson (2013). Traditional monitoring tools from the field of statistical process control (SPC) such as the cumulative sum (CUSUM) and the exponentially weighted moving average (EWMA) charts are conducted based on network metrics such as average betweenness and average closeness. Neil et al. (2013) adopted a scan statistic for both time window and subgraphs to detect locally anomalous subgraphs in computer networks. Azarnoush, Paynabar, and Bekki (2016) proposed a likelihood method based on the

edge existence as a function of vertex attributes to monitor the underlying network model. Peel and Clauset (2015) combined a generalized hierarchical random graph model with a Bayesian hypothesis test to determine the change point in evolving networks. However, these works are not applicable to multilayer networks.

To the best of our knowledge, there is no work proposed for monitoring multilayer networks, especially for count-weighted multilayer networks with sparse interactions. In the following, a model focusing on characterizing count-weighted multilayer networks will be introduced, with a monitoring scheme based on this model to detect the community-level behavioral changes in such networks.

3. MZIP-SBM: A Stochastic Block Model Based on MZIP Distribution

In this section, we first introduce the representation of a multilayer network as well as related notations, then we introduce the SBM for multilayer directed weighted networks based on a multivariate zero-inflated Poisson (MZIP) distribution, called MZIP-SBM. By using both the multilayer structure of the vertices as well as the weight and direction information on the edges, the proposed model could characterize a multilayer network more comprehensively.

3.1. Network Representation

We consider a directed multilayer graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ with M layers, where the vertex (also called node, these two expressions are equivalent in this article) set $\mathcal{V} = \{v_i\}_{i=1, \dots, N}$ consists of N vertices that are presented in each layer of the network, and the edge set $\mathcal{E} = \{\mathcal{E}^1, \mathcal{E}^2, \dots, \mathcal{E}^M\}$ consists of directed edges in this multilayer network, with \mathcal{E}^m representing the edges in layer m . Each edge connects two vertices in one specific layer representing one kind of interaction. For example, phone calls, text messages, and face-to-face contacts are multiple measures of social interaction and could be represented as separate layers in a multilayer network.

It is common to represent networks using adjacency matrices. Throughout this article, we use bolded letters to represent vectors and matrices, and use German letters to represent tensors. In an adjacency matrix \mathbf{A} , the element in the i th row and the j th column represents the interaction count from vertex i to vertex j during the specific time period. Since we consider directed weighted networks, the adjacency matrix of each network should be asymmetric and integer-valued. Note that generally we do not allow loops (namely, the case where an edge connects one vertex to itself), so the diagonal elements of such an adjacency matrix should be 0.

In the multilayer extension, we could use a three-dimensional tensor to represent such a network: $\mathfrak{A} = \{\mathbf{A}_{ij}\}, i = 1, \dots, N; j = 1, \dots, N$. Here \mathbf{A}_{ij} is a vector of length M : $\mathbf{A}_{ij} = (A_{ij}^{(1)}, A_{ij}^{(2)}, \dots, A_{ij}^{(M)})$, where M is the total number of layers in the network. For example, we use tensor \mathfrak{A} of dimension $3 \times 3 \times 2$ to represent the two-layer network in Figure 2, with each element of this tensor represented by A_{ij}^m where $i, j \in \{1, 2, 3\}, m \in \{1, 2\}$.

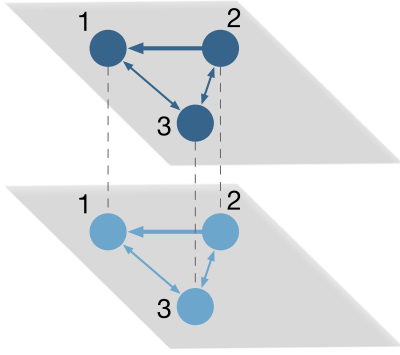


Figure 2. Illustration of a multilayer network. This network can be represented as a tensor \mathfrak{A} of dimension $3 \times 3 \times 2$.

3.2. Formulation of MZIP-SBM

In the proposed model, we use the common settings of a SBM with K blocks (communities). A random vector $Z_i \in \{1, \dots, K\}$, $i = 1, \dots, N$ is used to represent the community label of vertex v_i , with $Z_i = k$ if node i belongs to the k th community. Note that we assume each vertex only belongs to one community, and all layers share the same community structure. This assumption is reasonable for depicting several kinds of interactions among a group of individuals falling into different communities.

Different from ordinary SBMs for unweighted networks where the corresponding element for vertices v_i and v_j in the adjacency matrix follows a Bernoulli distribution (Holland, Laskey, and Leinhardt 1983), we aim to characterize the sparse interaction counts over the multilayer network and the correlations among different layers while preserving the community structure, and thus we choose to model the multivariate weighted edges using multivariate zero-inflated Poisson (MZIP) distribution proposed by Li et al. (1999). The observed edge weights between vertex i and vertex j in all the M layers denoted by a vector \mathbf{A}_{ij} of length M are determined by an MZIP distribution. There are two groups of parameters in the MZIP distribution: $\boldsymbol{\lambda}$ and \mathbf{p} where $\boldsymbol{\lambda} = (\lambda_{10}, \dots, \lambda_{M0}, \lambda_{00})$ and $\mathbf{p} = (p_0, p_1, \dots, p_M, p_{11})$. The parameters of this MZIP distribution only depend on which communities vertex i and vertex j come from

$$\mathbf{A}_{ij} | Z_i = q, Z_j = \ell \sim \text{MZIP}(\boldsymbol{\lambda}^{q\ell}, \mathbf{p}^{q\ell}), \quad (1)$$

where $\boldsymbol{\lambda}^{q\ell}$ and $\mathbf{p}^{q\ell}$ represent the MZIP parameters for edge weights between community q and community ℓ . Additionally, if \mathbf{A}_{ij} follows an MZIP distribution with parameters $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \mathbf{p})$, then

$$\begin{aligned} & (A_{ij}^{(1)}, A_{ij}^{(2)}, \dots, A_{ij}^{(M)}) \\ & \sim (0, 0, \dots, 0) \quad \text{with probability } p_0 \\ & \sim (\text{Poisson}(\lambda_{10} + \lambda_{00}), 0, \dots, 0) \quad \text{with probability } p_1 \\ & \sim (0, \text{Poisson}(\lambda_{20} + \lambda_{00}), 0, \dots, 0) \quad \text{with probability } p_2 \\ & \vdots \\ & \sim (0, 0, \dots, \text{Poisson}(\lambda_{M0} + \lambda_{00})) \quad \text{with probability } p_M \\ & \sim \text{multivariate Poisson}(\lambda_{10}, \lambda_{20}, \dots, \lambda_{M0}, \lambda_{00}) \\ & \quad \text{with probability } p_{11}, \end{aligned} \quad (2)$$

where $p_0 + p_1 + p_2 + \dots + p_M + p_{11} = 1$. Here λ_{00} characterizes the correlation among interaction counts across layers. The parameters of the above MZIP distribution consist of two groups: $\boldsymbol{\lambda}$ and \mathbf{p} . They play different roles in this model. Parameter vector $\mathbf{p}^{q\ell}$ determines the proportion of the following situations that may occur between community q and ℓ : absolutely no interaction in any layer, interactions following a Poisson distribution in only one layer, and interactions following a multivariate Poisson distribution in all layers; the other parameter vector $\boldsymbol{\lambda}^{q\ell}$ controls the specific rate or intensity of the Poisson-distributed interactions. They generally characterize the interaction density and intensity between community q and community ℓ , respectively. This configuration provides great flexibility in characterizing count-weighted multilayer networks. On one hand, \mathbf{p} gives a flexible representation of the sparsity level in all layers. On the other hand, λ_{00} represents the correlation between any two layers in the multivariate Poisson case, and thus could flexibly describe the over all correlations among different layers. In the next section, a variational-EM algorithm is introduced to estimate both the community labels $\{Z_i\}_{1 \leq i \leq N}$ and the MZIP parameters in this model.

