


Modeling and monitoring multilayer attributed weighted directed networks via a generative model

Hao Wu, Qiao Liang & Kaibo Wang


To cite this article: Hao Wu, Qiao Liang & Kaibo Wang (06 Sep 2023): Modeling and monitoring multilayer attributed weighted directed networks via a generative model, IISE Transactions, DOI: [10.1080/24725854.2023.2256369](https://doi.org/10.1080/24725854.2023.2256369)


To link to this article: <https://doi.org/10.1080/24725854.2023.2256369>

 View supplementary material [↗](#)

 Published online: 06 Sep 2023.

 Submit your article to this journal [↗](#)

 Article views: 193

 View related articles [↗](#)

 View Crossmark data [↗](#)



Modeling and monitoring multilayer attributed weighted directed networks via a generative model

Hao Wu^a, Qiao Liang^b , and Kaibo Wang^{a,c} 

^aDepartment of Industrial Engineering, Tsinghua University, Beijing, China; ^bSchool of Statistics, Southwestern University of Finance and Economics, Chengdu, China; ^cVanke School of Public Health, Tsinghua University, Beijing, China

ABSTRACT

As data with network structures are widely seen in diverse applications, the modeling and monitoring of network data have drawn considerable attention in recent years. When individuals in a network have multiple types of interactions, a multilayer network model should be considered to better characterize its behavior. Most existing network models have concentrated on characterizing the topological structure among individuals, and important attributes of individuals are largely disregarded in existing works. In this article, first, we propose a unified static Network Generative Model (static-NGM), which incorporates individual attributes in network topology modeling. The proposed model can be utilized for a general multilayer network with weighted and directed edges. A variational expectation maximization algorithm is developed to estimate model parameters. Second, to characterize the time-dependent property of a network sequence and perform network monitoring, we extend the static-NGM model to a sequential version, namely, the sequential-NGM model, with the Markov assumption. Last, a sequential-NGM chart is developed to detect shifts and identify root causes of shifts in a network sequence. Extensive simulation experiments show that considering attributes improves the parameter estimation accuracy and that the proposed monitoring method also outperforms the three competitive approaches, static-NGM chart, score test-based chart (ST chart) and Bayes factor-based chart (BF chart), in both shift detection and root cause diagnosis. We also perform a case study with Enron E-mail data; the results further validate the proposed method.

ARTICLE HISTORY

Received 8 March 2023
Accepted 26 August 2023

KEYWORDS

Generative model;
multilayer attributed
network; root cause
diagnosis; statistical process
control

1. Introduction

With the rapid development of information science and data collection technology, the types and structures of available data are becoming increasingly diverse and complex. Rich data creates the possibility of identifying relationships among various entities in a system and provides a powerful ability to describe interactions among these entities. A wide range of applications in supply chain, social science, and computer science (Kivelä *et al.* 2014; Ebrahimi *et al.* 2021) have demonstrated great needs for the modeling and monitoring of data with network structures. Therefore, the modeling and monitoring of network data have drawn much attention in recent years.

Early research on data-driven network modeling was usually limited to a single-layer structure, for example, research by Motalebi, Stevens and Steiner (2021) and Zhang *et al.* (2021). However, single-layer networks can only describe systems with one type of interaction among entities, while it is common that multitype interactions occur in a complex system. Such a multitype interaction mode reflected in a network is a multilayer structure, and each layer represents a type of interaction. For example, in a communication

network, people can communicate by telephone, instant message or E-mail, and each communication method could be considered one layer of connection between individuals. One traditional way to address multilayer networks is to aggregate the interactions from different layers and then analyze the resulting single-layer networks (De Domenico *et al.*, 2013). However, this approach obviously discards considerable layer-specific information about multilayer systems. Additionally, a multilayer network is not a simple combination of several single-layer networks, and there may be a certain correlation between different layers (Nicosia and Latora, 2015). Therefore, constructing a model to properly describe such multilayer structures is important to analyze these systems.

Similar to a single-layer network, a multilayer network also has community structures, and edges connecting individuals in the network can be weighted and directed (Dong *et al.*, 2020). As described in Newman and Park (2003), communities are divided into two main types: assortative communities and disassortative communities. The individuals in the first type of communities interact more densely with the nodes within the communities than the individuals outside the communities, whereas the other type of communities are

contradictory. We take assortative communities into account in our work. Moreover, individuals in a network are generally sparsely connected; that is, an individual frequently interacts with only a few individuals and not the majority. To characterize such multilayer weighted and directed networks, Dong *et al.* (2020) proposed a modeling method based on the Multivariate Zero-Inflated Poisson (MZIP) distribution and Stochastic Block Model (SBM). However, their model may suffer from an identifiability issue (Motalebi, Stevens and Steiner, 2021), which is illustrated in Appendix C.6.

In addition to the network topology structure formed by interactions among individuals, network data sometimes contain important attributes of individuals. For example, in a network that describes the relationships of all the attorneys in a law firm, each node is an attorney whose attributes include age, sex, number of years in the profession, and school of graduation (Liu *et al.*, 2022). It is common to assume that the nodes in the same community generally share similar attributes (Chang *et al.*, 2019).

In single-layer network modeling, some researchers have simultaneously considered individual attributes and network topology to better characterize interactions among individuals (Farahani *et al.*, 2017; Gahrooei and Paynabar, 2018) or to improve the accuracy of community detection (Chang *et al.*, 2019; Chunaev, 2020; Liu *et al.*, 2022). Thus, integrating the information of the network structure and individual attributes can also assist in the identification of network communities, thus improving the accuracy of capturing the overall features of a multilayer network and lays a foundation for its monitoring, which is of great significance. However, to the best of our knowledge, node attributes have not been considered in existing works on the modeling of a multilayer weighted network with sparse connections. Due to the structural complexity of multilayer weighted networks and the data heterogeneity of individual attributes and multilayer network topology structures, it is difficult to take into account individual attributes in multilayer weighted network modeling by simply modifying the existing model.

The focus of this article is to construct a unified model to describe the Multilayer Attributed Weighted Directed Network (MAWDN) and develop a phase-II monitoring method to detect shifts in an MAWDN sequence. The contribution of this article is threefold. First, we propose a unified static Network Generative Model (static-NGM) based on SBM and the Hurdle model, which combines the multilayer network topology and individual attributes and takes into account the community structure through the SBM framework and the sparse connection property via the Hurdle model. Also, the Hurdle model addresses the identifiability of an MZIP model. A Variational Expectation Maximization (VEM) algorithm is applied to perform parameter estimation and learn the in-control (IC) pattern. Second, we extend the static-NGM approach to a sequential version (sequential-NGM) with the Markov assumption to describe the generative process of an MAWDN sequence to characterize the time dependence of the sequence. Third, a monitoring scheme named the sequential-NGM chart based

on sequential-NGM is proposed to monitor anomalies and simultaneously identify root causes of anomalies.

The remainder of this article is organized as follows: Section 2 reviews existing works about network modeling and monitoring. In Section 3, we introduce the method proposed to model and monitor MAWDNs in detail. Section 4 verifies the validity of the model and monitoring scheme by simulation. A real case study is conducted in Section 5. In Section 6, we present a brief conclusion for this study and discuss several possible future directions.

2. Literature review

Due to the availability of network-type data in different applications, the statistical modeling and monitoring of networks has drawn much attention in the literature. In fact, network modeling and monitoring have a close relationship, and we provide a brief review of existing works on them.

The main objective of network monitoring is to detect sudden changes in a network sequence, and some reviews are available (Savage *et al.* 2014; Woodall *et al.* 2017; Noorossana *et al.* 2018; Hewapathirana 2019). Notably, network monitoring has a close connection with statistical process control, and the approaches for it can be divided into feature-based and model-based approaches (Stevens *et al.*, 2021).

Feature-based methods summarize a set of centrality measures from a network, such as degree, betweenness, and closeness centrality, and monitor them using multi-CUSUM or multi-EWMA control charts (McCulloh and Carley, 2011). These methods also define additional complex features for a network by scan statistics and use them to monitor networks (Priebe *et al.*, 2005). Another idea that belongs to this class is the direct monitoring of adjacent matrices or tensors using some of their features, such as eigenvalues (Hazrati-Marangaloo and Noorossana, 2021). However, such methods are not suitable for attributed network monitoring.

Model-based approaches generally utilize a statistical or probability model to explain interaction patterns among nodes and then use control charts to monitor model parameters, likelihood, or prediction residuals. The key to these approaches is constructing proper models that can closely fit the network data.

To describe the attributed networks, one type of model is based on the SBM (Chang *et al.*, 2019; Contisciani *et al.*, 2020; Zhang *et al.*, 2021; Liu *et al.*, 2022), which assumes that the individuals in the same community share the same interaction pattern and similar attributes. However, these models are mostly employed in community detection, but not network monitoring, and edges in networks are generally unweighted.

The other type of model is based on Generalized Linear Models (GLMs). These models link the edge distribution between two individuals with their attributes via GLMs, and Poisson regression and Bernoulli regression can be utilized in integer weighted networks and unweighted networks (Azarnoush *et al.*, 2016; Farahani *et al.*, 2017; Gahrooei and Paynabar, 2018). Considering the sparsity property of

weighted networks, the edge weight distribution between two individuals can be replaced with a Zero-inflated Poisson (ZIP) distribution (Motalebi Owlia, Amiri, and Fallahnezhad, 2023) or a Hurdle model (Ebrahimi *et al.*, 2021). To characterize the naturally time-evolution of a network stream in the context of network monitoring, Gahrooei and Paynabar (2018) and Ebrahimi *et al.* (2021) used a state space model on the regression coefficients of GLMs and monitored the network stream via one-step prediction residuals. Differently, Azarnoush *et al.* (2016) considered a dynamic-reference sliding window approach and performed monitoring via the likelihood ratio test. However, this type of model cannot be easily applied in the field of multilayer network modeling and monitoring.

Latent space models are another type of method to model network systems; they map the individuals of the network into a low-dimensional real space and represent them through some real scalars or vectors (Zou and Li, 2017; Sun *et al.*, 2020; Wang and Xie, 2021; Lee *et al.*, 2022). However, these models cannot be applied to model a weighted attributed network, especially a multilayer network.

