HB-DSBM: Modeling the Dynamic Complex Networks from Community Level to Node Level

Peng-Fei Jiao, Tian-Peng Li, Hua-Ming Wu, Chang-Dong Wang, Dong-Xiao He, and Wen-Jun Wang*

***Abstract*—A variety of methods have been proposed for modeling and mining dynamic complex networks, in which the topological structure varies with time. As the most popular and successful network model, Stochastic Block Model (SBM) has been extended and applied to community detection, link prediction, anomaly detection and evolution analysis of dynamic networks. However, all current models based on the SBM for modeling dynamic networks are designed at the community level, assuming that nodes in each community have the same dynamic behavior, which usually results in poor performance on temporal community detection and loses the modeling of node abnormal behavior. To solve the above problem, this paper proposes a Hierarchical Bayesian Dynamic Stochastic Block Model (HB-DSBM) for modeling the node-level and community-level dynamic behavior in a dynamic network synchronously. Based on the SBM, we introduce a hierarchical Dirichlet generative mechanism to associate the global community evolution with the microscopic transition behavior of nodes near-perfectly and generate the observed links across the dynamic networks. Meanwhile, an effective variational inference algorithm is developed and we can easy to infer the communities and dynamic behaviors of the nodes. Furthermore, with the two-level evolution behaviors, it can identify nodes or communities with abnormal behavior. Experiments on simulated and real-world networks demonstrate that HB-DSBM has achieved state-of-the-art performance on community detection and evolution. In addition, abnormal evolutionary behavior and events on dynamic networks can be effectively identified by our model.**

***Index Terms*—Temporal community detection, Node-level behavior, Community-level behavior, Dynamic SBM, Variational inference**

I. INTRODUCTION

COMPLEX networks have been widely applied to model a variety of real-world phenomena, e.g., social relations [1] and biological systems [2]. As an emerging interdisciplinary subject, network science has received increasing attention from different fields. Community structure, as one of the important statistical characteristics, plays an important role in understanding the formation and function, link prediction and spreading dynamics of complex networks. The community is usually denoted as a sub-network with a higher inner density in a given network. Community detection drives the development of link prediction, information spreading and anomaly detection of complex networks, it also can be used for event detection, criminal organization identification and terrorist group mining. Therefore, a large number of methods and models for community detection [3]–[6] have been proposed, e.g., modularity optimization-based, spectral clustering and statistical inference [7].

All of the methods described above are only designed for static networks with constant structures. However, complex networks are usually time-varying, i.e., the structure of the networks changes over time, which is referred to as dynamic or temporal complex networks. For instance, in a collaboration network based on DBLP data, nodes and edges denote the authors and their cooperative relationships, as well as changes in research fields, and the number of nodes and links changes over time. For this network, the communities usually correspond to different areas of research, such as the area of *Data Mining*. With the varying structures of dynamic networks, the communities present different evolution forms [8], including *Growth*, *Contraction*, *Merge*, *Split*, etc., and the evolution forms are early denoted by [9] and further expanded by [3].

With the various forms of community evolution in dynamic networks, it poses new challenges for detecting temporal communities. It is widely acknowledged that a dynamic network is usually represented as a series of snapshots, each of which can be regarded as a static network. As a result, community detection in dynamic networks has three subproblems, namely, community detection in each snapshot, evolution across snapshots and abnormal behavior or change point identification. These complex and interwoven challenges make it full of vitality. In essence, these subproblems can mutually promote and are related to each other. At the same time, the temporal community detection can also be used for exploring functional brain networks [10], change point detection [11], identification of influential individuals [12], [13], fraud detection [14], etc. It can also help understand the evolution mechanism of dynamic networks and make better link predictions.

A variety of methods [3], [15] for community detection in dynamic networks have been developed, including methods based on modularity, spectrum, multi-objective optimization, dynamics, dynamic stochastic block models and matrix decomposition. Each method focuses either on detection, evolution or abnormal behavior. For the detection task, they improve the community quality by considering the topological structure of some successive snapshots, which could increase the robust-

---

P. Jiao is with the School of Cyberspace, Hangzhou Dianzi University, Hangzhou, 310018, China. Email: cspjiao@gmail.com.

T. Li, D. He and W. Wang are with the College of Intelligence and Computing, Tianjin University, Tianjin, 300350, China. W. Wang is also with the College of Information Science and Technology, Shihezi University, Shihezi, 832003, China and the Shenzhen Research Institute of Tianjin University, Shenzhen, 518063, China. Email: {ltpnimeia, hedongxiao, wjwang}@tju.edu.cn.

H. Wu is with the Center for Applied Mathematics, Tianjin University, Tianjin, 300072, China. Email: whming@tju.edu.cn.

C. Wang is with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China. Email: changdongwang@hotmail.com.

(Corresponding author: Wen-Jun Wang)

ness of the community structure. The most representative one is the evolutionary clustering framework, which assumes that there should be no obvious mutation in community structure abrupt change and leads to a number of temporal communities detection methods like FaceNet [16]. For the community evolution task, the methods mainly focus on the evolution of communities across the snapshots with given community structures. For instance, event matching-based methods establish the corresponding relationship between communities by defining some similarity indexes across the snapshots. However, all these methods and models are focusing on community detection or community evolution, independently. The abnormal behavior is more related to the community evolution, if we have learned the dynamic behaviors on nodes and communities, it is easy to identify the abnormal nodes of communities or some events on networks [17]. Similarly, some critical events and behaviors are conducive to community detection and evolution.

In recent years, some models have tried to solve community detection and evolution simultaneously by modeling the dynamic network. DSBM [18] is the first generative model for the tasks by defining a probability transition matrix between two snapshots based on a stochastic block model. Then, some studies [19], [20] further extended DSBM to generate the dynamic network with an evolutionary community structure, which relaxed the constraint of fixed connectivity probabilities. Liu *et al.* [21] proposed DECS based on the evolutionary clustering and multi-objective optimization algorithm. At the same time, some methods of integrating community structure and abnormal behavior have been proposed gradually. Cheung *et al.* [22] proposed to detect both change points and community structures simultaneously based on each individual sub-network following a Stochastic Block Model (SBM), and utilized Minimum Description Length (MDL) Principle for minimizing objective criterion.

However, all these methods for modeling dynamic networks, whether based on DSBM or heuristics, are designed at the community level, i.e., they assume that nodes in each community have the same dynamic behavior. For example, the nodes in one community have the same transition probability from the current snapshot to the next, which has been proved by some empirical studies [23]. Thus, the lack of some important dynamic patterns of different nodes and communities usually leads to poor performance on temporal community detection and ignores certain node-level dynamic behaviors. It is impossible to recognize abnormal node behaviors in the community network.