4. Parameter Estimation

Given the dataset of a multilayer network, we can identify which community each vertex belongs to as well as estimate the model parameters for the underlying MZIP distribution. Because traversing discrete community labels $\mathbf{Z} = (Z_1, \dots, Z_N)$ is of exponential complexity with regard to the number of nodes N , obtaining the exact maximum likelihood estimates of MZIP-SBM is intractable. Therefore, we adopt variational-EM methods to achieve the approximate maximum likelihood estimates for SBMs (Mariadassou, Robin, and Vacher 2010). The proposed variational-EM algorithm is a combination of the EM algorithm and variational inference methods where the evidence lower bound (ELBO) is used to approximate the conditional distribution of community labels \mathbf{Z} during an EM iteration procedure.

Let \mathbf{A} denote the adjacency tensor of a multilayer network as formulated in Section 3.2: $\mathbf{A} = \{\mathbf{A}_{ij}\}_{i,j=1,\dots,N}$, $= \mathbf{A}_{ij}(A_{ij}^{(1)}, A_{ij}^{(2)}, \dots, A_{ij}^{(M)})$. The community label of any node i is treated as a latent random variable that follows a multinomial distribution

$$Z_i \stackrel{\text{iid}}{\sim} \mathcal{M}(1; \pi_1, \dots, \pi_K), \quad (3)$$

where $\mathcal{M}(1; \pi_1, \dots, \pi_K)$ denotes the multinomial distribution, $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$ with $0 < \pi_k < 1$ and $\sum_k \pi_k = 1$ represents the probability for a given vertex to belong to class k , $k = 1, \dots, K$. We assume that the number of communities K is known. Choosing the proper number of communities is another model selection problem, which we will briefly discuss at the end of this section.

Then (\mathbf{A}, \mathbf{Z}) is treated as the complete dataset where \mathbf{Z} is treated as unobserved latent variables. The problem is addressed by the estimation of both community labels \mathbf{Z} and parameters $\boldsymbol{\gamma} = (\boldsymbol{\pi}, \boldsymbol{\theta})$, where $\boldsymbol{\pi}$ denotes the parameters of the multinomial distribution in (3), and $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \mathbf{p})$ is the set of MZIP parameters in (2).

According to the conditional independence of the interaction counts over edges given \mathbf{Z} , we can decompose the log-likelihood of the complete dataset as

$$\begin{aligned} \log \mathbb{P}(\mathbf{Z}, \mathbf{A}; \boldsymbol{\gamma}) &= \log \mathbb{P}(\mathbf{Z}; \boldsymbol{\gamma}) + \log \mathbb{P}(\mathbf{A}|\mathbf{Z}; \boldsymbol{\gamma}) \\ &= \log \mathbb{P}(\mathbf{A}) + \log \mathbb{P}(\mathbf{Z}|\mathbf{A}; \boldsymbol{\gamma}). \end{aligned} \quad (4)$$

For convenience, we here denote $\mathbb{P}(Z_i = k)$ as $Z_i(k)$. Note that $Z_i(k)$ is a probability, which is different from $Z_i \in \{1, \dots, K\}$. Then according to (1) and (3),

$$\begin{aligned} \log \mathbb{P}(\mathbf{Z}, \mathbf{A}; \boldsymbol{\gamma}) &= \sum_{i=1}^N \sum_{q=1}^K Z_i(q) \log \pi_q \\ &+ \sum_{i \neq j} \sum_{q, \ell=1}^K Z_i(q) Z_j(\ell) \log f_{q\ell}(\mathbf{A}_{ij}), \end{aligned} \quad (5)$$

where $f_{q\ell}(\cdot) := f_{q\ell}(\cdot; \boldsymbol{\theta})$ is the probability density function of the MZIP distribution over the interaction count vector of length M between community q and community ℓ in (1).

Under the variational inference framework, we first find the ELBO of the log-likelihood of the incomplete dataset $\mathbb{P}(\mathbf{A})$

$$\begin{aligned} \text{ELBO}(R_A, \boldsymbol{\gamma}) &= \mathbb{E}(\log \mathbb{P}(\mathbf{Z}, \mathbf{A}; \boldsymbol{\gamma})) - \mathbb{E}(\log R_A(\mathbf{Z})) \\ &= \mathbb{E}(\log \mathbb{P}(\mathbf{A}|\mathbf{Z}; \boldsymbol{\gamma})) - \text{KL}(R_A(\mathbf{Z}), \mathbb{P}(\mathbf{Z}|\mathbf{A}; \boldsymbol{\gamma})), \end{aligned} \quad (6)$$

where KL denotes the Kullback–Leibler divergence and $R_A(\cdot)$ denotes the distribution we use to approximate $\mathbb{P}(\mathbf{Z}|\mathbf{A})$. By maximizing this ELBO, we try to search for an optimal $R_A(\mathbf{Z})$ that makes the conditional likelihood of the observed data large and at the same time makes the distance between $R_A(\mathbf{Z})$ and $\mathbb{P}(\mathbf{Z}|\mathbf{A})$ small.

To give an efficient maximization result of the ELBO, we limit the search of $R_A(\mathbf{Z})$ to the class of completely factorized distributions (namely, the mean-field variational family)

$$R_A(\mathbf{Z}) = \prod_i \mathcal{M}(Z_i; \boldsymbol{\tau}_i), \quad (7)$$

where \mathcal{M} denotes the multinomial distribution. We treat $\boldsymbol{\tau}_i$ as variational parameters to be optimized so that $R_A(\mathbf{Z})$ fits $\mathbb{P}(\mathbf{Z}|\mathbf{A}; \boldsymbol{\gamma})$ as well as possible: $\boldsymbol{\tau}_i = (\tau_{i1}, \dots, \tau_{iK})$ with $\sum_{q=1}^K \tau_{iq} = 1$, $\mathbb{E}_{R_A}(Z_i(q)) = \tau_{iq}$, and $\mathbb{E}_{R_A}(Z_i(q)Z_j(\ell)) = \tau_{iq}\tau_{j\ell}$.

With the mean-field family restriction, the ELBO can be rewritten as

$$\begin{aligned} \text{ELBO}(R_A, \boldsymbol{\gamma}) &= - \sum_i \sum_q \tau_{iq} \log \tau_{iq} + \sum_i \sum_q \tau_{iq} \log \pi_q \\ &+ \sum_{i \neq j} \sum_{q, \ell} \tau_{iq} \tau_{j\ell} \log f_{q\ell}(\mathbf{A}_{ij}). \end{aligned} \quad (8)$$

To maximize the ELBO with respect to both $R_A(\cdot)$ and $\boldsymbol{\gamma}$, we adopt the following iterative procedure: denoting by $R_A^{(r)}$ and $\boldsymbol{\gamma}^{(r)}$ the estimates after r steps, we compute

$$\begin{cases} R_A^{(r+1)} = \arg \max_{R_A \text{ factorized}} \text{ELBO}(R_A, \boldsymbol{\gamma}^{(r)}), \\ \boldsymbol{\gamma}^{(r+1)} = \arg \max_{\boldsymbol{\gamma}} \text{ELBO}(R_A^{(r+1)}, \boldsymbol{\gamma}). \end{cases} \quad (9)$$

Given $\boldsymbol{\gamma}$, we denote $\hat{\boldsymbol{\tau}}$ as the variational parameter defining the distribution

$$\hat{R}_A(\cdot) = \arg \max_{R_A(\cdot) \text{ factorized}} \text{ELBO}(R_A, \boldsymbol{\gamma}). \quad (10)$$

After some deviation, we can obtain that the optimal variational parameter $\hat{\boldsymbol{\tau}}$ satisfies the fixed point relation

$$\hat{\tau}_{iq} \propto \pi_q \prod_{j \neq i} \prod_{\ell=1}^K [f_{q\ell}(\mathbf{A}_{ij}) f_{\ell q}(\mathbf{A}_{ij})]^{\hat{\tau}_{j\ell}}. \quad (11)$$

Therefore, we can get $\hat{\boldsymbol{\tau}}$ by iterating the relation in (11) until convergence, and thus the distribution R_A is updated with this step.