Recall that one key feature of networks that we consider is that they have multiple layers. Kivelä *et al.* (2014) provided a detailed description of the history of multilayer networks and some terminologies. Han *et al.* (2015) and Paul and Chen (2016) constructed models based on the SBM to characterize multilayer networks with binary edges. The model of Contisciani *et al.* (2020) is similar, but considers individual attributes. To address weighted edges, De Bacco *et al.* (2017) further applied a Poisson distribution to model edge weight, and the basis of this model is the mixed membership SBM. Dong *et al.* (2020) noted that multilayer networks are sparse and emphasized that some correlations exist between different layers. Thus, the authors proposed an SBM based on Multivariate Zero Inflated Poisson Distribution (MZIP-SBM) to describe a multilayer network and constructed a control chart using score statistics to test whether a multilayer network is generated from the model with IC parameters by MZIP-SBM, however, they assumed that community labels are known as prior knowledge.

Although there are a lot of methods to model network systems and monitoring a network sequence based on the model, methods for modeling MANWDN and characterizing the time-dependent property of a multilayer network are still lacking. Therefore, methods for monitoring a MAWDN sequence and identifying the key factor that leads to a sudden change are also lacking. To address these issues, we propose a unified generative model to integrate both the network topology and attributes and extend it to a sequential version to model a network sequence. Then, the sequential-NGM chart scheme is proposed to simultaneously detect changes and identify the root causes of changes.

3. Methodology

In this section, first, we introduce the MAWDN representation. Second, we propose a static-NGM model to model an MAWDN and develop a VEM algorithm to estimate the IC

parameters offline. Third, we introduce a sequential version of static-NGM, i.e., sequential-NGM, to model an MAWDN sequence and introduce the monitoring and diagnosis scheme based on this model. Last, we provide important remarks on model implementation.

3.1. MAWDN representation

We extend the representation provided in Dong *et al.* (2020). Systematically, we refer to individuals in a network as nodes hereafter. We consider an MAWDN $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{W}\}$ with L layers, where $\mathcal{V} = \{v_i\}_{i=1, \dots, N}$ is a node set consisting of N nodes present at each layer of the network, $\mathcal{W} = \{\mathcal{W}^1, \dots, \mathcal{W}^N\}$ is the attribute set for all the nodes in the network, with \mathcal{W}^i representing all the attributes possessed by node i , and $\mathcal{E} = \{\mathcal{E}^1, \dots, \mathcal{E}^L\}$ represents an edge set consisting of all directed edges in the network, with \mathcal{E}^l consisting of all the edges at layer l .

For an MAWDN, we have its adjacent tensor \mathcal{A} and attribute matrix \mathbf{W} . \mathcal{A} is in dimension $N \times N \times L$, whose element n_{ijl} represents the number of interactions from node i to node j at layer l , and thus is integer-valued. Loops are not allowed in the context of this article. \mathbf{W} is in dimension $N \times T$, whose element w_{it} is binary, and $w_{it} = 1$ represents node i possessing attribute t , while $w_{it} = 0$, otherwise.

It is true that some attributes are not binary, even not discrete, but we can discretize them first and transform them into binary variables (Liu *et al.*, 2022). Hence, we only consider binary attribute in this work. We use an MAWDN with two layers, 15 nodes and 18 node attributes as an example. In this network, community 1 is strongly related to attributes 1–6, community 2 is strongly related to attributes 7–12 and community 3 is strongly related to the remaining 6 attributes. Figure 1(a) provides the topology of this network, the darker edge represents the greater edge weight, and Figure 1(b) is its node attribute matrix, the shaded position represents a 1 and the blank position represents a 0.

3.2. Static-NGM model and offline learning

In this section, we propose a static MAWDN generative model to describe the IC pattern. In the proposed model, we follow the idea of a SBM and assume that the nodes in one network can be partitioned into R communities, and that communities are disjoint with each other, we assume R is known in this work. For each node i , a categorical variable $c_i \in \{1, \dots, R\}$ represents its community ID, $c_i = r$ if node i belongs to community r . We follow the assumption proposed by Dong *et al.* (2020) that all network layers share the same community structure.

Moreover, we only consider the attributes that are highly related to communities and network independent. Selecting the coauthor network constructed by Pandhre *et al.* (2016) as an example, nodes represent authors, and an undirected edge exists between two nodes if the two authors represented by the two nodes are coauthors of an article. The attributes for a node is a vector that records the number of papers published in 20 different journals or conferences.

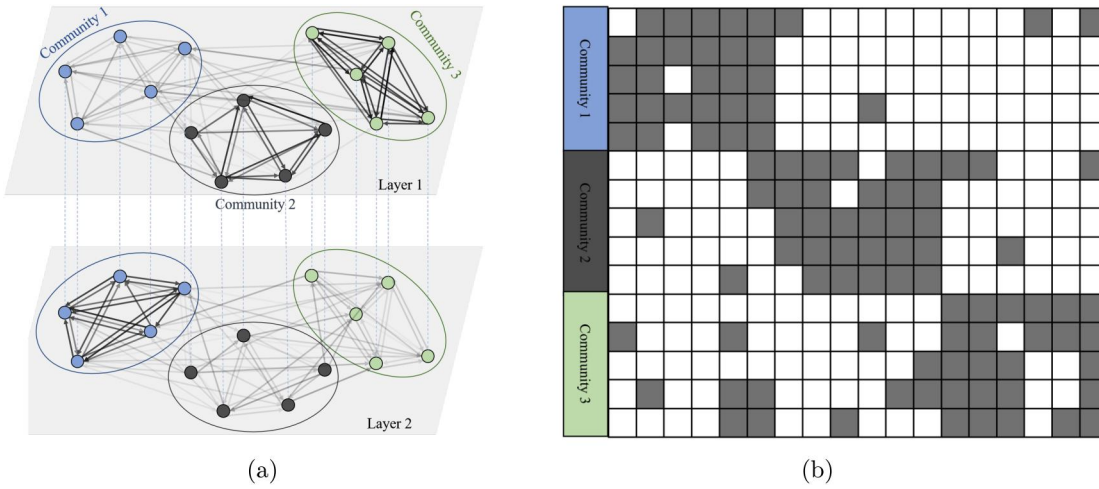


Figure 1. Illustration example of MAWDN. (a) The topology of the MAWDN, darker edges means greater edge weights. (b) The node attribute matrix of the MAWDN, each shaded cells represents a “1” and an empty cells represents a “0”.

Intuitively, we treat different research fields as different communities. Generally, authors in different communities select different journals or conferences to publish their works, whereas authors in the same communities have similar selections. Thus, the attributes in the coauthor network are highly related to communities. On the other hand, since the main journals and conferences in each research field are relatively fixed, the attribute distribution in different networks are considered the same, i.e., network independence. We further explain network independence in Section 3.2.1.

3.2.1. Model formulation

We consider a set of MAWDNs $\{\mathcal{G}_g, g = 1, \dots, G\}$. Given an index $g, g \in \{1, \dots, G\}$, MAWDN \mathcal{G}_g consists of N nodes, and we have its adjacent tensor \mathcal{A}^g and attribute matrix W^g . We here model the generation process of \mathcal{A}^g and W^g . Following the common assumptions of SBM models, we believe that the interaction behaviors among nodes and the attribute distributions of nodes both only depend on their latent community ID. For each MAWDN \mathcal{G}_g , we use a vector γ^g with dimension R to define a multinomial distribution that describes the community distribution of its nodes. For each pair of community $(r, s) \in \{1, \dots, R\}^2$, we use π_{rs}^g and θ_{rs}^g to define a Hurdle model, which can characterize the sparsity property, to describe the distribution of total edge number from a node in community r to a node in community s . Here, π_{rs}^g is the probability that the edge weight from a node in community r to a node in community s is deterministically set to zero, and θ_{rs}^g defines a positive Poisson distribution to describe the total edge number if it is greater than zero. The probability mass function of the Hurdle model is given by Equation (5).

We use a vector φ_{rs}^g with dimension L to define a multinomial distribution that describes the probability of a layer to which an edge from a node in community r to a node in community s in MAWDN g belongs. Meanwhile, for each community $r \in \{1, \dots, R\}$, we use a vector μ_r with dimension T to define a multinomial distribution to describe its attribute distribution. Please note that in a mathematical

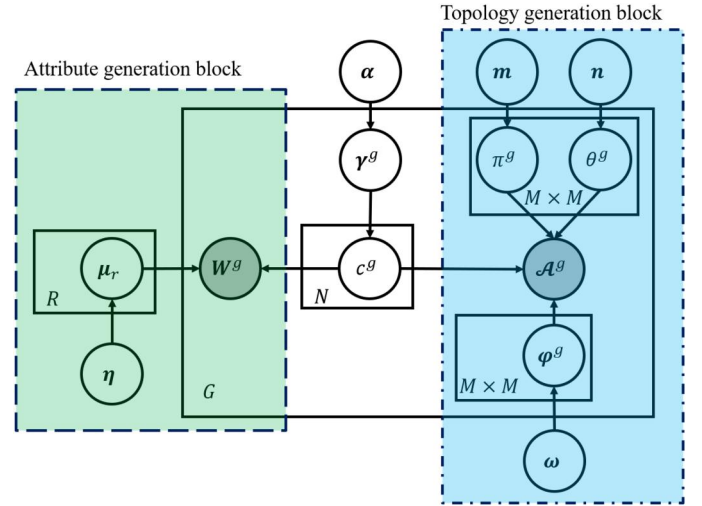


Figure 2. Graphical representation of static-NGM model.

view, network independent attribute distribution assumption is equivalent to that μ_r is independent of networks indices g .

Based on the definitions above, the MAWDN set $\{\mathcal{G}_g, g = 1, \dots, G\}$ generation process in the static-NGM model is shown below, and Figure 2 is the corresponding graphical representation.