In this paper, we focus on modeling the evolution of node-level and community-level of the dynamic network synchronously. There exist two major challenges: 1) how to accurately model both node-level and community-level dynamic behavior? 2) how to quantify the relationship between node-level and community-level transitions? With this information, we can infer the community structure, dynamic evolution and nodes with abnormal behaviors. In order to solve these problems in principle, we propose a Hierarchical Bayesian Dynamic Stochastic Block Model (HB-DSBM) to model both community structure and evolution from the perspective of statistical models. We characterize the changes of the generated network by defining several latent variables, such as community-level transition, node-level transition, and community membership of nodes. At the same time, we introduce a hierarchical Dirichlet generation mechanism to associate the mesoscopic community evolution with the micro-transition behavior of nodes in dynamic networks. Then, we show a detailed generated process of the temporal network and drive an effective optimization algorithm based on the variational inference. Experiments on temporal community detection, evolution, abnormal behavior and some case studies show the superiority of HB-DSBM. The main contributions of this paper can be summarized as follows:

- We propose a full Bayesian generative model called HB-DSBM, which is a well-designed generation mechanism based on DSBM and can generate and model the evolution of nodes and communities, and the changes of dynamic networks.
- The focus of HB-DSBM is to model dynamic communities at the node and community levels synchronously from a hierarchical Bayesian perspective. Moreover, HB-DSBM can help improve community detection and evolution, and further identify abnormal behaviors.
- We also propose an effective variational inference algorithm for HB-DSBM. Extensive experiments have been conducted to show its superior performance and effectiveness on simulated and real-world dynamic networks.

II. RELATED WORK

Recently, several innovative methods have been developed for community mining in dynamic networks. Here, we divide these methods into three categories, namely, heuristic optimization-based for temporal community detection, generative models for modeling community evolution, and abnormal behavior of nodes and communities identification on dynamic networks.

A. Heuristic Optimization-based Methods

In general, heuristic optimization-based methods can be divided into three categories, namely, two-step methods, incremental clustering-based methods and evolutionary clustering.

Two-Step Methods: These methods treat community detection and evolution as two independent problems. The key idea is to detect the communities, and then use similarity measures to match communities across different snapshots. GraphScope [24] first used the Minimum Description Length (MDL) principle to extract communities and detect community evolutions. Besides, there are also many methods based on the embedding technology with side information like node feature, e.g., the author information in a co-author network. For instance, TRNN [25] utilizes the multi-head self-attention into a transformer-style neural network to capture the dynamic information in the dynamic network, then observe the embedding vector of every node to support downstream tasks like link prediction, node classification and so on. But these methods take side information as input in the models. If there is no side information in the network, it is ubiquitous in the real world due to the privacy policy, the accuracy of these methods will drop drastically.

Incremental Clustering-based Methods: These methods typically renew the communities according to the varying of the dynamic network based on the community structure at the first snapshot, which is obtained via one static network. DynaMo [26] is an incremental modularity-based clustering method, which is faster than Louvain algorithm. Tajeuna *et al.* [11] proposed an approach for automatically detecting the size of the snapshot to adopt when identifying and tracking communities over time. TILES [27] used label propagation for network changes, which is very efficient for large-scale networks.

Evolutionary Clustering-based Methods: Evolutionary clustering is the most popular framework for detecting the communities of the snapshot by adding regularization. FacetNet [16] was a typical model based on a unified framework for community detection and evolution with evolutionary clustering. DYNMOGA [28] was proposed by turning the community detection into a multi-objective optimization problem. GenLouvain [29] exploited a novel measure of dynamic networks based on the modularity, and then used the well-known Louvain method [30] to calculate the community structure. Seifikar *et al.* [31] further introduced a new Louvain-based dynamic community detection algorithm which is relied on the previous snapshot of the network evolution. PisCES [32] used historical observations to predict future events of a network, which put the two tasks into a unified framework and made them mutually constrained.

In summary, some of these methods ignore the evolution for community mining in dynamic networks, some treat community detection and evolution as independent problems, and some lack theoretical explanations that simultaneously solve community detection and evolution.

B. Generative Model

It can be divided into Dynamic Latent Space Model (DLSM), Dynamic Stochastic Block Model (DSBM) and other generative models.

DLSM embeds the dynamic network into a latent Euclidean space and assigns each node a temporal trajectory with time [33] and uses it to conduct community detection [34]. Considering a large number of parameters in such models, they are usually optimized by the Markov Chain Monte–Carlo (MCMC) sampling and are only suitable for small networks [35]. The classic DSBM [18] was first proposed based on Stochastic Block Model (SBM) to analyze dynamic networks by providing a unified framework to capture both communities and their evolution simultaneously. Xu *et al.* [19] proposed a state-space model for dynamic networks based on the SBM and used an Extended Kalman Filter (EKF) augmented with a local search to optimize the model. Furthermore, Xu [36] constructed a stochastic block transition model (SBTM) to model the direct influence of connections between snapshots. Becker and Holzmann [37] analyzed the nonparametric identification in the dynamic stochastic block model. DBTDP [38] was developed for community detection and evolution tracking, and the number of communities per snapshot was automatically determined by a Dirichlet process. Similar works are proposed in [39], [40]. In addition, other generative models of dynamic networks have also been proposed, for example [41] proposed a Poisson gamma probabilistic model based on the Bernoulli Poisson link function. Furthermore, some methods based on the Nonnegative Matrix Factorization (NMF) for community detection in dynamic networks can also be regarded as model-based [42]–[44].

In summary, all these generative models are based on the community-level transition tendency, assuming that the nodes within the same community have identical dynamic behavior. This assumption causes different dynamic behaviors of nodes lost, thus failing to capture of dynamic behavior heterogeneity of nodes.

C. Abnormal Behavior Identification

Abnormal behavior identification in dynamic networks can indicate profound underlying network structure changes. It can be divided into two types:

Node Abnormal Behavior: It refers to abnormalities that occur due to changes in node features. Its key idea is to calculate and compare the node feature of different time slices. DeltaCon [45] handles streaming graphs, gradational similarity updates for time-evolving graphs by using the L2 norm, as well as graphs with node attributes for expressions that utilize fast belief propagation [46] to derive node affinity. CICPD [17] encodes nodes' importance characteristics of each time slice through PageRank defines a new network by using Jensen-Shannon (JS) divergence to compute the distance of snapshots. NetWalk [47] utilizes the popular network embedding method based on deep autoencoder and clique embedding to dynamically capture node abnormal, and by using reservoir sampling, NetWalk can compute the vector representations with constant space requirements.

Community or Motif Abnormal Behaviors [48], [49]: They are obviously caused by the huge change in the group. Peel *et al.* [50] introduced GHRG which laconically models nested community structure at all scales in a network. It considers a fixed-length sliding window and uses the generalized likelihood ratio to evaluate whether and the type of changes. The coding method of GraphScope [17] not only considers the community structure, but also considers their change points in time. Community detection and segment partition are obtained and their resemblance is measured by the Minimum Description Length (MDL) principle.

In summary, these methods focus on identifying the abnormal nodes or communities, or the change points on the temporal networks. As a result, they all failed in community detection and its evolution.

III. THE PROPOSED MODEL

In this section, we first introduce the notations used in this paper, and then give the problem definition and the details of our proposed HB-DSBM, including its generation process and the joint probability distribution.