For updating $\text{ELBO}(R_A, \boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma} = (\boldsymbol{\pi}, \boldsymbol{\theta})$ for a given distribution R_A , the optimal $\boldsymbol{\pi}$ and $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{p})$ can be calculated as below

$$\hat{\pi}_q = \frac{1}{N} \sum_i \tau_{iq}, \quad \hat{\boldsymbol{\theta}}^{q\ell} = \arg \max_{\boldsymbol{\theta}} \sum_{i \neq j} \tau_{iq} \tau_{j\ell} \log f(\mathbf{A}_{ij}; \boldsymbol{\theta}). \quad (12)$$

Now that both steps in (9) are computable, we could iteratively maximize the ELBO until convergence. An algorithmic sketch of the procedure is shown in Algorithm 1. More details on the derivation of this algorithm is in the supplementary materials.

Note that the number of communities K is assumed to be known in the proposed algorithm. However, there may be many situations where we do not know K . Here, we provide several off-the-shelf methods for choosing an appropriate K when we do not know it.

This issue has been discussed by a number of community detection works (see, e.g., Wang and Bickel 2017). A simple method is to define an objective function that represents the goodness of fit for any partition of all the individuals. Then, this objective function is optimized to find the best partition for a series of K values. From these K values we choose the K value that performs best under this criterion. For example, the integrated classification likelihood (ICL) proposed by Biernacki, Celeux, and Govaert (2000) could be used for this purpose, which approximates the complete data likelihood. Other criteria could also be considered for choosing a proper number of communities, for example, cross-validation for large-sized networks (Airoldi et al. 2008). In general, this issue is still a challenging,

Algorithm 1: Variational EM

Data: Observations $\mathbf{A} = \{A_{ij}^{(1)}, A_{ij}^{(2)}, \dots, A_{ij}^{(M)}\}_{i,j=1,\dots,N}$;
Number of communities K ;

Result: Community labels $\mathbf{Z} = \{Z_i\}_{i=1,\dots,N}$; Model parameters $\boldsymbol{\gamma} = (\boldsymbol{\pi}, \boldsymbol{\theta})$ with $\boldsymbol{\theta} = \{(\boldsymbol{\lambda}^{q\ell}, \boldsymbol{p}^{q\ell})\}$ for $q = 1, \dots, K; \ell = 1, \dots, K$

1 Initialization: Set $t = 0$; Set $\boldsymbol{\tau}_i^{(t)} = \underbrace{(1/K, \dots, 1/K)}_{\text{length } K}$,

$i = 1, \dots, N$; $\text{ELBO}^{(t)} = 0$;

2 Set $t = t + 1$;

3 E-step: Calculate $\boldsymbol{\tau}^{(t)}$ according to (11);

4 M-step: Calculate $\boldsymbol{\gamma}^{(t)}$, $\text{ELBO}^{(t)}$ according to (12) and (8);

5 **if** $|\text{ELBO}^{(t)} - \text{ELBO}^{(t-1)}|$ does not converge **then**

6 | go back to Step 2;

7 **else**

8 | Set $Z_i = \max_{k=1,\dots,K} \tau_{ik}^{(t)}, i = 1, \dots, N$;

9 | Return $\mathbf{Z}, \boldsymbol{\gamma}^{(t)}$;

10 **end**

unsolved problem, which is beyond the scope of the proposed model. More effective methods for choosing the correct number of communities will be investigated in the future.

Another notion is that, the same as almost all the algorithms for SBMs, this algorithm does not necessarily guarantee the identified communities are more densely interacted inside than with nodes outside (called assortative communities) (Newman and Park 2003). In fact, we allow the communities to have higher external interactions (called disassortative communities). Whether the communities are assortative or disassortative can be checked by comparing the MZIP parameters of internal interactions with external interactions. Therefore, we focus more on clustering those nodes with similar behaviors into communities to interpret the network behavior.

5. Monitoring the Multilayer Stochastic Block Model

The main purpose of network monitoring lies in detecting sudden changes across the network. These abrupt changes are also referred to as anomalies in the literature (Savage et al. 2014). Current works in the literature are mostly on single-layer networks and are not directly applicable to multilayer networks (Woodall et al. 2017). Furthermore, these works are not designed for multilayer count-weighted networks considering the sparsity and correlations among layers.

In this section, we propose a monitoring scheme based on the proposed MZIP-SBM model. A control chart is established based on the score test statistic of the proposed MZIP-SBM model to detect possible changes in a series of multilayer networks.

In the proposed method, the model parameters of several MZIP distributions are monitored to reflect the overall status of the whole network. The problem of monitoring the whole multilayer network is reduced to monitoring parameters of the proposed MZIP-SBM model. More specifically, we aim to monitor multiple multivariate ZIP processes simultaneously with score test statistics.

Formally, the change-point detection problem on a multilayer network is formulated as follows. Suppose we observe a series of multilayer networks \mathcal{G}_t 's, where each \mathcal{G}_t being the observed multilayer network at time t , which is uniquely represented by the three-dimensional adjacency tensor \mathbf{A}_t as explained in Section 3. The change-point model is

$$\mathcal{G}_t \stackrel{\text{iid}}{\sim} \begin{cases} F(\mathcal{G}; \mathbf{Z}, \boldsymbol{\theta}^0) & \text{for } t = 1, \dots, \tau \\ F(\mathcal{G}; \mathbf{Z}, \boldsymbol{\theta}^1) & \text{for } t = \tau + 1, \dots, \end{cases} \quad (13)$$

where τ is the unknown change point, F is the distribution developed from the proposed MZIP-SBM model, and $\boldsymbol{\theta}^0, \boldsymbol{\theta}^1$ are the in-control and out-of-control parameters for this model, respectively, with $\boldsymbol{\theta}^0 = \{\boldsymbol{\theta}^{q\ell(0)}\} = \{(\boldsymbol{\lambda}^{q\ell(0)}, \mathbf{p}^{q\ell(0)})\}$, $\boldsymbol{\theta}^1 = \{\boldsymbol{\theta}^{q\ell(1)}\} = \{(\boldsymbol{\lambda}^{q\ell(1)}, \mathbf{p}^{q\ell(1)})\}$, $q, \ell = 1, \dots, K$. We assume the community labels \mathbf{Z} are known in this setting, which can be regarded as the prior knowledge of the community structure of the unchanged network.

Due to the underlying community assumption of the SBM, developing a monitoring statistic for detecting change points in this framework can be decomposed into two levels: (1) developing a monitoring statistic for each of the K^2 community

pairs (they each have a set of parameters for the corresponding MZIP distribution); and (2) aggregating these statistics as a global monitoring statistic that fires an alarm when the network becomes abnormal. For the first level, we adopt the score test statistic for detecting the change in parameters of a specific MZIP distribution; then, for the second level, we aggregate these test statistics via adding them up to a global test statistic.