- For each community $r \in \{1, 2, \dots, R\}$:
 - Draw a multinomial attribute distribution with parameter vector $\mu_r \sim \text{Dirichlet}(\eta_r)$.
- For each network G_g :
 - Draw a multinomial community distribution with parameter $\gamma^g \sim \text{Dirichlet}(\alpha)$
- For each pair of community (r, s) :
 - Draw a Hurdle parameter $\pi_{rs}^g \sim \text{Beta}(m_{rs,1}, m_{rs,2})$
 - Draw a Hurdle parameter $\theta_{rs}^g \sim \text{Gamma}(n_{rs,1}, n_{rs,2})$
 - Draw a multinomial edge layer distribution with parameter vector $\varphi_{rs}^g \sim \text{Dirichlet}(\omega_{rs})$
- For each node i in the network G_g :
 - Draw a community ID $c_i^g \sim \text{multinomial}(\gamma^g)$
- For each attribute:
 - Draw a value $t \sim \text{multinomial}(\mu_{c_i^g})$

- For each pair of nodes $(i, j), i \neq j$:
 - Draw a value from Bernoulli $(\pi_{c_i^g, c_j^g}^g)$ to indicate whether $n_{ij}^g > 0$
 - if $n_{ij}^g > 0$, draw the total number of edges $n_{ij}^g \sim \text{Positive Poisson}(\theta_{c_i^g, c_j^g}^g)$ from node i to node j
 - if $n_{ij}^g > 0$, for each edge $e_{ij, k}^g, k = 1, \dots, n_{ij}^g$:
 - draw a layer $e_{ij, k}^g \sim \text{multinomial}(\varphi_{c_i^g, c_j^g}^g)$

This model mainly includes two blocks, namely, an attribute generation block (shown in rectangle with green shade and dash-line sides in Figure 2) and a topology generation block (shown in rectangle with blue shade and dot-and-dash-line sides in Figure 2). Using this model, we integrate the network node attributes and network topology, and depict the complex MAWDN data through a set of model parameters $\beta_1 = \{\gamma, \mu, \Theta, \Pi, \varphi\}$, where $\mu = \{\mu_r\}_{r=1, \dots, R}$, $\Theta = \{\theta_{rs}\}_{r, s=1, \dots, R}$, $\Pi = \{\pi_{rs}\}_{r, s=1, \dots, R}$ and $\varphi = \{\varphi_{r, s=1, \dots, R}\}$. We propose a VEM algorithm to estimate these model parameters in Section 3.2.2, $\beta_2 = \{\alpha, m, n, \omega, \eta\}$ are hyperparameters that can be set according to some prior information (see Appendix C for detailed guidelines).

3.2.2. Parameter estimation

We denote that $\mathcal{A} = \{\mathcal{A}^g\}_{g=1, \dots, G}$, $\mathbf{W} = \{\mathbf{W}^g\}_{g=1, \dots, G}$ and that $\mathbf{c} = \{\mathbf{c}^g\}_{g=1, \dots, G}$, where $\mathbf{c}^g = [c_1^g, c_2^g, \dots, c_N^g]$, let $I_i^g(r) = I(c_i^g = r) = 1$ if $c_i^g = r$ and $I_i^g(r) = 0$ otherwise. Based on the description above, we need to estimate the model parameters β_1 given the MAWDN data and hyperparameters β_2 and to infer the community ID of each node in each network. The likelihood function corresponding to the proposed model is

$$\begin{aligned} P(\mathcal{A}, \mathbf{W} | \beta_1) &= \sum_{\mathbf{c}} P(\mathcal{A}, \mathbf{W}, \mathbf{c} | \beta_1) \\ &= \sum_{\mathbf{c}} P(\mathcal{A} | \mathbf{c}, \Pi, \Theta, \varphi) P(\mathbf{W} | \mathbf{c}, \mu) P(\mathbf{c} | \gamma) \end{aligned} \quad (1)$$

Where

$$\begin{aligned} P(\mathcal{A} | \mathbf{c}, \Pi, \Theta, \varphi) &= \prod_{g=1}^G \prod_{i \neq j} \prod_{r, s} \prod_{l=1}^L \\ &\quad \left(P(n_{ij, l}^g | \varphi_{rs}^g, n_{ij}^g) P(n_{ij}^g | \pi_{rs}^g, \theta_{rs}^g) \right)^{I_i^g(r) I_j^g(s)} \end{aligned} \quad (2)$$

$$\begin{aligned} P(\mathbf{W} | \mathbf{c}, \mu) &= \prod_{g=1}^G \prod_{i=1}^N \prod_{r=1}^R \prod_{t=1}^T P(w_{it}^g | \mu_r) I_i^g(r) \\ &= \prod_{g=1}^G \prod_{i=1}^N \prod_{r=1}^R \prod_{t=1}^T (\mu_{rt})^{I_i^g(r) w_{it}^g} \end{aligned} \quad (3)$$

$$P(n_{ij}^g | \pi_{rs}^g, \theta_{rs}^g) = \begin{cases} \pi_{rs}^g & \text{if } n_{ij}^g = 0 \\ (1 - \pi_{rs}^g) \frac{(\theta_{rs}^g)^{n_{ij}^g} \exp(-\theta_{rs}^g)}{n_{ij}^g! (1 - \exp(-\theta_{rs}^g))} & \text{if } n_{ij}^g > 0 \end{cases} \quad (4)$$

$$P(\mathbf{c} | \gamma) = \prod_{g=1}^G \prod_{i=1}^N P(c_i^g | \gamma^g) = \prod_{g=1}^G \prod_{i=1}^N \prod_{r=1}^R (\gamma_r^g)^{I_i^g(r)} \quad (5)$$

We perform estimation and inference under a Bayesian framework. The key problem we need to address is computing the joint posterior distribution of \mathbf{c} and β_1 given the observation MAWDN data and prior distributions, that is

$$P(\beta_1, \mathbf{c} | \mathcal{A}, \mathbf{W}, \beta_2) = \frac{P(\beta_1, \mathbf{c}, \mathcal{A}, \mathbf{W} | \beta_2)}{\sum_{\mathbf{c}} \int_{\beta_1} P(\mathbf{c}, \mathcal{A}, \mathbf{W} | \beta_1) P(\beta_1 | \beta_2) d\beta_1} \quad (6)$$

Because the integration in the denominator of the right-hand of Equation (6) is in high complexity, obtaining a closed-form expression for this posterior distribution is intractable. Hence, we propose a VEM algorithm to find a variational distribution $Q(\beta_1, \mathbf{c})$ with an easy form to approximate $P(\beta_1, \mathbf{c} | \mathcal{A}, \mathbf{W}, \beta_2)$. The objective is minimizing the Kullback–Leiber (KL) divergence (Kullback, 1997) between $Q(\beta_1, \mathbf{c})$ and $P(\beta_1, \mathbf{c} | \mathcal{A}, \mathbf{W}, \beta_2)$, that is,

$$Q^*(\beta_1, \mathbf{c}) = \arg \min_{Q \in \mathcal{F}} D_{KL}(Q(\beta_1, \mathbf{c}) || P(\beta_1, \mathbf{c} | \mathcal{A}, \mathbf{W}, \beta_2)) \quad (7)$$

where \mathcal{F} is a given variational distribution family.

In our work, we consider a mean field distribution family, i.e., all the distributions are independent of each other, then $Q(\beta_1, \mathbf{c})$ has a form:

$$\begin{aligned} Q(\beta_1, \mathbf{c}) &= \prod_{r=1}^R q(\mu_r | \delta_r^\mu) \prod_{g=1}^G q(\gamma^g | \mathbf{v}^g) \prod_{i=1}^N q(c_i^g | \tau_i^g) \prod_{r, s} q(\pi_{rs}^g | \delta_{rs}^{\pi g}) \\ &\quad q(\theta_{rs}^g | \delta_{rs}^{\theta g}) q(\varphi_{rs}^g | \delta_{rs}^{\varphi g}) \end{aligned} \quad (8)$$

where $\beta_3 = \{\mathbf{v}, \tau, \delta^\mu, \delta^\pi, \delta^\theta, \delta^\varphi\}$ is the variational parameter set. Additionally, we restrict the form of each variational distribution as (9):

$$\begin{aligned} q(\gamma^g | \mathbf{v}^g) &= \text{Dirichlet}(\mathbf{v}^g); & q(c_i^g | \tau_i^g) &= \text{multinomial}(\tau_i^g) \\ q(\mu_r | \delta_r^\mu) &= \text{Dirichlet}(\delta_r^\mu); & q(\pi_{rs}^g | \delta_{rs}^{\pi g}) &= \text{Beta}(\delta_{rs, 1}^{\pi g}, \delta_{rs, 2}^{\pi g}); \\ q(\theta_{rs}^g | \delta_{rs}^{\theta g}) &= \text{Gamma}(\delta_{rs, 1}^{\theta g}, \delta_{rs, 2}^{\theta g}); & q(\varphi_{rs}^g | \delta_{rs}^{\varphi g}) &= \text{Dirichlet}(\delta_{rs}^{\varphi g}) \end{aligned} \quad (9)$$

By Jensen's inequality, we can obtain the lower bound of the log-likelihood function under Bayesian framework, which is always named as Evidence Lower Bound (ELBO):

$$\begin{aligned} \log P(\mathcal{A}, \mathbf{W} | \beta_2) &\geq \mathbb{E}_Q[\log P(\beta_1, \mathbf{c}, \mathcal{A}, \mathbf{W} | \beta_2)] \\ &\quad - \mathbb{E}_Q[\log Q(\beta_1, \mathbf{c})] \triangleq \text{ELBO}(\beta_3) \end{aligned} \quad (10)$$

By Equations (7) and (10), we can obtain the relationship among the log-likelihood function, KL-divergence between $Q(\beta_1, \mathbf{c})$ and $P(\beta_1, \mathbf{c} | \mathcal{A}, \mathbf{W}, \beta_2)$, and ELBO:

$$\log P(\mathcal{A}, \mathbf{W} | \beta_2) - \text{ELBO}(\beta_3) = D_{KL}(Q(\beta_1, \mathbf{c}) || P(\beta_1, \mathbf{c} | \mathcal{A}, \mathbf{W}, \beta_2)) \quad (11)$$

which indicates that minimizing KL-divergence is equivalent to maximizing ELBO given prior distributions and observed MAWDN data. Therefore, we here optimize variational parameters by maximizing the ELBO. To achieve this,

we split β_3 into $\beta_{31} = \{\tau\}$ and $\beta_{32} = \{v, \delta^\mu, \delta^\pi, \delta^\theta, \delta^\varphi\}$, and adopt an iterative procedure shown in Equation (12) to update these two set of parameters until it converges

$$\begin{cases} \beta_{31}^{(r+1)} = \operatorname{argmax}_{\beta_{31}} ELBO(\beta_{31}, \beta_{32}^{(r)}) \\ \beta_{32}^{(r+1)} = \operatorname{argmax}_{\beta_{32}} ELBO(\beta_{31}^{(r+1)}, \beta_{32}) \end{cases} \quad (12)$$