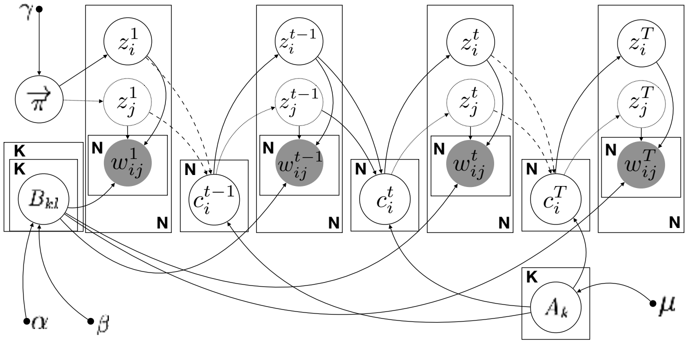


Fig. 1: Graphical model of HB-DSBM. w_{ij}^t is the similarity of nodes i and j at snapshot t and the observation variable. A and C are evolutionary variables at the community and node levels, respectively, representing their dynamic behaviors. B and Z are the latent variables of DSBM, and α, β, μ and γ are hyperparameters.

A. Problem Formulation

Given a dynamic unweighted and undirected network with T snapshots, we use $W = \{W^1, W^2, \cdots, W^T\}$ to represent a series of adjacency matrices of the temporal network. We also denote $Z = \{Z^1, Z^2, \cdots, Z^T\}$ as the community assignment for all nodes in each snapshot, i.e., $z_i^t \in \{1, \cdots, K\}$ is used to denote the community ownership of node i, where K (assumed to be constant) is the number of communities, $t = 1, \cdots, T$ and $i = 1, \cdots, N$. In other words, $z_i^t = k$ means that node i belongs to the k-th community at snapshot t. Important notations used in this paper are listed in Tab. I. Therefore, the problem can be formulated as: given W, how can we get the community structure, its evolutionary trajectory and nodes across the network? Furthermore, which nodes and communities may have abnormal behaviors? The HB-DSBM comes into being.

The graph model of HB-DSBM is illustrated in Fig. 1, which generates the observed links at each snapshot based on the community and node levels. It can be described in the following three parts.

- In this model, π is the prior probability of Z^1 and follows a Dirichlet distribution with parameter γ, B is the probability matrix denoted on the communities, i.e., B_{kl} is the probability that two nodes belonging to communities k and l will establish a link, and it follows a Beta distribution with hyperparameters α_{kl} and β_{kl}. With the community membership and the probability matrix, it can generate the links at each snapshot.
- Let $A \in [0,1]^{K \times K}$ express the global community-level transition matrix, and A_k, each row of A, follows a Dirichlet distribution with μ as its parameter, so $\sum_l A_{kl} = 1$. It denotes the global dynamic behaviors of communities and models their evolution.
- We introduce $C = \{C^1, C^2, \cdots, C^T\}$ to handle the community transition tendency for each node across the snapshots, where each C^t is generated from the community-level transition matrix A and C^1 has no real meaning just for unified formalization. At snapshot $t > 1$, node i follows its unique transition vector $C_i^t \in [0,1]^K$, which is a probability vector following a Dirichlet distribution with parameter $A_{z_i^{t-1}}$, so $\sum_k c_{ik}^t = 1$. It can model the microscopic evolution behavior of nodes and help improve community detection and abnormal behaviors.

TABLE I: Notations and their description

Notation	**Description**
K, N	The number of communities and nodes of the dynamic network, respectively
W^t	The adjacency matrix at snapshot t
π_k	The probability of node i belonging to community k at snapshot 1
z_i^t	The community node i belonging to community at snapshot t
A_k	The community-level transition vector of k
C_i^t	The node-level transition vector of node i at snapshot t
B_{kl}	The probability of connection between community k and l at any snapshot
γ, μ	The parameters of Dirichlet distribution of π and A_k
α, β	The parameters of Beta distribution of B

The first step is a general generation process for each link of all snapshots. The second and third processes are the core components of the model from the community and node levels, which form the evolution dynamics of the temporal network. The third step is very important and exquisite to associate the community evolution with node behavior, which is known as the hierarchical Dirichlet generative mechanism.

Compared to DSBM [18] and SBTM [36], our HB-DSBM inherits all their advantages and has the following innovations. From the microscopic view, we use C_i^t ($t \geq 2$) to describe the transition probability vector of node i at snapshot t, which can represent the temporal trajectory of nodes. From the mesoscopic view, we denote A as the probability matrix of community transition, which is the dynamic behavior patterns of communities across the dynamic networks. Furthermore, we integrate these two pieces of information to construct and model the dynamic network.

Based on the above discussion, it can generate the temporal community structure and its evolution for the observed links of the dynamic network. The complete generation process is as follows:

1) Generate community initialize probability $\pi \sim Dir(\gamma)$.
2) Generate block matrix $B \sim Beta(\alpha, \beta)$.
3) For each node i at snapshot $t = 1$:
    a) Generate every node's community assignment $z_i^1 \sim Mult(\pi)$.
    b) Generate link $\omega_{ij}^1 \sim Bernoulli(\cdot|B_{z_i^1, z_j^1})$.
4) Generate every community-level transition probability vector $A_k \sim Dir(\mu)$.
5) For each node i at snapshot $t > 1$:
    a) Generate every node-level community transition vector $C_i^t \sim Dir(A_{z_i^{t-1}})$.
    b) Generate every node's community assignment $z_i^t \sim Mult(C_i^t)$.
    c) Generate link $\omega_{ij}^t \sim Bernoulli(\cdot|B_{z_i^t, z_j^t})$.

According to the graphical model in Fig. 1 and the above generation process, the joint probability distribution of HB-DSBM can be written as:

$$
\begin{aligned}
&P(W, Z, C, B, A, \pi|\alpha, \beta, \gamma, \mu) \\
=&\prod_{t=1}^{T} Pr(W^t|Z^t, B) Pr(Z^1|\pi) \prod_{t=2}^{T} Pr(Z^t|C^t) \\
&\prod_{t=2}^{T} Pr(C^t|A, Z^{t-1}) Pr(A|\mu) Pr(\pi|\gamma) Pr(B|\alpha, \beta) \\
=&\prod_{t=1}^{T}\Big[\prod_{w_{ij}=1} B_{z_i^{(t)} z_j^{(t)}} \prod_{w_{ij}=0} (1 - B_{z_i^{(t)} z_j^{(t)}})\Big] \prod_{i=1}^{N} \pi^{z_i^1} \\
&\prod_{t=2}^{T}\prod_{i=1}^{N} C_i^{t z_i^t} \prod_{t=2}^{T}\prod_{i=1}^{N} \frac{\Gamma(\sum_l A_{z_i^{t-1} l})}{\prod_l \Gamma(A_{z_i^{t-1} l})} \prod_l C_{il}^{A_{z_i^{t-1} l} - 1} \\
&\prod_k \frac{\Gamma(\sum_l (\mu_{kl}))}{\prod_l \Gamma(\mu_{kl})} \prod_l A_{kl}^{\mu_{kl}-1} \frac{\Gamma(\sum_k \gamma_k)}{\prod_k \Gamma(\gamma_k)} \prod_k \pi_k^{\gamma_k - 1} \\
&\prod_{k, l \geq k} \frac{\Gamma(\alpha_{kl} + \beta_{kl})}{\Gamma(\alpha_{kl})\Gamma(\beta_{kl})} B_{kl}^{\alpha_{kl}-1} (1 - B_{kl})^{\beta_{kl}-1}. \qquad (1)
\end{aligned}
$$

When $t = 1$, we could generate the community z_i^1 of node i from a multinomial distribution with parameter π, and then generate links between every pair of nodes following a Bernoulli distribution, i.e., nodes i and j have a link between them with a probability of $Bernoulli(\cdot|B_{z_i^1, z_j^1})$. When $t > 1$, the nodes' community ownership Z^t follows a multinomial distribution with parameter C^t, i.e., node i follows $Mult(C_i^t)$ to choose its community at t and C_i^t is generated from $Dir(A_{z_i^{t-1}})$. Each A_k is generated from a Dirichlet distribution with parameter μ, it denotes the global transfer tendency of community k.