Since the null hypothesis is that the multilayer network is in control, that is, $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, we can test the likelihood that the observed network is generated according to $\boldsymbol{\theta}^0$, and ring an alarm when this likelihood is smaller than a threshold. This is the main intuition for the proposed monitoring method. The reason why we use the score test statistic is that the score test only requires the parameter estimates under the null hypothesis and thus avoids the heavy computation loads required to re-estimate the parameters at each step. When we have no prior knowledge on how the network will change, we test $\boldsymbol{\theta}^{q\ell} = (\boldsymbol{\lambda}^{q\ell}, \mathbf{p}^{q\ell})$ for every community pair (q, ℓ) using the proposed test statistic based on the following score and Fisher information

$$U(\hat{\boldsymbol{\theta}}^{q\ell}) = \sum_{i=1}^N \sum_{j \neq i} \frac{\partial \log Z_i(q) Z_j(\ell) L(\hat{\boldsymbol{\theta}}^{q\ell} | \mathbf{A}_{ij})}{\partial \boldsymbol{\theta}^{q\ell}}, \quad (14)$$

$$I(\hat{\boldsymbol{\theta}}^{q\ell}) = -\mathbb{E} \left(\sum_{i=1}^N \sum_{j \neq i} \frac{\partial^2 \log Z_i(q) Z_j(\ell) L(\hat{\boldsymbol{\theta}}^{q\ell} | \mathbf{A}_{ij})}{\partial \boldsymbol{\theta}^{q\ell} \partial \boldsymbol{\theta}^{q\ell T}} \right). \quad (15)$$

And the test statistic $s_{\theta^{q\ell}}$ is

$$s_{\theta^{q\ell}} = U^T(\hat{\boldsymbol{\theta}}^{q\ell}) I^{-1}(\hat{\boldsymbol{\theta}}^{q\ell}) U(\hat{\boldsymbol{\theta}}^{q\ell}). \quad (16)$$

This test statistic measures the consistency of the observed network with the in-control parameter values. Theoretically this test statistic asymptotically follows a chi-square distribution with d degrees of freedom where d is the total number of parameters used for the test. In practice, if the sample size is not large enough, the quantile of this statistic could be approximated by an empirical quantile of in-control samples.

When there are K communities, there are K^2 community pairs, each of them having separate parameter values in the MZIP model. Therefore, in all the three situations mentioned above, we have to aggregate K^2 local test statistics to obtain a global monitoring statistic. For the whole of a network with K communities, the summation of all the K^2 test statistics is used to establish the control chart

$$s_{\phi} = \sum_{q=1}^K \sum_{\ell=1}^K s_{\phi^{q\ell}}. \quad (17)$$

The upper control limit (UCL) is determined from $\mathbb{P}(s_{\phi} > \text{UCL}) \leq \alpha$, where α is a predefined false alarm rate. Note that asymptotically $s_{\phi^{q\ell}}$ follows a chi-square distribution with d degrees of freedom, where d is the number of parameters used to test for the underlying MZIP distribution of each community pair. The sum of these K^2 random variables, each following a chi-square distribution of degree d still follows a chi-square distribution with $d \times K^2$ degrees of freedom. Therefore, the control limits could be theoretically obtained by the in-control distribution of s_{ϕ} . Practically, if the sample size is not

large enough, the quantile of this statistic could be approximated by an empirical quantile of in-control samples. For a series of observations $\mathbf{A}_t, t = 1, 2, \dots$, the global test statistic $\{s_\phi(t)\}$ is calculated at each time point. Once $s_\phi(t) > \text{UCL}$, a change is declared. In summary, this monitoring method contains three steps: (1) estimate the in-control model parameters and determine the UCL; (2) for each time slot, calculate s_ϕ according to (17); and (3) raise an alarm as long as $s_\phi > \text{UCL}$.

Once an alarm is raised, a follow-up diagnosis procedure can be used for identifying the source of the change in the network: since the K^2 community-level pairwise test statistics $s_{\phi^{q\ell}}, q, \ell = 1, \dots, K$ are of the same scale, we can develop a common UCL for all the test statistics $s_{\phi^{q\ell}}$ according to the chi-square distribution or historical observations. Similar to the global UCL for s_ϕ , for the interactions between community q and ℓ , $\text{UCL}_{q\ell}$ can be determined from $P(s_{\phi^{q\ell}} > \text{UCL}_{q\ell}) \leq \alpha'$, where α' is again the predefined false alarm rate for test statistics $s_{\phi^{q\ell}}$. If any $s_{\phi^{q\ell}} > \text{UCL}_{q\ell}$ is found, then probably the interaction behavior between community q and ℓ is abnormal.

In addition to taking the summation of each communal pairwise test statistic $s_{\phi^{q\ell}}$ in (17), other methods could be considered to integrate these pairwise statistics. For example, Wilson, Stevens, and Woodall (2016) chose to model the parameters of each community pair separately, which is more convenient for tracking anomalies but may lead to too many monitoring statistics as K increases. Finding a cleverer way of combining these pairwise statistics remains a challenge for future work.

Given this test method, an alternative to testing all the MZIP parameters is to test a subset of parameters that is more likely to change or that corresponds to a suspected change in behavior according to domain experience. By simply replacing $\theta^{q\ell}$ with another subset of parameters $\phi^{q\ell} \subseteq \theta^{q\ell}$ in (14) and (15), different test statistics regarding to $\phi^{q\ell}$ are expressed and could be easily calculated numerically. For example, we may monitor $\phi^{q\ell} = \lambda^{q\ell}$ or $\phi^{q\ell} = p^{q\ell}$ separately for each community pair (q, ℓ) to detect abrupt changes in the density or intensity of interactions, respectively. Using a subset of parameters could make the monitoring more sensitive for a specific type of changes. Nevertheless, it may miss the changes on the parameters excluded from the test statistic. Extensive experiments of

monitoring all and a subset of parameters are conducted in Section 6.1.1.

6. Performance Study

6.1. Simulation Experiments

In this part, simulation experiments are carried out to validate the performance of the proposed parameter estimation algorithm as well as to evaluate the performance of the proposed monitoring scheme.

6.1.1. Evaluation of Parameter Estimation

To evaluate the performance of the proposed parameter estimation method, we simulate multilayer networks with known parameters and assess the accuracy in estimating the parameters with the proposed variational-EM algorithm.

Two sample sizes are considered here: $n = 1$ and $n = 4$, where n is the number of multilayer networks used as samples for estimating the model parameters. For each sample size, we let the number of layers $M = 3$, the number of vertices $N = 30$, and the number of communities $K = 2$. An example of generated multilayer networks is shown in Figure 3. The two-community block structure is clear in this three-layer network, and the interactions are correlated across these layers.

We repeat the estimation algorithm 100 times to assess its estimation performance. When there are $K = 2$ communities, we have four groups of MZIP parameters: (λ^{11}, p^{11}) , (λ^{12}, p^{12}) , (λ^{21}, p^{21}) , and (λ^{22}, p^{22}) . The superscripts represent the community pair that corresponds to each set of MZIP parameters. We set the parameters of the simulated multilayer network to reflect a common phenomenon in real networks: interactions between members from the same community are more frequent than interactions between members from two different communities, which is also called assortative mixing (Newman and Park 2003). Therefore, we set the intensity parameters to be $\lambda_{10}^{11} = \lambda_{20}^{11} = \lambda_{30}^{11} = \lambda_{10}^{22} = \lambda_{20}^{22} = \lambda_{30}^{22} = 15$, $\lambda_{10}^{12} = \lambda_{20}^{12} = \lambda_{30}^{12} = \lambda_{10}^{21} = \lambda_{20}^{21} = \lambda_{30}^{21} = 5$. The degree of interaction correlation among layers is set as $\lambda_{00}^{q\ell} = 15$ when $q = \ell$, $\lambda_{00}^{q\ell} = 5$ when $q \neq \ell$. For interaction density p , we set $p_0^{11} = 0.2$, $p_1^{11} = p_2^{11} = p_3^{11} = 0.1$, $p_1^{11} = 0.5$; $p_0^{12} = p_0^{21} = 0.5$, $p_1^{12} = p_2^{12} = p_3^{12} = p_1^{21} = p_2^{21} = p_3^{21} = 0.15$, $p_{11}^{12} = p_{11}^{21} = 0.05$; $p_0^{22} = 0.1$, $p_1^{22} = p_2^{22} = p_3^{22} = 0.1$, $p_{11}^{22} = 0.6$.