After some derivation, we can obtain the update formula for each component in β_3 . Given β_{32} , the optimal τ meets the fixed point relation (13):

$$\begin{aligned} \tau_{ir}^g \propto \exp \left\{ \sum_{j \neq i} \sum_{r=1}^R \tau_{js}^g \left(T_{ij}^{rs}(g) + T_{ji}^{sr}(g) \right) + w_{it}^g \sum_{t=1}^T \mathbb{E}_Q[\log \mu_{rt}] \right. \\ \left. + \mathbb{E}_Q[\log \gamma_r^g] \right\} \end{aligned} \quad (13)$$

Thus, we can get optimal τ by iterating the relation above until convergence. Define indicator function $I(n_{ij}^g = 0) = 1$ if $n_{ij}^g = 0$ and $I(n_{ij}^g = 0) = 0$, otherwise. For simplicity, we denote $I(n_{ij}^g = 0)$ as I_{hij}^- and let $I_{gij}^+ = 1 - I_{hij}^-$. Then, given τ , the optimal value for elements in β_{32} are:

$$\nu_r^g = \sum_{i=1}^N \tau_{ir}^g + \alpha_r; \quad \delta_{rt}^\mu = \sum_{g=1}^G \sum_{i=1}^N \tau_{ir}^g w_{it}^g + \eta_{rt} \quad (14a)$$

$$\begin{aligned} \delta_{rs,1}^{\pi g} &= \sum_{i=1}^N \sum_{j=1, j \neq i}^N \tau_{ir}^g \tau_{js}^g I_{gij}^- + m_{rs,1}; \\ \delta_{rs,2}^{\pi g} &= \sum_{i=1}^N \sum_{j=1, j \neq i}^N \tau_{ir}^g \tau_{js}^g I_{gij}^+ + m_{rs,2} \end{aligned} \quad (14b)$$

$$\delta_{rs}^{\theta g*} = \operatorname{argmax}_{\delta_{rs}^{\theta g}} L(\delta_{rs}^{\theta g}); \quad \delta_{rs,l}^{\varphi g} = \sum_{i=1}^N \sum_{j=1, j \neq i}^N \tau_{ir}^g \tau_{js}^g I_{gij}^+ n_{ij,l}^g + \omega_{rs,l} \quad (14c)$$

We then use means of variational posterior distributions as point estimations of model parameters:

$$\begin{aligned} \hat{\gamma}_r^g &= \frac{\nu_r^g}{\sum_{r=1}^R \nu_r^g}, \hat{\mu}_{rt} = \frac{\delta_{rt}^\mu}{\sum_{h=1}^T \delta_{rt}^\mu}, \hat{\pi}_{rs}^g = \frac{\delta_{rs,2}^{\pi g}}{\delta_{rs,1}^{\pi g} + \delta_{rs,2}^{\pi g}} \\ \hat{\theta}_{rs}^g &= \frac{\delta_{rs,1}^{\theta g}}{\delta_{rs,2}^{\theta g}}, \hat{\varphi}_{rs,l}^g = \frac{\delta_{rs,l}^{\varphi g}}{\sum_{l=1}^L \delta_{rs,l}^{\varphi g}} \end{aligned} \quad (15)$$

We summarize this algorithm in Algorithm 1 in Appendix E.2. More details about the derivation process and the closed-form expressions of expectations are provided in Appendix A. We can observe that the τ “softly” estimates the community ID of each node, and the optimal values of the other variational parameters totally depend on the optimal value of τ . “Softly” means that the τ provides the distribution of community ID of each node, but does not give the exact community to which each node belongs. Therefore, the more accurate the community estimation is, the more accurate the model parameter estimation is. In other words, if a node i belongs to community r , it is better to get the result that $\tau_{ir} \approx 1$ and $\tau_{is} \approx 0, s \neq r$. Thus, the

structure of our model and parameter estimation process also indicate the network community detection using network topology and node attributes. However, these two absolutely different types of data can bring different information gains to community detection and the two corresponding terms in the likelihood function can be on different scales. To better balance the contributions of the two types of data and improve the accuracy of community detection, we consider using weight likelihood function (Liang and Wang, 2022) to replace the original likelihood function (1) by introducing a weight factor σ .

$$\begin{aligned} P(\mathcal{A}, W | \beta_1) &= \sum_{\mathbf{c}} P(\mathcal{A}, W, \mathbf{c} | \beta_1) \\ &= \sum_{\mathbf{c}} P(\mathcal{A} | \mathbf{c}, \mathbf{\Pi}, \mathbf{\Theta}, \boldsymbol{\varphi}) P(W | \mathbf{c}, \boldsymbol{\mu})^\sigma P(\mathbf{c} | \boldsymbol{\gamma}) \end{aligned} \quad (16)$$

Using the almost same derivation process, we get the new update formula for τ and δ^μ

$$\begin{aligned} \tau_{ir}^g \propto \exp \left\{ \sum_{j \neq i} \sum_{r=1}^R \tau_{js}^g \left(T_{ij}^{rs}(g) + T_{ji}^{sr}(g) \right) \right. \\ \left. + \sigma w_{it}^g \sum_{t=1}^T \mathbb{E}_Q[\log \mu_{rt}] + \mathbb{E}_Q[\log \gamma_r^g] \right\} \end{aligned} \quad (17)$$

$$\delta_{rt}^\mu = \sigma \sum_{g=1}^G \sum_{i=1}^N \tau_{ir}^g w_{it}^g + \eta_{rt} \quad (18)$$

The update formulas for the other variational parameters remain unchanged. σ is the weight coefficient of node attributes relative to network topology, $\sigma = 0$ means ignoring the information in node attributes, $\sigma \rightarrow \infty$ means using information in node attributes only. We provide a guideline to select a proper σ in Appendix C.1.

We note that all the optimal variational parameters related to the Beta distribution and Dirichlet distribution can be viewed as the sum of the prior pseudo count and the real count in MAWDN, for instance, ν_r^g is the sum of the prior pseudo count of nodes in community r and real number of nodes belonging to community r in MAWDN g . Such findings make our model more explanatory and valid. Another note is that, by estimating the attribute distribution $\boldsymbol{\mu}_r$ for community $r \in \{1, \dots, R\}$, we want to identify the attributes that the community r is strongly dependent (Chang *et al.*, 2019; Liu *et al.*, 2022) with, namely, the attributes with a larger μ_{rt} . Consequently, we are not particularly bothered about the exact value of μ_{rt} because it does not affect the estimated value of the other parameters, but rather focus on the attributes with relatively large μ_{rt} , which can help us to efficiently infer the node community, so as to accurately estimate the model parameters.

3.3. Sequential-NGM model and online monitoring

In this stage, the IC parameters $\boldsymbol{\gamma}^{(0)}$, $\mathbf{\Pi}^{(0)}$, $\mathbf{\Theta}^{(0)}$, and $\boldsymbol{\varphi}^{(0)}$, are known. Also, the attribute distribution $\boldsymbol{\mu}$ for each community is known. They can be estimated offline from IC MAWDNs using static-NGM model in Section 3.2.

We further assume that μ does not change over the observation time. This assumption is reasonable, due to the network independent property of node attributes. Even if there is a shift of μ , we can treat it as the appearance of new communities, which is do not considered in this work. To realize the aim of monitoring an MAWDN sequence online, we extend the static-NGM model to its sequential version (sequential-NGM) to describe the sequential generative process of an MAWDN sequence. Through estimating and monitoring the model parameters at each observation time, we can control the network sequence.

3.3.1. Model formulation

We define a sequence of MAWDNs $\mathcal{G}_g, g = 1, 2, \dots, G$, where \mathcal{G}_g is the observed MAWDN at time g , represented by its adjacent tensor \mathcal{A}^g and attribute matrix \mathbf{W}^g . To model the time dependent property of the network sequence, we include the Markov assumption in the sequential-NGM model, i.e., the prior distributions of model parameters φ, γ, Π and Θ at time g only depend on those at time $g - 1$. Specifically, the model parameters at time 1 are generated from the prior distribution defined by the target parameters $\gamma^{(0)}, \Pi^{(0)}, \Theta^{(0)}$ and $\varphi^{(0)}$. The sequential generative process of MAWDN sequence $\mathcal{G}_g, g = 1, 2, \dots, G$ is described as below (see Figure 3 for its graphical representation):

- For each network \mathcal{G}_g :
 - Draw a multinomial community distribution with parameter vector $\gamma^g \sim \text{Dirichlet}(\rho N \gamma^{g-1})$
- For each pair of community (r, s) :
 - Draw a parameter $\pi_{rs}^g \sim \text{Beta}(\rho N(N-1) \gamma_r^{g-1} \gamma_s^{g-1} \pi_{rs}^{g-1}, \rho N(N-1) \gamma_r^{g-1} \gamma_s^{g-1} (1 - \pi_{rs}^{g-1}))$.
 - Draw a parameter $\theta_{rs}^g \sim \text{Gamma}(\rho N(N-1) \gamma_r^{g-1} \gamma_s^{g-1} (1 - \pi_{rs}^{g-1}) \theta_{rs}^{g-1}, \rho N(N-1) \gamma_r^{g-1} \gamma_s^{g-1} (1 - \pi_{rs}^{g-1}))$.
 - Draw a multinomial edge layer distribution with parameter vector $\varphi_{rs}^g \sim \text{Dirichlet}(\rho N(N-1) \gamma_r^{g-1} \gamma_s^{g-1} (1 - \pi_{rs}^{g-1}) \varphi_{rs}^{g-1})$
- For each node i in the network \mathcal{G}_g :

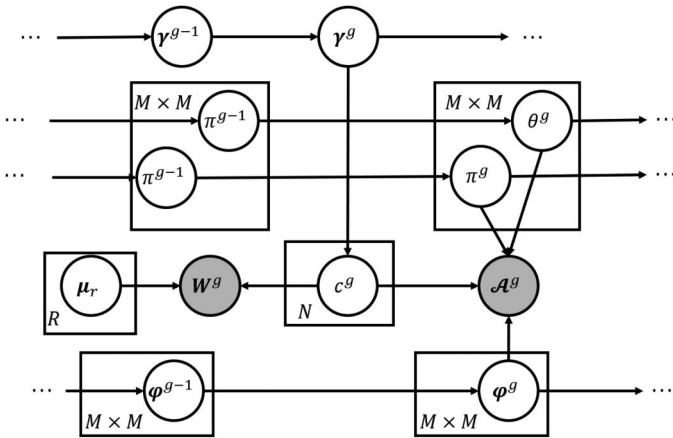


Figure 3. Graphical representation of sequential-NGM model.