To optimize this model, our goal is to calculate the posterior distribution $P(W, Z, C, B, A, \pi|\alpha, \beta, \gamma, \mu)$, which can be written as:

$$
P(Z, C, B, A, \pi|W) = \frac{P(W, Z, C, B, A, \pi|\alpha, \beta, \gamma, \mu)}{P(Z, C, B, A, \pi|W, \alpha, \beta, \gamma, \mu)}, \qquad (2)
$$

where the parameters α, β, γ and μ are ignored for convenience. In the next section, we will introduce how to calculate and optimize this posterior.

It is important to emphasize that although we have assumed that the number of communities and nodes in the network is fixed, these communities and nodes are unweighted and undirected. It can be easily extended to complex situations similar to DSBM [18]. On the other hand, we can also extend the link generation process by replacing SBM with a more refined model, such as the degree preserving [51] or scale-free characteristic [52] SBM model. Although they can help improve the generative capability and community detection, our focus is modeling the dynamic network from the community and node levels, and these extensions can be used as follows.

IV. LEARNING THE MODEL

Intuitively, the proposed model is complex and difficult to be optimized. So in this section, we propose an efficient variational expectation-maximization algorithm to infer parameters of the model. Then, on the basis of the learned parameters Z, C, B, A, and π, we can infer community structure, community evolution, block matrix and abnormal behaviors in the dynamic network.

A. Variational Inference

It is usually difficult to directly calculate the posterior distribution $P(Z, C, B, A, \pi|W, \alpha, \beta, \gamma, \mu)$, for it needs to integrate all the hidden variables Z, C, B and A. Although we take some conjugate prior distributions, its calculation is also exponential. For complex probability graph models, variational inference is usually used as an effective learning method. With this framework, we can approximate the posterior with a decomposable distribution $q(Z, C, B, A, \pi)$ based on mean-field theory [41] as follows:

$$
q(\Delta) = \prod_{t=1}^{T}\prod_{i=1}^{N} q(z_i^t) \prod_{t=2}^{T}\prod_{i=1}^{N} q(c_i^t) q(B) q(A) q(\pi), \qquad (3)
$$

where Δ represents parameters $\{Z, C, B, A, \pi\}$ for simplicity, the block matrix variational parameter $q(B|\widetilde{\alpha}, \widetilde{\beta}) = \prod_{k,l \geq k} Beta(\widetilde{\alpha}_{kl}, \widetilde{\beta}_{kl})$, and community-level transition matrix variational parameter $q(A|\widetilde{\mu}) = \prod_{k=1}^{K}\prod_{l=1}^{K} Dir(\widetilde{\mu}_{kl})$. $q(z_i^t|\widetilde{\phi}_i^t)$ follows a multinomial distribution with $\widetilde{\phi}_i^t$ as its parameter. $q(c_i^t|\widetilde{\xi}_i^t)$ and $q(\pi|\widetilde{\gamma})$ both follow the Dirichlet distribution with parameters $\widetilde{\xi}_i^t$ and $\widetilde{\mu}_{kl}$, respectively. We need to note that all of these settings are based on the conjugate distribution and it can help infer the variational parameters based on the coordinate ascent algorithm.

After a series of derivation [53] and variational inference, we have the following identity (APPENDIX A):

$$
\log P(W) = KL(q(\Delta)||P(Z, C, B, A, \pi|W)) + \widetilde{L}(q). \qquad (4)
$$

To learn the model, our goal is to optimize the KL divergence between $P(Z, C, B, A, \pi|W, \alpha, \beta, \gamma, \mu)$ and $q(\Delta) = q(Z, C, B, A, \pi)$. According to Eq. 4, minimizing the divergence is equivalent to maximizing $\widetilde{L}(q)$, referred to as the evidence lower bound (ELBO). So our goal is to optimize $\widetilde{L}(q)$ with respect to the variational parameters.

Under the variational inference framework, the variational evidence lower bound $\widetilde{L}(q)$ of the model can be written as:

$$
\begin{aligned}
\widetilde{L}(q) =& \sum_z \int_{\pi, B, A, C} q(\Delta) \log \frac{p(\Delta, W)}{q(\Delta)} \mathrm{d}\Delta \\
=& \mathbb{E}_{\widetilde{\phi}, \widetilde{\alpha}, \widetilde{\beta}} \sum_{t=1}^{T} [\log P(W^t|Z^t, B)] \\
&+ \mathbb{E}_{\widetilde{\gamma}, \widetilde{\phi}} [\log P(Z^1|\pi)] + \mathbb{E}_{\widetilde{\phi}, \widetilde{\xi}} \sum_{t=2}^{T} [\log P(Z^t|C^t)] \\
&+ \mathbb{E}_{\widetilde{\xi}, \widetilde{\phi}, \widetilde{\mu}} \sum_{t=2}^{T} [\log P(C^t|A, Z^{t-1})] \\
&+ \mathbb{E}_{\widetilde{\mu}} [\log P(A)] + \mathbb{E}_{\widetilde{\gamma}} [\log P(\pi)] + \mathbb{E}_{\widetilde{\alpha}, \widetilde{\beta}} [\log P(B)] \\
&- \mathbb{E}_{\widetilde{\gamma}} [\log q(\pi)] - \mathbb{E}_{\widetilde{\alpha}, \widetilde{\beta}} [\log q(B)] - \mathbb{E}_{\widetilde{\mu}} [\log q(A)] \\
&- \sum_{t=2}^{T}\sum_{i=1}^{N} \mathbb{E}_{\widetilde{\xi}} [\log q(C_i^t)] - \sum_{t=1}^{T}\sum_{i=1}^{N} \mathbb{E}_{\widetilde{\phi}} [\log q(z_i^t)], \qquad (5)
\end{aligned}
$$

where $\widetilde{\phi}$, $\widetilde{\xi}$, $\widetilde{\alpha}$, $\widetilde{\beta}$, $\widetilde{\mu}$ and $\widetilde{\gamma}$ are the variational parameters. For simplicity, we omit the conditional parts of the $q(\cdot)$ distributions. For example, we abbreviate $q(\Delta|\widetilde{\phi}, \widetilde{\xi}, \widetilde{\alpha}, \widetilde{\beta}, \widetilde{\mu}, \widetilde{\gamma})$ and $q(z_i^t|\widetilde{\phi}_i^t)$ to $q(\Delta)$ and $q(z_i^t)$.