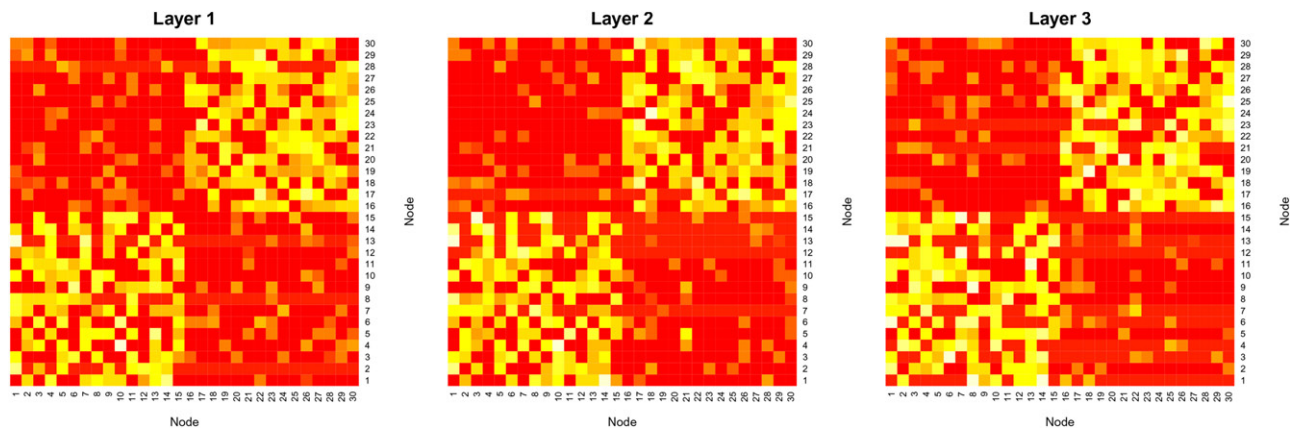


Figure 3. Illustration of a simulated multilayer network.

Table 1 displays the absolute estimation errors of λ and \mathbf{p} , with the standard deviations shown in brackets. We could tell from the results that the proposed algorithm provides a satisfactory estimation of the MZIP-SBM model. Moreover, when we have more samples of multilayer networks, we could, in general, estimate parameters with higher accuracy and lower bias. Moreover, the identification accuracy of community identity \mathcal{Z} in this parameter setting is 100% for both $n = 1$ and $n = 4$, indicating our proposed algorithm performs quite well in community identity classification.

6.1.2. Performance of Network Monitoring

To evaluate the performance of the proposed monitoring scheme, we generate temporal multilayer networks at each sampling epoch and compare the average time to signal of the proposed monitoring scheme with the other five competitors: the first one (ZIP) models each layer of the network independently as a zero-inflated Poisson distribution and uses the sum of score test statistics for each layer as the monitoring statistic; the second (Poisson) is similar to the first method except that it models the interactions in each layer with a Poisson distribution; the third (SCAN) is a modification of the method proposed by Priebe et al. (2005) for monitoring E-mail networks based on the distribution of network scan statistics; the rest two methods (BT and CL) are based on average betweenness and closeness of nodes, which are ordinary network metrics for clustering effects. Note that currently few monitoring methods for multilayer count-weighted networks have been proposed, and existing monitoring methods for single-layer binary networks suffer with substantial information loss in count-weighted multilayer networks. Therefore, here we just adapt three naive methods (SCAN, BT, and CL) to this setting to provide a monitoring performance baseline. We calculate the sum of the maximum number of neighbors, betweenness and closeness in all layers for the monitoring statistics for SCAN, BT, and CL methods.

We first compare these methods to detect changes within one community, which aims to justify the effectiveness of the score test under the proposed MZIP distribution. Then we compare the performance of the proposed method when members forming $K = 2$ communities with equal sizes. Different types of structural changes in the community behaviors are considered. We replicate each simulation 1000 times, and we then estimate the average run length (ARL) from these 1000 simulations.

Multilayer networks with $N = 10$ vertices and $M = 3$ layers are generated according to the MZIP distribution in (2). We intend to choose parameters to mimic the inter-

actions among an active community. Therefore, we assume the interaction intensity to be at a relatively high level. The parameter setting of the MZIP model is the same as that of within community interactions for the first community in Section 6.1.1: $\lambda^0 = (\lambda_{10}, \lambda_{20}, \lambda_{30}, \lambda_{00}) = (15, 15, 15, 15)$, and $\mathbf{p}^0 = (p_0, p_1, p_2, p_3, p_{11}) = (0.2, 0.1, 0.1, 0.1, 0.5)$. Since there is a constraint in \mathbf{p} : $\sum_{i=0}^m p_i + p_{11} = 1$, we only use the elements excluding p_{11} to calculate the score test statistics. We consider different possible changes in interaction patterns in reality, and evaluate the out-of-control ARL performance of the proposed monitoring method. We set the in-control ARL to be 200 approximately.

A variety of possible shift types and shift sizes are evaluated for when the process goes out-of-control. These shift types represent possible scenarios when the interaction patterns go through structural changes: a change in λ indicates the change in the intensity or frequency among individuals who have fixed interaction channels (nonzero interaction counts); a change in \mathbf{p} indicates the overall interaction density, or the degree of sparsity in the network.

The resulting ARLs when λ changes are shown in Table 2. We could tell from the results that the proposed MZIP method is able to detect all the change types we consider in this experiment. It is especially efficient in detecting changes in λ_{00} , which indicates the correlation of interactions among layers. The ZIP method is competitive in detecting changes in $(\lambda_{10}, \lambda_{20}, \lambda_{30})$, but it is less effective at detecting changes in λ_{00} . Both the Poisson and SCAN methods are ineffective in most situations. Generally, only when the change in λ occur across all layers and is significant enough could the Poisson method have even limited ability to detect these changes. The SCAN, BT, and CL methods could not detect any changes in λ , because they all neglect the weight information of the edges.

Table 3 shows the ARLs when \mathbf{p} changes. When the proportion of zero interaction counts decreases and the proportion of multivariate Poisson counts increases, all these methods can detect these changes. However, when the decrease in the proportion of zero interaction counts transfer to the proportion of interactions in each of the layers, only the proposed MZIP method could quickly detect this change. Considering that the structural changes in networks in applications usually occur in both λ and \mathbf{p} , the proposed MZIP method could generally detect this kind of change faster than other competitors when there is only one community in the multilayer network.

Table 1. Parameter estimation results for $N = 30, M = 3, K = 2$.