- Draw a community ID $c_i^g \sim \text{multinomial}(\gamma^g)$
- for each attribute of node i :

*Draw a value $t \sim \text{Dirichlet}(\mu_{c_i^g}^g)$

- For each pair of nodes (i, j) :
 - Draw a value from Bernoulli $(\pi_{c_i^g, c_j^g}^g)$ to indicate whether $n_{ij}^g > 0$
 - if $n_{ij}^g > 0$, Draw the total number of edges $n_{ij}^g \sim \text{Positive Poisson}(\theta_{c_i^g, c_j^g}^g)$ from node i to node j
 - if $n_{ij}^g > 0$, for each edge $e_{ij,k}^g$:
- *Draw a layer $e_{ij,k}^g \sim \text{multinomial}(\varphi_{c_i^g, c_j^g}^g)$

where ρ is a smoothing factor that adjust the influence of the prior. A larger ρ means a stronger influence of the last network on the present network. See Appendix C.1 for the discussion of selecting ρ .

According to the above model structure, given the parameters at time $g - 1$ and the MAWDN representation $\{\mathcal{A}^g, \mathbf{W}^g\}$, we can estimate the parameter at time g .

3.3.2. Parameter estimation

We use the VEM algorithm as we do in Section 3.2.2 to perform parameter estimation. The variational distribution at time g is

$$Q^g(\gamma^g, \Pi^g, \Theta^g, \varphi^g, \mathbf{c}) = q(\gamma^g | \mathbf{v}^g) \prod_{i=1}^N q(c_i^g | \tau_i^g) \prod_{r,s} q(\pi_{rs}^g | \delta_{rs}^g) q(\theta_{rs}^g | \delta_{rs}^g) q(\varphi_{rs}^g | \delta_{rs}^g) \quad (19)$$

Note that μ is known, we do not include the terms related to it in Q^g . By some similar derivation to it in Appendix A, we obtain the update formula for each variational parameter.

We also use the mean of the variational distributions as the estimation results of model parameters:

$$\hat{\gamma}_r^g = \frac{\nu_r^{g*}}{\sum_{r=1}^R \nu_r^{g*}} = \frac{1}{1 + \rho} \times \frac{\sum_{i=1}^N \tau_{ir}^g}{N} + \frac{\rho}{1 + \rho} \times \gamma_r^{g-1} \quad (20a)$$

$$\begin{aligned} \hat{\pi}_{rs}^g &= \frac{\delta_{rs,2}^g}{\delta_{rs,1}^g + \delta_{rs,2}^g} \\ &= \frac{\sum_{i=1}^N \sum_{j=1, j \neq i}^N \tau_{ir}^g \tau_{js}^g I_{ij}^g + \rho N(N-1) \gamma_r^{g-1} \gamma_s^{g-1} \pi_{rs}^{g-1}}{\sum_{i=1}^N \sum_{j=1, j \neq i}^N \tau_{ir}^g \tau_{js}^g + \rho N(N-1) \gamma_r^{g-1} \gamma_s^{g-1}} \end{aligned} \quad (20b)$$

$$\begin{aligned} &\approx \frac{1}{1 + \rho} \times \frac{\sum_{i=1}^N \sum_{j=1, j \neq i}^N \tau_{ir}^g \tau_{js}^g I_{ij}^g}{\sum_{i=1}^N \sum_{j=1, j \neq i}^N \tau_{ir}^g \tau_{js}^g} + \frac{\rho}{1 + \rho} \times \pi_{rs}^{g-1} \\ &\hat{\theta}_{rs} = \frac{\delta_{rs,1}^g}{\delta_{rs,2}^g} \end{aligned} \quad (20c)$$

$$\hat{\varphi}_{rs,l}^g = \frac{\sum_{i=1}^N \sum_{j=1, j \neq i}^N \tau_{ir}^g \tau_{js}^g I_{ij,l}^g + \rho N(N-1) \gamma_r^{g-1} \gamma_s^{g-1} (1 - \pi_{rs}^{g-1}) \theta_{rs}^{g-1} \varphi_{rs,l}^{g-1}}{\sum_{l=1}^L \left(\sum_{i=1}^N \sum_{j=1, j \neq i}^N \tau_{ir}^g \tau_{js}^g I_{ij,l}^g + \rho N(N-1) \gamma_r^{g-1} \gamma_s^{g-1} (1 - \pi_{rs}^{g-1}) \theta_{rs}^{g-1} \varphi_{rs,l}^{g-1} \right)}$$

$$\approx \frac{1}{1+\rho} \times \frac{\sum_{i=1}^N \sum_{j=1, j \neq i}^N \tau_{ir}^g \tau_{js}^g I_{ij,l}^g}{\sum_{i=1}^N \sum_{j=1, j \neq i}^N \tau_{ir}^g \tau_{js}^g I_{ij}^g} + \frac{\rho}{1+\rho} \times \varphi_{rs,l}^{g-1} \quad (20d)$$

In Equation (20), we notice that the estimation results of γ^g , Π^g and φ^g can be approximately viewed as in a EWMA type, where $\frac{1}{1+\rho}$ is equivalent to λ in EWMA.

3.3.3. Monitoring and root cause diagnosis scheme

The main objective of MAWDN monitoring and diagnosis is detecting the sudden change of an MAWDN sequence and diagnosing the root cause of it. To achieve this objective, we propose a four-chart control scheme and name it as the sequential-NGM chart.

Considering the fact that the feature of an MAWDN can be reflected by the model parameters $\beta = \{\gamma, \Pi, \Theta, \varphi\}$, we can conduct the following hypothesis test to check whether there is a shift at time g :

$$H_0 : \beta_g = \beta^{(0)}; H_1 : \beta_g \neq \beta^{(0)} \quad (21)$$

where $\beta_g = \{\gamma^g, \Pi^g, \Theta^g, \varphi^g\}$ and $\beta^{(0)} = \{\gamma^{(0)}, \Pi^{(0)}, \Theta^{(0)}, \varphi^{(0)}\}$. One note is that the networks may slowly evolve with some pattern in some systems, and the objective in such a situation is detecting the abrupt shifts, but ignoring the natural evolution. To tackle this situation, Azarnoush *et al.* (2016) considered using a sliding window of reference networks that is updated dynamically. In this method, they dynamically updated the IC parameters by estimating them using the sliding window of reference networks at each time point. In our work, an idea inspired by Azarnoush *et al.* (2016) can also be used to deal with the natural evolution. A detailed discussion is provided in Appendix D.3. There are four types of shift that can be reflected by these parameters. Given an MAWDN sequence:

1. the first type is change of γ , which means a shift of community distribution;
2. the second type is change of Π , which indicates the shift of overall interaction probability among nodes;
3. the third type is change of Θ , which reflects a shift of overall interaction intensity among nodes;
4. and the last type is change of φ , which is an indication of the shift of the interaction type distribution.

To perform the test above, we can compare the distributions parameterized by β_g and distributions parameterized by $\beta^{(0)}$. The KL-divergence can be applied to measure the distance between two distributions. To simultaneously detect the change point and diagnose which one of the four types of shift leads to the change, we monitor the four types of parameters individually, and the charting statistics are defined as:

$$Q_g^\gamma = 2N(1+\rho) \times D_{KL}(P(c|\hat{\gamma}^g), P(c|\gamma^{(0)}))$$

$$= 2N(1+\rho) \times \sum_{r=1}^R \gamma_r^g \log \left(\frac{\hat{\gamma}_r^g}{\hat{\gamma}_r^{(0)}} \right) \quad (22a)$$

$$Q_g^\pi = 2 \sum_{r,s} \left(\sum_{i \neq j} \tau_{ir}^g \tau_{js}^g + \rho N(N-1) \hat{\gamma}_r^{g-1} \hat{\gamma}_s^{g-1} \right)$$

$$\times D_{KL}(\text{Bern}(1 - \hat{\pi}_{rs}^g), \text{Bern}(1 - \pi_{rs}^{(0)})) \quad (22b)$$

$$Q_g^\theta = 2 \sum_{r,s} \left(\sum_{i \neq j} \tau_{ir}^g \tau_{js}^g I_{ij}^g + \rho N(N-1) \hat{\gamma}_r^{g-1} \hat{\gamma}_s^{g-1} (1 - \hat{\pi}_{rs}^{g-1}) \right)$$

$$\times D_{KL}(\text{Poi}(\hat{\theta}_{rs}^g), \text{Poi}(\theta_{rs}^{(0)})) \quad (22c)$$

$$Q_g^\varphi = 2 \sum_{r,s} \left(\sum_{i \neq j} \tau_{ir}^g \tau_{js}^g I_{ij}^g + \rho N(N-1) \hat{\gamma}_r^{g-1} \hat{\gamma}_s^{g-1} (1 - \hat{\pi}_{rs}^{g-1}) \hat{\theta}_{rs}^{g-1} \right)$$

$$\times D_{KL}(P(l|\hat{\varphi}_{rs}^g), P(l|\varphi_{rs}^{(0)})) \quad (22d)$$

We chart these four statistics separately and the monitoring scheme trigger an Out-Of-Control (OOC) signal once one of the statistic goes out of its Upper Control Limit (UCL). Please note that this monitoring scheme considers historical information. We use simulations to determine the UCL for each individual chart such that a pre-specified overall IC average run length (ARL_0) is attained, and the simulation method is introduced in Appendix C.1. We denote these UCLs as $U^\gamma, U^\pi, U^\theta$ and U^φ . The MAWDN online monitoring algorithm is summarized in Algorithm 2 in Appendix E.2.