We maximize the ELBO to learn the variational parameters of the latent variables Z, π, B, A, C and the model parameters $\gamma, \alpha, \beta, \mu$. We take the derivatives of $\widetilde{L}(q)$ with respect to the variational parameters $\widetilde{\phi}, \widetilde{\gamma}, \widetilde{\alpha}, \widetilde{\beta}, \widetilde{\mu}, \widetilde{\xi}$, and set these derivatives to zeros and get the update rules as:

$$\nabla \widetilde{L}(q) = \left\{ \frac{\partial \widetilde{L}}{\partial \widetilde{\gamma}}, \frac{\partial \widetilde{L}}{\partial \widetilde{\alpha}}, \frac{\partial \widetilde{L}}{\partial \widetilde{\beta}}, \frac{\partial \widetilde{L}}{\partial \widetilde{\mu}}, \frac{\partial \widetilde{L}}{\partial \widetilde{\xi}}, \frac{\partial \widetilde{L}}{\partial \widetilde{\phi}} \right\} = 0. \tag{6}$$

Next, we will introduce the iteration rules on the variational parameters $\widetilde{\phi}, \widetilde{\gamma}, \widetilde{\alpha}, \widetilde{\beta}, \widetilde{\mu}$ and $\widetilde{\xi}$, respectively. We also need to add that, $\widetilde{\phi}$ and $\widetilde{\xi}$ are denoted at the node level, which are referred to as local variational parameters. On the contrary, the parameters $\widetilde{\alpha}$, $\widetilde{\beta}$ and $\widetilde{\gamma}$ are global for they are designed at the community level.

B. Parameter Learning

The iteration rules on the variational parameters is given by:

$$\widetilde{\gamma}_k = \gamma_k + \sum_{i=1}^{N} \widetilde{\phi}_{ik}^1, \tag{7}$$

$$\widetilde{\xi}_{ik}^t \propto \widetilde{\phi}_{ik}^t + \sum_l \widetilde{\phi}_{il}^{t-1} \Big(\frac{\widetilde{\mu}_{kl}}{\sum_l \widetilde{\mu}_{kl}} - 1 \Big) + 1, \tag{8}$$

$$\widetilde{\alpha}_{kk} = \alpha_{kk} + \frac{\sum_t \sum_{i<j} \widetilde{\phi}_{ik}^t \widetilde{\phi}_{jk}^t w_{ij}^t}{T}, \tag{9}$$

$$\widetilde{\beta}_{kk} = \beta_{kk} + \frac{\sum_t \sum_{i<j} \widetilde{\phi}_{ik}^t \widetilde{\phi}_{jk}^t (1 - w_{ij}^t)}{T}, \tag{10}$$

$$\widetilde{\alpha}_{kl} = \alpha_{kl} + \frac{\sum_t \sum_{i \neq j} \widetilde{\phi}_{ik}^t \widetilde{\phi}_{jl}^t w_{ij}^t}{T}, \tag{11}$$

$$\widetilde{\beta}_{kl} = \beta_{kl} + \frac{\sum_t \sum_{i \neq j} \widetilde{\phi}_{ik}^t \widetilde{\phi}_{jl}^t (1 - w_{ij}^t)}{T}, \tag{12}$$

$$\widetilde{\mu}_{kl} \propto \mu_l + \frac{\sum_{t=2}^{T} \sum_i \widetilde{\phi}_{ik}^{t-1}}{T - 1}. \tag{13}$$

For parameter $\widetilde{\phi}$:

1) When $t = 1$*:*

$$\begin{aligned} \widetilde{\phi}_{ik}^1 \propto \exp \Big\{ & \sum_j \sum_l \widetilde{\phi}_{jl}^1 \Big[w_{ij}^1 [\psi(\widetilde{\alpha}_{kl}) - \psi(\widetilde{\alpha}_{kl} + \widetilde{\beta}_{kl})] \\ & + (1 - w_{ij}^1)[\psi(\widetilde{\beta}_{kl}) - \psi(\widetilde{\alpha}_{kl} + \widetilde{\beta}_{kl})] \Big] \\ & + \psi(\widetilde{\gamma}_k) - \psi(\sum_l \widetilde{\gamma}_l) + \sum_l \psi(\widetilde{\mu}_{kl}) - \psi(\sum_l \widetilde{\mu}_{kl}) \\ & + \sum_l (\frac{\widetilde{\mu}_{kl}}{\sum_l \widetilde{\mu}_{kl}} - 1)\big(\psi(\widetilde{\xi}_{ik}^2) - \psi(\sum_l \widetilde{\xi}_{il}^2)\big) \Big\}. \end{aligned} \tag{14}$$

2) When $1 < t < T$*:*

$$\begin{aligned} \widetilde{\phi}_{ik}^t \propto \exp \Big\{ & \sum_j \sum_l \widetilde{\phi}_{jl}^t \Big[w_{ij}^t [\psi(\widetilde{\alpha}_{kl}) - \psi(\widetilde{\alpha}_{kl} + \widetilde{\beta}_{kl})] \\ & + (1 - w_{ij}^t)[\psi(\widetilde{\beta}_{kl}) - \psi(\widetilde{\alpha}_{kl} + \widetilde{\beta}_{kl})] \Big] \\ & + \psi(\widetilde{\xi}_{ik}^t) - \psi(\sum_l \widetilde{\xi}_{il}^t) + \sum_l \psi(\widetilde{\mu}_{kl}) - \psi(\sum_l \widetilde{\mu}_{kl}) \\ & + \sum_l (\frac{\widetilde{\mu}_{kl}}{\sum_l \widetilde{\mu}_{kl}} - 1)\big(\psi(\widetilde{\xi}_{ik}^{t+1}) - \psi(\sum_l \widetilde{\xi}_{il}^{t+1})\big) \Big\}. \end{aligned} \tag{15}$$

3) When $t = T$*:*

$$\begin{aligned} \widetilde{\phi}_{ik}^T \propto \exp \Big\{ & \sum_j \sum_l \widetilde{\phi}_{jl}^T [w_{ij}^T [\psi(\widetilde{\alpha}_{kl}) - \psi(\widetilde{\alpha}_{kl} + \widetilde{\beta}_{kl})] \\ & + (1 - w_{ij}^T)[\psi(\widetilde{\beta}_{kl}) - \psi(\widetilde{\alpha}_{kl} + \widetilde{\beta}_{kl})]] \\ & + \psi(\widetilde{\xi}_{ik}^T) - \psi(\sum_l \widetilde{\xi}_{il}^T) \Big\}, \end{aligned} \tag{16}$$

where $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)} = \frac{d \log \Gamma(x)}{dx}$.

The inference detail of variational parameters can be seen in Supplemental Materials (APPENDIX A). As we can see, the hyperparameters of the model such as μ and α are constants, and their values have no significant effect on the performance of the model and algorithm.

C. The Algorithm

As mentioned above, we have given all the update rules of parameters in our model. Here the optimization process of the model is given as Algorithm 1, where $\widetilde{\xi}$ and $\widetilde{\phi}$ are local variational parameters because they are related to each node, and $\widetilde{\alpha}$, $\widetilde{\beta}$, $\widetilde{\gamma}$ and $\widetilde{\mu}$ are global variational parameters.