	$(q, \ell) = (1, 1)$		$(q, \ell) = (1, 2)$		$(q, \ell) = (2, 1)$		$(q, \ell) = (2, 2)$	
	$n = 1$	$n = 4$	$n = 1$	$n = 4$	$n = 1$	$n = 4$	$n = 1$	$n = 4$
λ_{00}	0.28(1.31)	0.03(1.39)	0.18(2.72)	0.14(2.56)	0.04(2.77)	0.00(2.84)	0.04(1.26)	0.02(1.34)
λ_{10}	0.06(0.53)	0.02(0.44)	0.03(0.65)	0.03(0.55)	0.07(0.59)	0.00(0.54)	0.01(0.47)	0.06(0.49)
λ_{20}	0.00(0.49)	0.01(0.44)	0.05(0.62)	0.02(0.56)	0.09(0.54)	0.05(0.69)	0.01(0.50)	0.06(0.42)
λ_{30}	0.06(0.44)	0.03(0.47)	0.07(0.62)	0.01(0.51)	0.07(0.58)	0.03(0.54)	0.01(0.43)	0.06(0.46)
p_0	0.060(0.053)	0.005(0.031)	0.006(0.034)	0.001(0.030)	0.002(0.032)	0.000(0.031)	0.062(0.055)	0.002(0.021)
p_1	0.001(0.020)	0.003(0.019)	0.001(0.023)	0.001(0.022)	0.004(0.024)	0.001(0.021)	0.004(0.021)	0.003(0.018)
p_2	0.001(0.020)	0.000(0.020)	0.000(0.022)	0.002(0.024)	0.002(0.022)	0.001(0.026)	0.002(0.020)	0.000(0.019)
p_3	0.002(0.022)	0.001(0.020)	0.001(0.023)	0.000(0.023)	0.000(0.025)	0.002(0.023)	0.002(0.020)	0.001(0.022)
p_{11}	0.061(0.064)	0.007(0.034)	0.003(0.016)	0.002(0.014)	0.001(0.014)	0.000(0.014)	0.063(0.064)	0.001(0.032)

Table 2. ARL results for λ when monitoring one community.

λ	MZIP	ZIP	Poisson	SCAN	BT	CL
(15, 15, 15, 15)	200	200	200	216.45	200	200
(16, 15, 15, 15)	66.93	56.92	251.32	225.41	333.33	250
(17, 15, 15, 15)	6.34	7.28	244.47	223.66	166.67	250
(18, 15, 15, 15)	1.60	1.84	194.66	222.02	250	333.33
(16, 16, 16, 15)	110.23	8.27	272.60	214.49	166.67	250
(17, 17, 17, 15)	5.15	1.49	99.40	221.22	250	166.67
(18, 18, 18, 15)	1.27	1.01	32.92	227.38	500	166.67
(15, 15, 15, 16)	76.54	173.54	194.69	213.59	125	200
(15, 15, 15, 17)	27.90	163.67	187.29	219.84	166.67	142.86
(15, 15, 15, 18)	8.55	153.11	190.92	236.02	333.33	250
(15, 15, 15, 19)	3.80	133.98	193.13	221.44	142.86	500
(15, 15, 15, 20)	1.91	126.10	188.94	230.40	200	333.33

Table 3. ARL results for p when monitoring one community.

p	MZIP	ZIP	Poisson	SCAN	BT	CL
(0.2, 0.1, 0.1, 0.1, 0.5)	200	200	200	216.45	199	200
(0.15, 0.1, 0.1, 0.1, 0.55)	61.71	18.09	125.32	50.50	16.95	15.87
(0.10, 0.1, 0.1, 0.1, 0.60)	6.65	3.17	10.84	13.87	2.67	2.67
(0.05, 0.1, 0.1, 0.1, 0.65)	1.29	1.23	2.31	5.21	1.16	1.16
(0.05, 0.15, 0.15, 0.15, 0.5)	1.28	25.92	251.79	78.23	25	23.81

Table 4. ARL results for λ when monitoring two communities.

λ	MZIP	ZIP	Poisson	SCAN	BT	CL
λ^0	200	200	200	215.52	200	200
$\lambda^{11} = (20, 15, 15, 15)$	21.52	1.04	130.22	214.21	214.03	211.46
$\lambda^{11} = (20, 20, 20, 15)$	18.13	1.00	17.26	211.03	213.71	217.37
$\lambda^{11} = (23, 15, 15, 15)$	1.08	1.00	47.18	222.89	208.91	220.94
$\lambda^{11} = (23, 23, 23, 15)$	1.01	1.00	3.33	212.82	219.06	206.25
$\lambda^{11} = (15, 15, 15, 20)$	26.46	198.20	209.07	210.86	200.92	212.85
$\lambda^{11} = (15, 15, 15, 23)$	2.56	146.62	191.25	221.28	203.93	212.43
$\lambda^{12} = (15, 10, 10, 5)$	193.29	19.60	170.46	231.39	214.90	218.81
$\lambda^{12} = (15, 15, 15, 5)$	92.71	1.88	97.30	219.86	208.59	226.03
$\lambda^{12} = (18, 10, 10, 5)$	37.70	2.55	103.50	222.91	210.11	199.84
$\lambda^{12} = (18, 18, 18, 5)$	1.18	1.01	17.84	213.25	204.46	222.39
$\lambda^{12} = (10, 10, 10, 10)$	77.04	141.70	204.65	217.73	207.51	218.25
$\lambda^{12} = (10, 10, 10, 13)$	25.53	140.63	203.89	218.61	213.84	202.59

We then extend this experiment to multilayer networks with multiple communities. More specifically, we generate 3-layer networks of 20 vertices, with these vertices distributed into $K = 2$ communities. Again, we use the parameter settings listed in Section 6.1.1. Similar to the single-community experiment, both shifts in λ and in p are considered. For changes in λ , we consider the increase in inside-community interaction intensity (including the increase in one layer and all layers), the increase in inside-community interaction correlation, the increase in intercommunal interaction intensity (including the increase in one layer and all layers) and the increase in intercommunal interaction correlation. These scenarios correspond to changes in $\lambda_{10}^{11}, (\lambda_{10}^{11}, \lambda_{20}^{11}, \lambda_{30}^{11}), \lambda_{00}^{11}, \lambda_{10}^{12}, (\lambda_{10}^{12}, \lambda_{20}^{12}, \lambda_{30}^{12}), \lambda_{00}^{12}$. For changes in p , we also investigate the density due to both inside-community and between-community interactions. The shift magnitudes are set as $\Delta(p) = (-0.1, 0, 0, 0, 0.1)$, $\Delta(p) = (-0.15, 0, 0, 0, 0.15)$, and $\Delta(p^{11}) = (-0.15, 0.05, 0.05, 0.05, 0)$ for these two types of changes.

The ARLs under these different kinds of shifts are shown in Tables 4 and 5. Similar to the results in the single-community experiment, when the within-community interaction intensity λ^{11} or the between-community interaction intensity λ^{12}

increases while the correlation among layers does not change, the ZIP method is more capable of detecting shifts quickly. However, the ZIP method could not detect any changes if the change occurs in correlations among the layers. The SCAN, BT, and CL methods still perform unsatisfactorily in detecting these changes. When the proportion of all-zero interactions decreases, and the proportion of nonzero interactions in one single layer increases, the proposed MZIP method would outperform the ZIP method. Both BT and CL methods are sensitive to the changes in p , but they are not capable to detecting changes in λ . The SCAN method performs badly in all the changes evaluated in this experiment. This result suggests that the proposed monitoring method could detect increases both in inside-community interactions and between-community counts and was especially ideal for detecting the changes in correlations among different layers.