Assuming that there is only one root cause, if there is an alarm, we first determine the type of shift according to the individual chart that triggers the alarm. If the shift is the first type, the diagnosis process finishes; If the shift falls into one of the other three types of shift, we need to further diagnose the parameter of which pair of communities have been changed. We note that the latter three charting statistics can be decomposed based on community pairs. Taking Q_g^π as an example, we can decompose it into $Q_g^{\pi_{rs}} = \sum_{r,s} Q_g^{\pi_{rs}}$, where

$$Q_g^{\pi_{rs}} = 2 \left(\sum_{i \neq j} \tau_{ir}^g \tau_{js}^g + \rho N(N-1) \hat{\gamma}_r^{g-1} \hat{\gamma}_s^{g-1} \right)$$

$$\times D_{KL}(\text{Bern}(1 - \hat{\pi}_{rs}^g), \text{Bern}(1 - \pi_{rs}^{(0)})) \quad (23)$$

We believe that the parameter change of the community pair (r, s) corresponding to the most significant increase of ratio $Q_g^{(\cdot)rs} / Q_g^{(\cdot)}$ from IC to OOC is the root cause of the alarm. As we have mentioned before, the four parameters reflect four types of shifts. If there are more than one of $Q_g^\gamma, Q_g^\pi, Q_g^\theta$ and Q_g^φ out of their control limits, we can conclude that all the shifts, related to the monitoring statistics which are OOC, exist in the system.

As a summary of Section 3.1–3.3, we show the overall modeling and monitoring framework proposed by us in Figure 4.

3.4. Discussion on the proposed model and its implementation

In this section, we summarize some important points on the proposed model and its implementation. First, some key parameters should be set properly, including hyperparameter

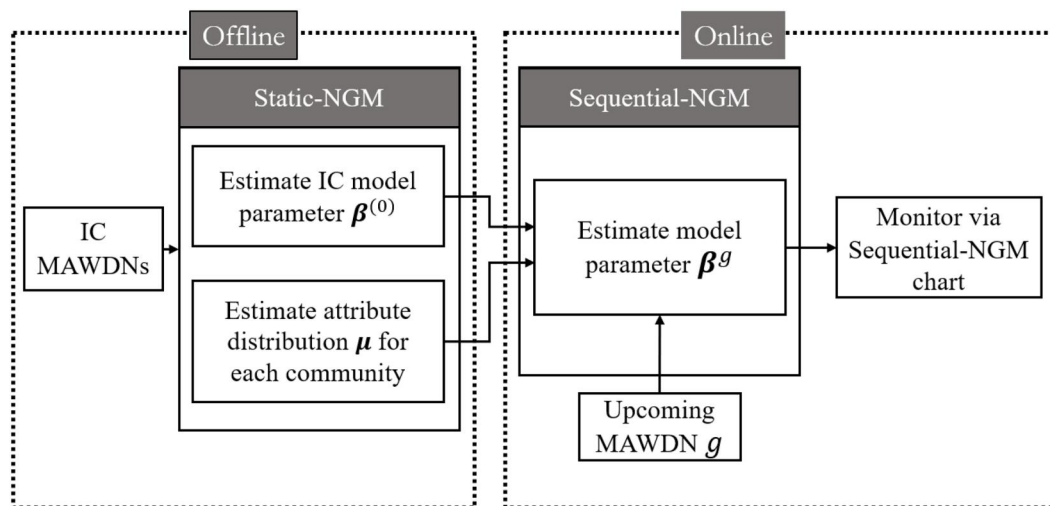


Figure 4. Framework of the proposed method.

β_2 , community number R , weighted coefficient σ , smooth parameter λ and UCL. In this work, β_2 and R are set according to some prior knowledge; σ and λ are set empirically; and UCL is determined by simulation. See Appendix C.1 for details. Second, one assumption about the community structure and three assumptions about the node attributes are made to complete our model. We provide some situations in which these assumptions are valid in Appendix C.2. Thirdly, we discuss the computation efficiency in Appendix C.3 and find that our method scales quadratically with the node number or community number, and scales linearly with the layer number or attribute number. Additionally, our method is tractable. Finally, we provide reasons why the single-layer network modeling methods are not suitable to model multilayer networks (see Appendix C.4) and why we use a multi-chart monitoring scheme in this work (see Appendix C.5).

4. Simulation study

In this section, we first use simulation experiments to validate the parameter estimation accuracy of the VEM algorithm. Then we evaluate the performance of the proposed sequential-NGM chart monitoring scheme and the three competitive schemes in detecting the four types of shifts and root cause diagnosis accuracy.

4.1. Evaluation of parameter estimation

To validate the static-NGM and evaluate the performance of the parameter estimation method, we generate MAWDNs through static-NGM and access the accuracy in estimating model parameters with VEM. It should be noted that the node attribute generation process is similar to He *et al.* (2017) and each attribute is generated by a Bernoulli distribution independently. Given the community ID of a node, the attributes of the node that are strongly related to the community are independent and identically distributed (i.i.d.) generated by a Bernoulli distribution with a large mean p_1 , while the remaining attributes are i.i.d. generated

by a Bernoulli distribution with a small mean p_2 . We set $p_1 = 0.9, p_2 = 0.1$. One discussion about selection of p_1 and p_2 is provided in Appendix D.1. We also estimate the model parameters only through network topology and compare the estimation results with those considering node attributes.

In the simulation, we set node number $N = 45$, community number $R = 3$, layer number $L = 3$, total attribute number is $T = 24$, the number of attributes strongly related to each community number is eight (attributes 1–8 are strongly related to community 1; attributes 9–16 are strongly related to community 2; attributes 17–24 are strongly related to community 3), attribute distribution $\gamma = (1/3, 1/3, 1/3)$. Other model parameters are set as below:

$$\pi = \begin{pmatrix} 0.20 & 0.50 & 0.50 \\ 0.50 & 0.23 & 0.50 \\ 0.50 & 0.50 & 0.20 \end{pmatrix}, \theta = \begin{pmatrix} 20 & 10 & 10 \\ 10 & 25 & 10 \\ 10 & 10 & 24 \end{pmatrix}$$

$$\varphi = \left\{ \begin{pmatrix} 1/7 & 1/3 & 1/3 \\ 1/3 & 3/5 & 1/3 \\ 1/3 & 1/3 & 3/5 \end{pmatrix}; \begin{pmatrix} 4/7 & 1/3 & 1/3 \\ 1/3 & 1/10 & 1/3 \\ 1/3 & 1/3 & 1/10 \end{pmatrix}; \begin{pmatrix} 2/7 & 1/3 & 1/3 \\ 1/3 & 1/10 & 1/3 \\ 1/3 & 1/3 & 1/10 \end{pmatrix} \right\}$$

We estimate the model parameters of 100 networks and report the mean and standard deviation of KL divergence between the distributions defined by the true model parameters and distributions defined by the estimated model parameters. Table 1 displays the mean and standard deviation (values in brackets) of KL divergence, the columns with “w.” represents the results considering node attributes, whereas the columns with “w.o.” represents the results without considering node attributes. Another simulation case with 80 nodes is provided in Appendix D.1.

The results in Table 1 and D1 in Appendix D.1 show that the VEM can provide a satisfactory estimation of parameters of static-NGM. In addition, the estimation results are more accurate when considering node attributes than those not considering node attributes. Also, we find that the parameter estimation results for $N = 45$ are generally more accurate than those for $N = 80$, no matter with or without node attribute. Such results are intuitive. On the one hand, the introduction of node attributes enables us to have more information, so that we can carry out more accurate community identification, thus improving the

Table 1. KL-divergence between the true distributions and the estimated distributions ($N = 45, R = 3$).

(r, s)	$\pi(\times 10^{-2})$		$\theta(\times 10^{-2})$		$\varphi(\times 10^{-2})$		γ	
	w.	w.o.	w.	w.o.	w.	w.o.	w.	w.o.
(1)	0.214(0.273)	1.772(2.694)	0.356(0.564)	16.028(24.530)	0.032(0.043)	32.542(31.877)	0.017(0.020)	0.150(0.161)
(1, 2)	0.218(0.299)	0.219(0.455)	0.536(0.651)	4.102(2.375)	0.081(0.091)	0.317(1.224)		
(1, 3)	0.162(0.216)	0.112(0.178)	0.479(0.608)	0.379(0.500)	0.104(0.127)	0.061(0.063)		
(1, 2)	0.167(0.262)	0.216(0.425)	0.547(0.712)	2.506(13.109)	0.083(0.092)	0.236(0.810)		
(2)	0.250(0.394)	0.201(0.339)	0.421(0.598)	24.992(39.425)	0.026(0.030)	9.344(11.642)		
(2, 3)	0.223(0.343)	0.276(0.800)	0.534(0.849)	13.156(55.573)	0.107(0.113)	0.667(2.437)		
(1, 3)	0.187(0.241)	0.130(0.214)	0.437(0.518)	0.384(0.453)	0.098(0.0103)	0.064(0.086)		
(2, 3)	0.178(0.243)	0.251(0.703)	0.653(0.997)	12.824(51.589)	0.102(0.112)	0.651(2.336)		
(3)	0.219(0.302)	1.092(2.298)	0.343(0.540)	27.017(20.488)	0.026(0.035)	40.619(38.830)		

Table 2. ARL and diagnosis accuracy for different shift scale of γ with different schemes

No.	KL divergence	γ	Sequential-NGM		Static-NGM		ST		BF	
			ARL	Acc.	ARL	Acc.	ARL	Acc.	ARL	Acc.
0	0	(1/3,1/3,1/3)	203.4(2.020)	–	200.4(1.990)	–	201.9(2.038)	–	198.5(1.984)	–
1	0.0003	(0.323,0.333,0.343)	180.6(1.741)	0.327	196.8(1.944)	0.261	197.0(1.980)	0.055	197.0(1.967)	0.029
2	0.0012	(0.313,0.333,0.353)	127.9(1.264)	0.523	181.8(1.837)	0.317	189.9(1.895)	0.065	187.2(1.904)	0.052
3	0.0027	(0.303,0.333,0.363)	80.60(0.771)	0.716	163.2(1.665)	0.394	176.3(1.794)	0.081	170.9(1.721)	0.085
4	0.0048	(0.293,0.333,0.373)	46.19(0.430)	0.848	138.7(1.402)	0.478	166.7(1.669)	0.099	153.1(1.552)	0.150
5	0.0075	(0.283,0.333,0.383)	27.02(0.234)	0.909	114.1(1.165)	0.570	152.4(1.535)	0.134	134.7(1.378)	0.255
6	0.0109	(0.273,0.333,0.393)	16.89(0.138)	0.952	89.85(0.916)	0.677	132.8(1.323)	0.168	115.8(1.153)	0.356
7	0.0148	(0.263,0.333,0.403)	11.56(0.088)	0.972	67.12(0.685)	0.760	116.3(1.163)	0.209	93.36(0.922)	0.487
8	0.0194	(0.253,0.333,0.413)	8.272(0.058)	0.984	49.22(0.500)	0.820	100.1(1.006)	0.256	76.60(0.764)	0.605
9	0.0246	(0.243,0.333,0.423)	6.304(0.041)	0.991	36.49(0.371)	0.875	84.96(0.849)	0.311	62.03(0.605)	0.705
10	0.0305	(0.233,0.333,0.433)	4.940(0.030)	0.993	26.75(0.273)	0.910	70.60(0.703)	0.368	48.41(0.472)	0.793

accuracy of parameter estimation. On the other hand, there are more samples for $N=80$ to perform parameter estimation compared with the cases for $N=40$, thus we can obtain more accurate parameter estimation. In addition, the introduction of node attributes can accelerate the convergence of the algorithm to some extent.