Algorithm 1 Optimization algorithm for HB-DSBM

```
Input: the adjacency matrix $W^t$ for each snapshot, hyper-parameters, max iteration times $n_{max}$ and the threshold $\varepsilon$.
Output: the variational parameters $\widetilde{\alpha}$, $\widetilde{\beta}$, $\widetilde{\gamma}$, $\widetilde{\mu}$, $\widetilde{\xi}$, $\widetilde{\phi}$.
parameters initialization
repeat
    given $\widetilde{\phi}$, update $\widetilde{\gamma}$, $\widetilde{\xi}$, $\widetilde{\alpha}$, $\widetilde{\beta}$ and $\widetilde{\mu}$ according to Eqs. 7, 8, 9, 10, 11, 12 and 13.
    given $\widetilde{\gamma}$, $\widetilde{\xi}$, $\widetilde{\alpha}$, $\widetilde{\beta}$ and $\widetilde{\mu}$, update $\widetilde{\phi}$ according to Eqs. 14, 15 and 16.
until $\widetilde{L}$ converges or iteration times $> n_{max}$
return $\widetilde{\xi}$, $\widetilde{\gamma}$, $\widetilde{\alpha}$, $\widetilde{\beta}$, $\widetilde{\mu}$ and $\widetilde{\phi}$
```

The computational complexity of the proposed algorithm mainly depends on three parts. The complexity of updating ϕ is $O(TN^2K^2)$, the complexity of the step that updates ξ is $O(TNK^2)$ and the complexity of the step that calculates the ELBO is $O(TN^2K^2)$, where T is the number of snapshots, N is the number of nodes in the network and K is the number of communities. In summary, the computational complexity of this algorithm is $O(TN^2K^2)$.

Considering that most real-world networks are sparse, we can further improve the efficiency and reduce running time and complexity as $O(mK^2)$, where m is the number of edges for all snapshots of the dynamic network. We can further reduce computing time by using some sampling methods based on stochastic optimization or parallelism for the proposed algorithm.

V. EXPERIMENTS

In this section, we conduct several experiments to demonstrate the performance of our proposed model on temporal communities detection, community and node evolution and abnormal behaviors identification of dynamic networks.

A. Baselines and Settings

We compare our HB-DSBM with typical and representative temporal community detection methods. They are designed either from spectral methods, incremental optimization, defined optimization functions, generative models or global views. These baselines could represent the best level of community analysis in dynamic networks, which are listed as follows:

- **ECD** [54]: It combines the proposed new genetic operator and classic genetic operators to exploit inter and intra connections between nodes. This approach improves the discovery of evolving community structures and finds the best balance between clustering accuracy and temporal smoothness.
- **DECS** [21]: It is a novel algorithm based on genome representation, employing Population Generation via Label Propagation (PGLP) for population initialization and decomposition framework for multi-objective optimization.
- **ESPRA** [55]: It is a density-based method by combining the resource allocation (RA) index in link prediction and structural perturbation model [56] to improve the community detection.
- **GenLouvain** [29]: It denotes a generalized network function, time-dependent modularity, and can model the temporal coupling across the snapshots with heuristic optimization.
- **AFFECT** [57]: It extends the classical evolutionary clustering [58] with adaptive evolution factor, so no balance parameter is needed to be designed. It is a general framework for temporal community analysis with different clustering methods.
- **DYNMOGA** [28]: It generalizes the evolutionary clustering based on a multiobjective optimization algorithm, and can compromise the snapshot quality and historical cost effectively and determine the number of communities in dynamic networks.
- **DSBM** [18]: It is the most successful generative model for dynamic community detection and evolution analysis based on SBM.
- **PisCES** [32]: From the perspective of spectral optimization, this is a global method that can infer the evolution by combining a series of networks, eigenvector smoothing and degree correction.

For a fair comparison, we take the open codes by the authors and set the default parameters in the original papers. For our HB-DSBM, since the values of the hyperparameters do not significantly affect the results, without loss of generality, we set $\alpha_{kl} = 10$ when $k \neq l$ and $\alpha_{kk} = N$, and set $\beta = \alpha kl, l \neq k = 10$. Besides, μ and γ are also set to the same value, i.e., $\mu_k = \gamma_k = \frac{1}{K}$ for $1 < k < K$ in our experiments.

B. Performance Metrics and Datasets

To compare the performance of HB-DSBM and baseline methods, we introduce some indicators as our performance metrics. We also show the synthetic and real dynamic networks.

1) Evaluation Index: Accuracy (AC) or error rate [59] is usually denoted as the distance between the ground truth and community membership of one method. Its definition is as follows:

$$AC = \|ZZ^T - Z'Z'^T\|_F, \tag{17}$$

where Z and Z' are the community membership of ground truth and one method, respectively. $\|\cdot\|_F$ is the Frobenius norm, the smaller AC value on each snapshot, the better the community results.

As most temporal community detection work [60] does, we also use Normalized Mutual Information (NMI) as one of our performance metrics to evaluate the proposed model and baselines for community detection in dynamic networks. Because NMI is specifically designed for static networks, we calculate it for different methods on each snapshot of the dynamic network. NMI is used when there exists ground truth, which measures the similarity between a given community partition and the true community structure. Let $Z = \{Z_1, \cdots, Z_K\}$ and $Z' = \{Z'_1, \cdots, Z'_K\}$ represent the true community partition and the community partition to be evaluated, respectively, where Z_k or Z'_k is the nodes set of community k. For Z and Z', we usually have $Z_k \cap Z_l = \emptyset, k \neq l$ and $\bigcup Z_k$ is the node set of dynamic networks. The NMI is denoted as:

$$\text{NMI}(Z, Z') = \frac{\sum_{Z,Z'} p(Z, Z') \log \frac{p(Z,Z')}{p(Z)p(Z')}}{\max(H(Z), H(Z'))}, \tag{18}$$

where $H(Z)$ and $H(Z')$ are the entropy of community Z and Z', respectively. The value of NMI is between 0 and 1. The higher the value of NMI is, the more similar the given community partition is to the true community partition.

Adjusted Rand index is another metric for clustering and community detection performance, which is defined as:

$$ARI = \frac{Index - \mathbb{E}[Index]}{\max(Index) - \mathbb{E}[Index]}, \tag{19}$$

where $Index$ is the Rand index value of a community, defined as $RI = \frac{a+b}{\binom{n}{2}}$, where a and b are the number of node pairs placed in the same cluster and in different clusters, respectively. $\mathbb{E}[Index]$ is the expectation of $Index$. A larger ARI value indicates better performance on community detection.