According to our simulation experiment, we find that our method is especially good at detecting changes in correlation among different layers in the network. Therefore, for general single-layer network monitoring without any community structure, some simple methods can be used, such as SCAN, BT, or CL; when each layer is almost independent with any other

Table 5. ARL results for p when monitoring two communities.

p	MZIP	ZIP	Poisson	SCAN	BT	CL
p^0	200	200	200	215.52	200	200
$p^{11} = (0.1, 0.1, 0.1, 0.1, 0.6)$	68.87	5.23	31.31	184.75	11.41	12.01
$p^{11} = (0.05, 0.1, 0.1, 0.1, 0.65)$	45.52	1.55	6.00	166.98	4.55	4.52
$p^{11} = (0.05, 0.15, 0.15, 0.15, 0.5)$	42.43	47.28	182.70	205.73	48.85	48.63
$p^{12} = (0.4, 0.15, 0.15, 0.15, 0.15)$	149.78	7.10	201.32	125.79	5.88	6.08
$p^{12} = (0.35, 0.15, 0.15, 0.15, 0.2)$	104.06	1.76	197.69	125.79	2.73	2.62
$p^{12} = (0.35, 0.2, 0.2, 0.2, 0.05)$	54.33	94.33	210.02	166.73	25.57	28.02

layer in the multilayer network, we would suggest practitioners to use the ZIP method. However, for most of the cases in real applications, the inter-layer dependence in multilayer networks is significant. In this case only our proposed method could perform well for detecting correlational changes.

6.2. Application on Enron Multilayer E-mail Networks

The Enron E-mail corpus is a unique and famous set of network data reflecting communications in a real energy trading company. The Enron scandal, publicized in October 2001, eventually led to the bankruptcy of Enron Corporation. Following a March 26, 2003 ruling by the Federal Energy Regulatory Commission of the United States of America, approximately 0.6 million E-mails to/from Enron employees were publicly released. This dataset consists of a collection of E-mails corresponding to the E-mail among 184 unique E-mail addresses representing senior management and other employees at Enron, collected over a period from 1998 to 2002 (Priebe et al. 2005).

We establish a two-layer weighted directed network for this E-mail dataset: the first layer represents the E-mails sent directly from the source node to the target node, while the second layer represents the cc/bcc E-mails, with each edge standing for an E-mail from a sender to a receiver using cc/bcc. Intuitively these two layers indicate different types of working relationships: sending E-mails directly indicates collaboration on specific projects in most cases, but cc/bcc is more likely to show the relation of superior and subordinate. The counts of E-mail communications are treated as weights over edges. We first apply the proposed variational-EM algorithm on this network to show that the direction of interaction as well as the edge count information could be well expressed by the proposed model. Then we use monthly interaction data to establish a control chart to verify that the anomalous time points detected by the proposed monitoring scheme correspond to known events that occurred in Enron Corporation.

To conduct the parameter estimation, we choose a subnetwork of 13 vertices consisting of 2 cliques from May to August in 2001 of this Enron E-mail network. These vertices are deliberately chosen so that they fall into two communities. The first community contains 7 vertices corresponding to managers in Enron, and the second contains 6 vertices that are ordinary employees in Enron. We conduct the variational-EM algorithm proposed in Section 4 on this two-layer network. The result shows that considering the correlation of different layers is helpful for discovering the true community structure.

The community labels for all these 13 vertices are correctly identified by the proposed algorithm, that is, the first 7 vertices

Table 6. Estimated parameters in Enron E-mail dataset.

var	$(q, l) = (1, 1)$	$(q, l) = (1, 2)$	$(q, l) = (2, 1)$	$(q, l) = (2, 2)$
λ_{00}	3.49	2.96e-04	3.82e-08	5.05e-07
λ_{10}	4.57	0.60	0.50	23.62
λ_{20}	0.01	5.98	0.05e-08	6.22
p_{00}	0.29	0.92	0.27	0.03
p_1	0.29	0.05	0.29	0.40
p_2	0.03	0.02	0.27	6.5e-05
p_{11}	0.41	0.01	0.18	0.57

belong to one community, and the remaining 6 vertices belong to the other. The estimated values of the parameters are shown in Table 6. λ_{00} reflects the correlation between the two layers in the MZIP model, and $\lambda_{10} + \lambda_{00}$ and $\lambda_{20} + \lambda_{00}$ reflect the interaction intensity in the two layers. p_{00} , p_1 , p_2 , and p_{11} embody the interaction density of the network; they are, respectively, the probabilities of both layers having no interactions, interactions that occurred only in layer 1, interactions that occurred only in layer 2, and interactions that occurred in both layers according to a bivariate Poisson distribution. (q, l) represents the community pair that the sender and the receiver of an interaction belong to, respectively. It is clear from the result that these 2 communities show different behaviors in different layers, both in within-community interactions and between-community interactions. However, if we conduct the variational-EM algorithm on the aggregated network or on these two layers separately, it could not even correctly identify the two underlying communities, that is, the estimated $\hat{\tau}_{iq}$ converges to be the same value for any q and i , which implies that every individual is equally likely to belong to any community. This result indicates that combining different types of information (multiple layers) with correlation could indeed help to find the true underlying community structure.

The monitoring scheme based on the score test statistic of the MZIP model for each community pair is also conducted on the Enron E-mail dataset. For each month from January 1999 to June 2002, we calculate the test statistic according to the proposed method. As is shown in Figure 4, the abnormal outbreaks first occurred in late 2000 and reached two peaks in May 2001 and October 2001. It is found that real events occurred in Enron during these special time periods. In October 2000, the Enron attorney discussed Tim Belden's strategies (Tim Belden was the head of trading at Enron and was one of the first executives to be prosecuted and to admit wrongdoings at Enron); in November 2000, Federal Energy Regulation Commission (FERC) exonerated Enron. These two events mainly correspond to a sudden increase in interaction intensity, which

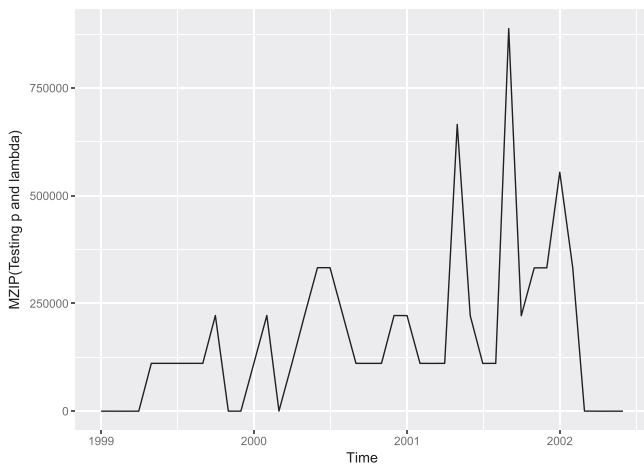


Figure 4. The Enron E-mail interaction patterns monitored with the proposed MZIP method.

is obvious in Figure 4. In addition, in May 2001, the Chief Executive Officer of the Division had more interactions with the employees although his “in-control” pattern in previous networks was more interaction with the Chief Financial Officer and Vice President of the Division; in October 2001, the Enron scandal was revealed, and every division of the corporation experienced changes, reflected in the change of both λ and p in the proposed model. These key events are consistent with the findings of other researchers, such as Priebe et al. (2005), Xu and Hero (2013), and Peel and Clauset (2015).

7. Conclusions

In this article, we propose a SBM for multilayer count weighted networks called MZIP-SBM. This work provides a solution to explicitly characterize the correlations among different layers as well as the sparsity of interactions on real networks. Then a monitoring scheme based on the proposed model is proposed for monitoring community-level behavioral changes in a multilayer network. We also propose a variational-EM algorithm to calculate the community memberships and the parameters for MZIP models of each community pair. Through simulation experiments and a case study on the Enron E-mail network, our proposed model appears to be effective in characterizing the correlation information among layers and the zero-inflated interaction counts in count-weighted multilayer networks. The proposed monitoring scheme is also demonstrated to perform well in detecting changes in different change scenarios.