4.2. Performance of monitoring and diagnosis scheme

To evaluate the performance of the proposed sequential-NGM monitoring scheme, we compare it with other three competitive schemes: static-NGM chart, score test chart (ST chart), and Bayes factor chart (BF chart). The detailed description for these charts are given in Appendix B. We consider different scale of different types of shifts in Section 3.3.3. Two metrics, ARL and diagnosis accuracy, are taken into account. The basic settings of each simulated networks are the same as those in the case in Section 4.1, also the IC parameters. The simulation data generating process is described in Appendix D.2. The overall ARL_0 for each monitoring scheme is set as 200. We assume that only one type of shift occurs at a time, and the degree of shift is measured by the KL-divergence between the distributions defined by OOC and IC parameters.

Table 2 displays the comparison results of four monitoring schemes under various degree of community distribution (γ) shift. Acc. represents the accuracy of root cause diagnosis, KL-divergence in the table is the degree of shift, and γ is the value of γ after a shift. Each ARL and diagnosis accuracy rate in the table is approximated based on 10,000 simulations, and the values in brackets are standard errors (SE). We observe that the four schemes can be ranked in the order of sequential-NGM > static-NGM > BF > ST according

to ARLs, and the order is the same according to diagnosis accuracy rate when the degree of shift of γ is not so small (Observations 3 to 10 in Table 2). Observations 1 and 2 in Table 2 show that diagnosis accuracy of ST is relatively larger than BF, but far less than the two four-chart schemes when the degree of shift of γ is very small, however, the detection and diagnosis capability of ST and BF charts is very weak in this situation. Therefore, when the shift of γ is very small, the comparison between ST chart and BF chart is not meaningful. The comparison results of the four monitoring schemes under various degree of π_{11} , θ_{11} or φ_{11} shifts are provided in Appendix D.2.

Generally, the proposed sequential-NGM chart outperforms the other three benchmarks monitoring schemes in both of the metrics. Further, the above results show that the sequential-NGM chart outperforms the static-NGM chart and the BF chart outperforms the ST chart, both in terms of ARLs and diagnosis accuracy rate. This is because the sequential-NGM chart and BF chart take into account information from the historical network and can accumulate small shifts, thus improving the efficiency of monitoring and diagnosis. Also, in most cases, comparing two monitoring schemes that take historical information into account (sequential-NGM chart and BF chart) or two monitoring schemes that do not take historical information into account (static-NGM chart and ST chart), the performance of the multi-chart monitoring scheme in root cause diagnosis accuracy is often better than that of the single-chart monitoring scheme, and in the former three types of anomalies, the performance of the multi-chart monitoring scheme is also better than that of the single-chart monitoring mode in ARLs. In addition, considering the historical information, the performance of the multi-chart scheme (sequential-

NGM chart) in ARLs for monitoring the fourth type of anomaly is also superior to that of single-chart scheme (BF chart). These findings show the advantages of multi-chart monitoring. The main reason is that a single-chart monitoring scheme integrates all randomness based on likelihood function values, reducing the sensitivity of monitoring and diagnosis, whereas the multi-chart monitoring scheme is based on the values of different types of model parameters, separating randomness from different sources, alleviating this problem to some extent. Another advantage of multi-graph monitoring is that different weights can be set for each monitoring statistic based on actual requirements. In this article, we consider each statistic to have the same weight.

Additionally, although the performance of the ST chart on ARLs is better than that of the static-NGM chart when the historical information is not taken into account for the fourth type of anomaly monitoring, ST chart requires the information of the node community label. If this information is unknown, it needs to be inferred, but without considering node attributes, the speed and accuracy of the inference about node community label are low, so the monitoring and diagnosis efficiency of ST chart will be reduced. We also perform a comparison study between our method and the method of Dong *et al.* (2020). The detailed results are shown in Appendix D.2 and show that our model has a better performance.

5. Case study

In this section, we apply our proposed method to the Enron E-mail interaction dataset (Priebe *et al.*, 2005). The October 2001 Enron scandal that led to the bankruptcy of Enron Corporation makes this dataset a typical case in the area of network modeling and monitoring (Zou and Li, 2017; Gahrooei and Pynabar, 2018; Motalebi, Owlia, Amiri, and

Fallhnezhad, 2023; Motalebi, Stevens and Steiner, 2021; Stevens and Steiner, 2021).

Following the design in Dong *et al.* (2020), we also construct E-mails among employees per month as a two-layer weighted directed network, with one layer representing the carbon copy/blind carbon copy (cc/bcc) E-mails and the other layer representing the directly sent E-mails. The direction of an edge is from sender to receiver, and its weight is the number of E-mails sent. Further treating the occupation of employees as node attributes, we then obtain the attributed networks. For simplicity, we only consider the interactions among CEOs, presidents, managers and directors; then, for each node, we use a four-dimensional binary vector with only one element being 1 to represent its attribute. We assume that there are three communities; additionally, we have prior knowledge about the communities that CEOs, presidents, managers and directors form (Gahrooei and Paynabar, 2018). With this knowledge and setting the hyper-parameters as we illustrate in Appendix C.1, we use the networks for June 2000 and July 2000 as IC samples to estimate the IC parameters and use the networks from August 2000 to June 2002 to perform online monitoring.

We calculate the four charting statistics from June 2000 to June 2002, and Figure 5 shows the results. Due to the unchanged community structure of the subnetwork in this case, we notice that Q^{γ} is approximately equal to zero; thus, the chart for it is uninformative. We also observe that there are several peaks in the charts for the remaining three statistics and that there are several real events during the periods corresponding to these peaks. In October 2000 (red line in the figure), Q^{θ} reaches the first peak, and Q^{ϕ} is in the stage of rapid rise. In this month, the strategies of Tim Belden, who was the first Enron executive to be indicted and admit wrongdoing, were discussed. The blue line marks May 2001, when Q^{π} , Q^{θ} and Q^{ϕ} all reach peaks. The former Enron Chairman Kenneth Lay secretly met with Arnold

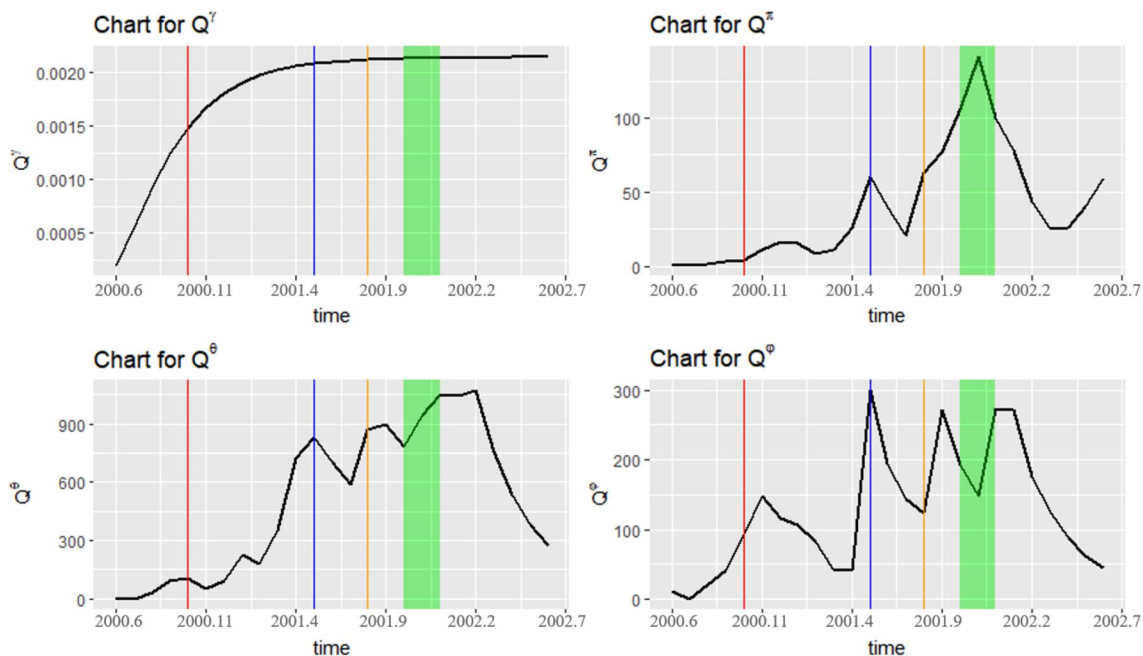


Figure 5. Sequential-NGM charting statistics of Enron E-mail data.

Schwarzenegger, former Los Angeles Mayor Richard Riordan and junk bond King Michael Milken in this month. Lay was touting a plan for solving Los Angeles's energy crisis at that time. In August 2001 (orange line in the figure), Vice President of Enron Corporation Sherron Watkins reported accounting irregularities within the company to then-CEO Ken Lay and warned him that Enron could collapse amid a wave of accounting scandals. We observe this anomaly in the charts of Q^π and Q^θ , i.e., a shift of both overall interaction probability and intensity among individuals exists, while the method of Dong *et al.* (2020) failed to detect this anomaly. Then, from October 2001 to December 2001, the Enron scandal was revealed, and bankruptcy occurred. This abnormal period is evident and marked as a green band in Figure 5. These events are consistent with the observation in Peel and Clauset (2015).