2) Datasets: We use the following synthetic and real-world datasets of dynamic networks to evaluate HB-DSBM and the baselines for temporal community analysis:

- **Synthetic Dataset 1**: It was firstly adopted by Lin *et al.* [16], [59] based on the GN network data. Specifically, it usually generates a temporal network with 128 or 256 nodes, 4 communities and 10 snapshots. A single

Fig. 2: Community detection results on the synthetic dataset, the x-axis is snapshot t. From left to right, the synthetic temporal networks are parameterized as synthetic dataset 1: (a) $\sigma = 5, nC = 9, aD = 20$; (b) $\sigma = 5, nC = 3, aD = 20$; (c) $\sigma = 4, nC = 9, aD = 16$; (d) $\sigma = 4, nC = 3, aD = 16$ and synthetic dataset 2: (e) switch; (f) merge-split. The top-down results are presented based on the different metrics AC, NMI and ARI, respectively. For each data, we randomly generate 20 networks with the same configuration parameters, and the values are the mean of corresponding results.

parameter σ representing the mean number of edges from a node to nodes in other communities is used to describe the data. And we use nC to indicate the number of nodes to leave their original community in each snapshot while aD to indicate the average degree of nodes. Relatively speaking, this data has s stable dynamic behaviors across the snapshots. With different parameter settings, we have different temporal networks to evaluate the models.

- **Synthetic Dataset 2**: It was proposed by [8] to describe community dynamics. For this dataset, several community-level events are introduced to make it more similar to real-world networks. Generally, each generated network contains 10 snapshots and $1,000$ nodes, with an average degree of 15 and a maximum degree of 50. The number of communities ranges from 20 to 50, and the probability of edges between communities is 0.2. Its node degree follows the power-law distribution. In our experiments, we select two classes of datasets, namely, switch and merge-split, which are the most representative temporal networks with dynamic behaviors at the node and community levels, respectively.
- **KIT-Email Dataset**: It is an email network [61] where nodes represent senders and recipients and the edge denotes the relationships. It has $1,097$ email IDs (the number of nodes) and $27,887$ messages. We use it to construct three temporal networks with time intervals of 2, 3 and 6 months, respectively, and the number of snapshots of the three temporal networks is 24, 16 and 8, respectively. The number of communities ranges from 23 to 32.
- **DBLP Dataset**: It comes from the DBLP bibliography[1]. Similar to the processing of DSBM [18], we select data from three major fields, i.e., data mining (DM), database (DB) and artificial intelligence (AI) from 28 conferences during nine years. After preprocessing, it contains $1,163$ authors and $26,986$ papers. We construct this temporal network based on the cooperative relationships and the ground truth is their research fields. We split this network into 9 time snapshots with each corresponding to one year.
- **A Patent Dataset**: We collect this data from WANFANG DATA[2] and focus on one university. It includes $3,427$ patents ranging from 2010 to 2020, belonging to the patent classification number $G06$ (COMPUTING; CALCULATING; COUNTING). We construct its author collaboration network. Due to the imbalance of patent counts, we divide it into 4 snapshots, which come from years ($2010-2014$, $2014-2016$, $2016-2018$, $2018-2020$), respectively, and each snapshot contains $1,942$ nodes and $15,000$ edges on average.

C. *Experiments on Community Detection*

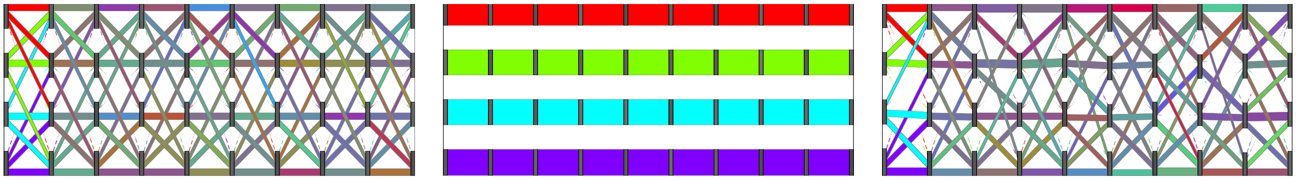
We compare HB-DSBM with ECD, DECS, ESPRA, DYNOMGA, DSBM, GenLou, and PisCES on two types of synthetic datasets and real-world temporal networks.

1) Synthetic Networks: For Synthetic Dataset 1, we generate four temporal networks, the specific parameter settings are $\sigma = 5, nC = 9, aD = 20$; $\sigma = 5, nC = 3, aD = 20$; $\sigma = 4, nC = 9, aD = 16$; $\sigma = 4, nC = 3, aD = 16$,

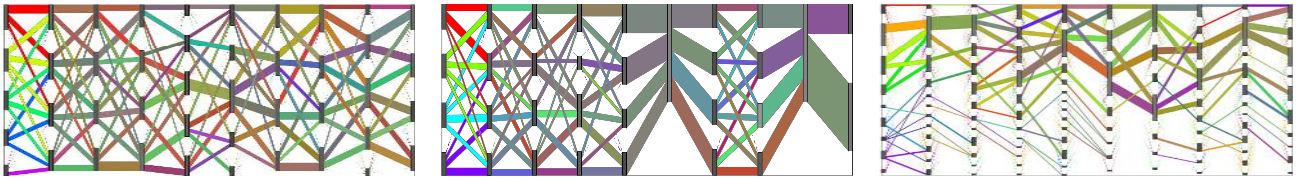
---

[1] https://dblp.uni-trier.de/

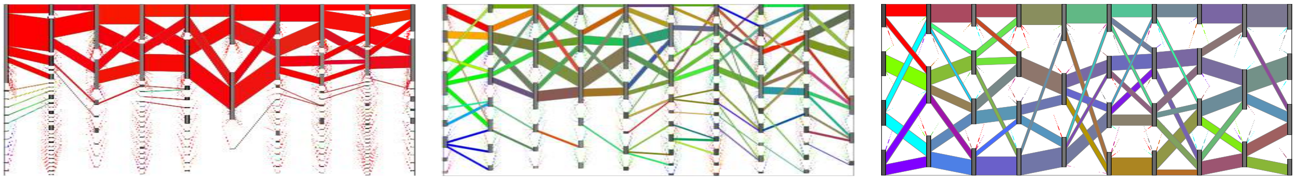
[2] http://www.wanfangdata.com.cn/index.html



(a) GroundTruth (b) DSBM (c) HB-DSBM



(d) GenLouvain (e) PisCES (f) DYNMOGA



(g) DECS (h) ESPRA (i) AFFECT

Fig. 3: Sankey diagram of nodes and communities evolution on Synthetic Dataset 1 with $\sigma = 5$, $nC = 9$ and $aD = 20$.