The major limitation of this proposed model is the computation complexity growth of parameter estimation algorithm with regard to K , N , and M . Currently, almost all statistical (generative) models for multilayer networks have the same problem of scaling in parameter estimation. Hopefully this issue will be tackled by faster optimization techniques.

Statistical models for multilayer networks have aroused great interest among researchers in recent years. Our proposed model motivates several avenues of future research. First, more exquisite monitoring methods based on the proposed model could be investigated, such as the CUSUM or EWMA control charts, to better capture the structural changes in

multilayer networks. Second, the SBM can be further extended for generalized multilayer networks where the community memberships can be different across all layers. Third, finding a clever method of combining the local test statistics for each community pair remains a topic for future work. Furthermore, the proposed method inherits the limitations of the SBM in that it assumes the weights on the edges are distributed conditionally independently and thus fails to capture the realistic propensity of the network to form triangles and other structures. The development of versatile models and monitoring methods remains an open challenge.

Acknowledgements

The authors would like to thank the editor and all anonymous referees for their valuable comments, which have helped us improve this work greatly.

Funding

Miss Hang Dong and Dr. Kaibo Wang’s work are supported by the National Natural Science Foundation of China (Key Program) under grant 71731008 and National Key R&D Program of China under grant No. 2017YFF0209400; Dr. Nan Chen’s work is partially supported by Singapore AcRF under grant R-266-000-123-114.

Supplementary Materials

Appendices Derivation of updating steps in the proposed variational-EM algorithm; parameter estimation results of the proposed model in simulation experiments when $K = 3$; confidence interval of MLE for the MZIP distribution; performance of network monitoring when $K = 3$.

Code Code for simulating multi-layer networks based on the proposed model; code for parameter estimation; a toy example of using the code.

References

- Aicher, C., Jacobs, A. Z., and Clauset, A. (2015), “Learning Latent Block Structure in Weighted Networks,” *Journal of Complex Networks*, 3, 221–248. [2]
- 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. [2,5]
- Azarnoush, B., Paynabar, K., and Bekki, J. (2016), “Monitoring Temporal Homogeneity in Attributed Network Streams,” *Journal of Quality Technology*, 48, 28. [3]
- Biernacki, C., Celeux, G., and Govaert, G. (2000), “Assessing a Mixture Model for Clustering With the Integrated Completed Likelihood,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 719–725. [5]
- Cheng, A., and Dickinson, P. (2013), “Using Scan-Statistical Correlations for Network Change Analysis,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Berlin, Heidelberg: Springer, pp. 1–13. [3]
- De Bacco, C., Power, E. A., Larremore, D. B., and Moore, C. (2017), “Community Detection, Link Prediction, and Layer Interdependence in Multilayer Networks,” *Physical Review E*, 95, 042317. [3]
- Erdős, P., and Rényi, A. (1960), “On the Evolution of Random Graphs,” *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5, 17–61. [2]
- Gao, Z.-K., Dang, W.-D., Xue, L., and Zhang, S.-S. (2017), “Directed Weighted Network Structure Analysis of Complex Impedance Measurements for Characterizing Oil-in-Water Bubbly Flow,” *Chaos*, 27, 035805. [1]
- Girvan, M., and Newman, M. E. (2002), “Community Structure in Social and Biological Networks,” *Proceedings of the National Academy of Sciences of the United States of America*, 99, 7821–7826. [1]

- Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010), “A Survey of Statistical Network Models,” *Foundations and Trends^o in Machine Learning*, 2, 129–233. [2]
- Han, Q., Xu, K., and Airoldi, E. (2015), “Consistent Estimation of Dynamic and Multi-layer Block Models,” in *International Conference on Machine Learning*, pp. 1511–1520. [3]
- Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002), “Latent Space Approaches to Social Network Analysis,” *Journal of the American Statistical Association*, 97, 1090–1098. [2]
- Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983), “Stochastic Blockmodels: First Steps,” *Social Networks*, 5, 109–137. [2,4]
- Jun, C., and Shun-zheng, Y. (2009), “Network Monitoring Based on Community Structure Change,” in *2009 International Symposium on Intelligent Ubiquitous Computing and Education, IUCE, IEEE*, pp. 337–340. [3]
- Karrer, B., and Newman, M. E. J. (2011), “Stochastic Blockmodels and Community Structure in Networks,” *Physical Review E*, 83, 016107. [2]
- Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014), “Multilayer Networks,” *Journal of Complex Networks*, 2, 203–271. [1,2,3]
- Li, C.-S., Lu, J.-C., Park, J., Kim, K., Brinkley, P. A., and Peterson, J. P. (1999), “Multivariate Zero-Inflated Poisson Models and Their Applications,” *Technometrics*, 41, 29–38. [4]
- Mariadassou, M., Robin, S., and Vacher, C. (2010), “Uncovering Latent Structure in Valued Graphs: A Variational Approach,” *The Annals of Applied Statistics*, 4, 715–742. [4]
- Miller, K. T., Griffiths, T. L., and Jordan, M. I. (2009), “Nonparametric Latent Feature Models for Link Prediction,” in *International Conference on Neural Information Processing Systems*, pp. 1276–1284. [2]
- Neil, J., Hash, C., Brugh, A., Fisk, M., and Storlie, C. B. (2013), “Scan Statistics for the Online Detection of Locally Anomalous Subgraphs,” *Technometrics*, 55, 403–414. [3]
- Newman, M. E. J., and Park, J. (2003), “Why social Networks Are Different From Other Types of Networks,” *Physical Review E*, 68, 036122. [6,7]
- Nicosia, V., and Latora, V. (2015), “Measuring and Modeling Correlations in Multiplex Networks,” *Physical Review E*, 92, 032805. [1]
- Paul, S., and Chen, Y. (2016), “Consistent Community Detection in Multi-relational Data Through Restricted Multi-layer Stochastic Blockmodel,” *Electronic Journal of Statistics*, 10, 3807–3870. [3]
- Peel, L., and Clauset, A. (2015), “Detecting Change Points in the Large-Scale Structure of Evolving Networks,” in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI’15, AAAI Press*, pp. 2914–2920. [3,11]
- Priebe, C. E., Conroy, J. M., Marchette, D. J., and Park, Y. (2005), “Scan Statistics on Enron Graphs,” *Computational & Mathematical Organization Theory*, 11, 229–247. [3,8,10,11]
- Savage, D., Zhang, X., Yu, X., Chou, P., and Wang, Q. (2014), “Anomaly Detection in Online Social Networks,” *Social Networks*, 39, 62–70. [3,6]
- Wang, Y. X. R., and Bickel, P. J. (2017), “Likelihood-Based Model Selection for Stochastic Block Models,” *The Annals of Statistics*, 45, 500–528. [5]
- Wilson, J. D., Stevens, N. T., and Woodall, W. H. (2016), “Modeling and Estimating Change in Temporal Networks via a Dynamic Degree Corrected Stochastic Block Model,” arXiv no. 1605.04049. [3,7]
- Woodall, W. H., Zhao, M. J., Paynabar, K., Sparks, R., and Wilson, J. D. (2017), “An Overview and Perspective on Social Network Monitoring,” *IIEE Transactions*, 49, 354–365. [3,6]
- Xu, K. S., and Hero, A. O. (2013), “Dynamic Stochastic Blockmodels: Statistical Models for Time-Evolving Networks,” in *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction, Berlin, Heidelberg: Springer*, pp. 201–210. [11]