6. Discussion and conclusion

In this article, we focus on modeling and monitoring multi-layer weighted directed networks with node attributes. To incorporate both the node attributes and network topology, first, this article proposes a static NGM model. This model well characterizes the community structure and sparse interaction properties of an MAWDN. Second, we extend it to a sequential version, i.e., sequential-NGM model, to characterize the time-dependent property of a network sequence. A VEM algorithm is developed to estimate the model parameters and use these parameters to describe the features of networks. A sequential-NGM chart monitoring scheme based on the sequential-NGM model is constructed to monitor four types of community-level shifts in a network sequence by the four sets of model parameters. A diagnosis process is introduced to identify the root cause of an OOC alarm.

We validate the proposed method in simulation experiments and observe that integrating the node attributes improves the accuracy of the parameter estimation, as introducing node attributes can improve the accuracy of community detection. Additionally, the proposed monitoring scheme sequential-NGM outperforms the three benchmark schemes (static-NGM, BF, ST) in both the efficiency of the four types of shift detection and the accuracy of root cause diagnosis. Specifically, under the condition of multi-chart (single-chart), the chart considering historical information generally outperforms that not considering historical information. Also, given the sequential-NGM (static-NGM) used, the multi-chart generally has a better performance than the single chart, especially for the root cause diagnosis accuracy. The reason may be that the multi-chart can separate randomness from different sources and increase the sensitivity of monitoring and diagnosis. We also apply the proposed method to Enron E-mail data; the results further validate the efficiency of the method.

There are still some limitations in this work to be addressed in future research:

1. The computation efficiency is limited as the node number, community number and layer number increase,

which is also a common problem for statistical models for networks (Dong *et al.*, 2020); thus, more advanced optimization techniques are needed.

2. The community number is treated as prior knowledge in this work, which may not be realistic in some cases. Selecting the proper community number for a multi-layer network deserves more attention.
3. The node attributes in this work are assumed to be highly related to communities, but how to select them from all available attributes remains an open topic.
4. Generally, a real network has specific features, for example, the power-law feature, which the model in this work disregards.

How to address these features should be further considered, and more realistic models are needed.

Acknowledgments

The authors greatly thank the Department Editor, the Associate Editor and anonymous referees for their helpful comments and suggestions, which have helped us improve this work greatly.

Funding

Dr. Wang's work was supported by the Key Program of the National Natural Science Foundation of China under Grant No. 71932006. Dr. Liang's work was supported by the National Natural Science Foundation of China under Grant No. 72201212.

Notes on contributors

Hao Wu is currently a PhD student at Department of Industrial Engineering, Tsinghua University. He received his BS degree in industrial engineering from Tsinghua University in 2021. His research focuses on network system modeling and monitoring.

Qiao Liang is currently an assistant professor in the School of Statistics, Southwestern University of Finance and Economics, Chengdu, China. She received her PhD and BS degrees in industrial engineering from Tsinghua University, Beijing, China. Her research interests are in the areas of statistical modeling and data analytics for manufacturing and service processes, with a focus on statistical process control based on text analytics.

Kaibo Wang is a professor in the Department of Industrial Engineering, jointly appointed by the Vanke School of Public Health, Tsinghua University, Beijing, China. He received his BS and MS degrees in mechatronics from Xi'an Jiaotong University, Xi'an, China, and his PhD in industrial engineering and engineering management from the Hong Kong University of Science and Technology, Hong Kong. His research focuses on statistical quality control and data-driven system modelling, monitoring, diagnosis, and control.

ORCID

Qiao Liang  <http://orcid.org/0000-0001-8758-291X>
Kaibo Wang  <http://orcid.org/0000-0001-9888-4323>

Data availability statement

The data that support the findings of this study are openly available at <http://www.cs.cmu.edu/~enron/>

References

- Azarnoush, B., Paynabar, K., Bekki, J. and Runger, G. (2016) Monitoring temporal homogeneity in attributed network streams. *Journal of Quality Technology*, **48**(1), 28–43.
- Chang, Z., Jia, C., Yin, X. and Zheng, Y. (2019) A generative model for exploring structure regularities in attributed networks. *Information Sciences*, **505**, 252–264.
- Chunaev, P. (2020) Community detection in node-attributed social networks: A survey. *Computer Science Review*, **37**, 100286.
- Contisciani, M., Power, E.A. and De Bacco, C. (2020) Community detection with node attributes in multilayer networks. *Scientific Reports*, **10**(1), 1–16.
- 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**(4), 042317.
- De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivela, M., Moreno, Y., Porter, M.A., Gómez, S. and Arenas, A. (2013) Mathematical formulation of multilayer networks. *Physical Review X*, **3**(4), 041022.
- Dong, H., Chen, N. and Wang, K. (2020) Modeling and change detection for count-weighted multilayer networks. *Technometrics*, **62**(2), 184–195.
- Ebrahimi, S., Reisi-Gahrooei, M., Paynabar, K. and Mankad, S. (2021) Monitoring sparse and attributed networks with online hurdle models. *IISE Transactions*, **54**(1), 91–104.
- Farahani, E.M., Baradaran Kazemzadeh, R., Noorossana, R. and Rahimian, G. (2017) A statistical approach to social network monitoring. *Communications in Statistics-Theory and Methods*, **46**(22), 11272–11288.
- Gahrooei, M.R. and Paynabar, K. (2018) Change detection in a dynamic stream of attributed networks. *Journal of Quality Technology*, **50**(4), 418–430.
- Han, Q., Xu, K. and Airolidi, E. (2015) Consistent estimation of dynamic and multi-layer block models, *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France, pp. 1511–1520.
- Hazrati-Marangaloo, H. and Noorossana, R. (2021) A nonparametric change detection approach in social networks. *Quality and Reliability Engineering International*, **37**(6), 2916–2935.
- He, D., Feng, Z., Jin, D., Wang, X. and Zhang, W. (2017) Joint identification of network communities and semantics via integrative modeling of network topologies and node contents, in *Proceedings of the AAAI Conference on Artificial Intelligence*, San Francisco, California, USA, pp. 116–124.
- Hewapathirana, I.U. (2019) Change detection in dynamic attributed networks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **9**(3), e1286.
- Kivela, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y. and Porter, M.A. (2014) Multilayer networks. *Journal of Complex Networks*, **2**(3), 203–271.
- Kullback, S. (1997) *Information Theory and Statistics*. Courier Corporation, Massachusetts, USA.
- Lee, W., McCormick, T.H., Neil, J., Sodja, C. and Cui, Y. (2022) Anomaly detection in large-scale networks with latent space models. *Technometrics*, **64**(2), 241–252.
- Liang, Q. and Wang, K. (2022) Ratings meet reviews in the monitoring of online products and services. *Journal of Quality Technology*, **54**(2), 197–214.
- Liu, W., Chang, Z., Jia, C. and Zheng, Y. (2022) A generative node-attribute network model for detecting generalized structure and semantics. *Physica A: Statistical Mechanics and its Applications*, **588**, 126557.
- McCulloh, I. and Carley, K.M. (2011) Detecting change in longitudinal social networks. Technical report, Military Academy West Point NY Network Science Center (NSC).
- Motalebi, N., Owlia, M.S., Amiri, A. and Fallahnezhad, M.S. (2023) Monitoring social networks based on zero-inflated Poisson regression model. *Communications in Statistics-Theory and Methods*, **52**(7), 2099–2115.
- Motalebi, N., Stevens, N.T. and Steiner, S.H. (2021) Hurdle blockmodels for sparse network modeling. *American Statistician*, **75**(4), 383–393.
- Newman, M.E. and Park, J. (2003) Why social networks are different from other types of networks. *Physical Review E*, **68**(3), 036122.
- Nicosia, V. and Latora, V. (2015) Measuring and modeling correlations in multiplex networks. *Physical Review E*, **92**(3), 032805.
- Noorossana, R., Hosseini, S.S. and Heydarzade, A. (2018) An overview of dynamic anomaly detection in social networks via control charts. *Quality and Reliability Engineering International*, **34**(4), 641–648.
- Pandhre, S., Gupta, M. and Balasubramanian, V.N. (2016) Community-based outlier detection for edge-attributed graphs. *arXiv preprint arXiv:1612.09435*.
- Paul, S. and Chen, Y. (2016) Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. *Electronic Journal of Statistics*, **10**(2), 3807–3870.
- Peel, L. and Clauset, A. (2015) Detecting change points in the large-scale structure of evolving networks, *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, Austin, Texas, USA, pp. 2914–2920.
- 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.
- Savage, D., Zhang, X., Yu, X., Chou, P. and Wang, Q. (2014) Anomaly detection in online social networks. *Social Networks*, **39**, 62–70.
- Stevens, N.T., Wilson, J.D., Driscoll, A.R., McCulloh, I., Michailidis, G., Paris, C., Parker, P., Paynabar, K., Perry, M.B., Reisi-Gahrooei, M., et al. (2021) Research in network monitoring: Connections with SPM and new directions. *Quality Engineering*, **33**(4), 736–748.
- Sun, H., He, F., Huang, J., Sun, Y., Li, Y., Wang, C., He, L., Sun, Z. and Jia, X. (2020) Network embedding for community detection in attributed networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, **14**(3), 1–25.
- Wang, J. and Xie, M. (2021) Modeling and monitoring unweighted networks with directed interactions. *IISE Transactions*, **53**(1), 116–130.
- Woodall, W.H., Zhao, M.J., Paynabar, K., Sparks, R. and Wilson, J.D. (2017) An overview and perspective on social network monitoring. *IISE Transactions*, **49**(3), 354–365.
- Zhang, S., Wang, Y., Wang, W., Jiao, P. and Pan, L. (2021) A unified Bayesian model of community detection in attribute networks with power-law degree distribution, *Collaborative Computing: Networking, Applications and Worksharing: 16th EAI International Conference, Proceedings, Part II 16*, Springer, Cham, Switzerland, pp. 518–529.
- Zou, N. and Li, J. (2017) Modeling and change detection of dynamic network data by a network state space model. *IISE Transactions*, **49**(1), 45–57.