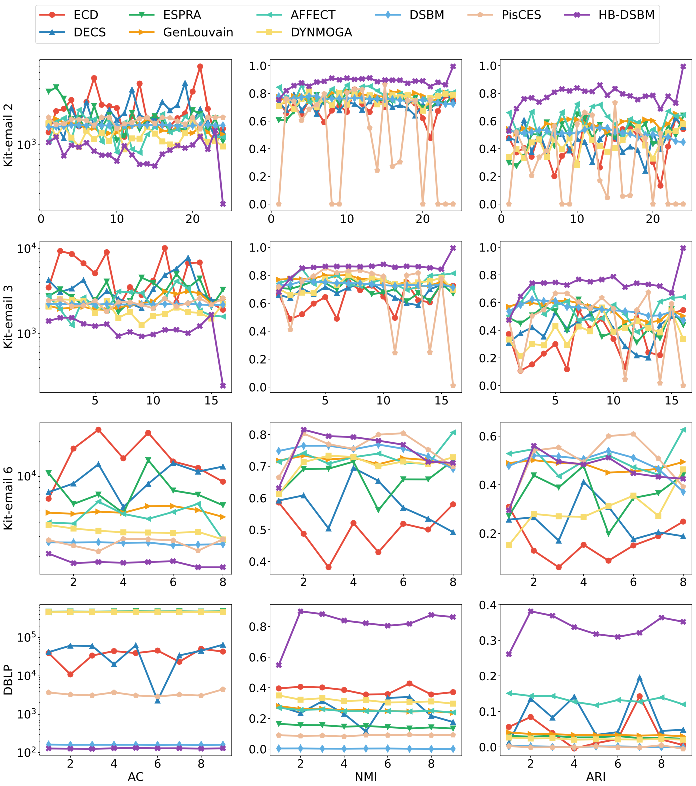
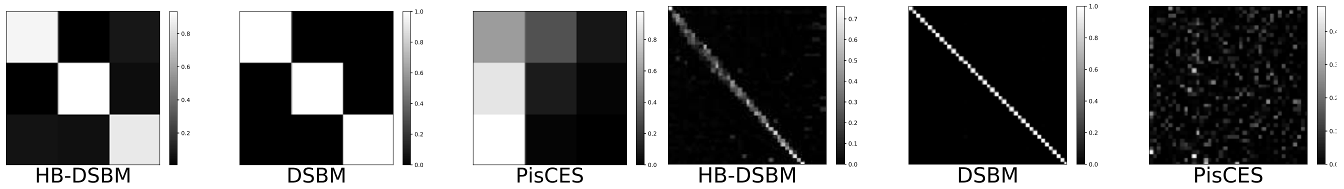


Fig. 4: Community detection results on real-world datasets. From left to right, the KIT email temporal networks with 24, 16 and 8 snapshots, respectively, and the DBLP network. For each line, the results are presented based on different metrics AC, NMI and ARI, respectively.

respectively. They all have 128 nodes and 4 communities, but different degrees of dynamic evolution behaviors. For Synthetic Dataset 2, we select two typical synthetic networks, one is specially designed for the dynamic behavior of nodes and the other is for communities, called switch and merge-split, respectively. They all have $1,000$ nodes and varying the number of communities and their degree distributions are all power-law. For each temporal network, we report the results based on the three metrics as shown in Fig. 2. It is easy to know that HB-DSBM has achieved the best performance based on the AC, NMI and ARI on all six different dynamic networks. This is because our model can model both the community detection and its evolution, so accurate community and node behaviors can help to improve the detection results. PisCES achieves the second-best performance because it is committed to improving community structure by incorporating the evolution of the temporal network across time. However, PisCES has very poor performance on the merge-split network, in which the communities have great varying behaviors, so it cannot cope with the community evolution. Although GenLouvain has also achieved good results since it can preserve the consistency of community structure over time by coupling constraints, it usually overfits the network structure and automatically determines that the number of communities exceeds the ground truth. The performance of ECD, DECS, ESPRA, DSBM and AFFEXT is somewhere in between and is difficult to distinguish on different networks.

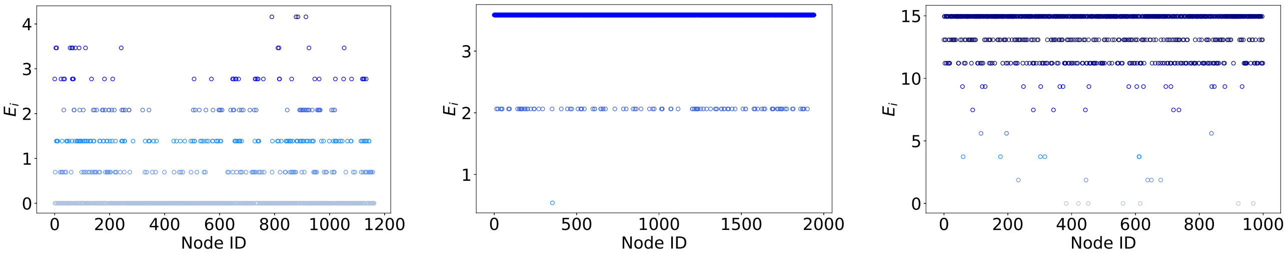
From Fig. 2 (a), (b), (c) and (d), HB-DSBM and other baselines all present the results of serialization on small-scale temporal networks. However, for Fig. 2 (e) and (f), these methods are polarized on community detection in large-scale temporal networks with more complex dynamic behaviors. To be specific, HB-DSBM still achieves the best performance



(a) DBLP

(b) Synthetic Dataset 2 of merge-split event

Fig. 5: Block matrix or community interaction on DBLP data and Synthetic Dataset 2 merge-split event, where the diagonal line indicates the link probability within the community.



(a) DBLP

(b) Patent

(c) Synthetic Dataset 2 merge-split event

Fig. 6: Anomaly index of node-level transform trends. Note that Patent and Synthetic datasets have the same anomaly index patterns, which indicates that the Patent network also contains a dramatic community evolution.

on both networks, PisCES has better performance than other baselines, while DSBM has a very low NMI because its sampling algorithm makes it inefficient in large networks, even with $1,000$ nodes at each snapshot. ECD, DECS and DYNOMGA also have poor performance, especially on the switch network. Without loss of generality, the models with the community and network evolution analysis usually have better results than others.

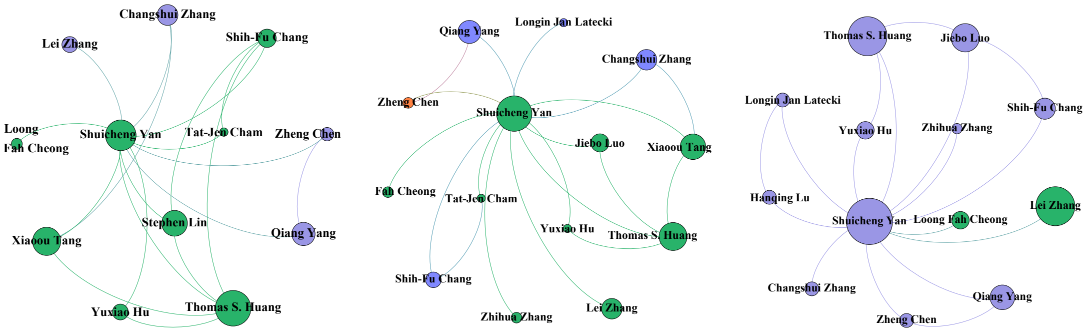
2) Real-World Networks: As for real-world datasets, we also evaluate the performance of different methods in terms of AC, NMI and ARI. We compare the results of our proposed method with the baselines of the KIT-email and DBLP dynamic networks. As we can see from Fig. 4, on the three KIT-email networks, HB-DSBM has the best performance than other baselines on two networks no matter based on the AC, NMI or ARI. On the temporal network with 6-month intervals, our model also has competitive results, for that in this network, each snapshot of which is relatively independent of the other two. So we can conclude that HB-DSBM not only has a stronger ability for community detection but also better simulates the characteristics of the real-world networks. We also note that the NMI and ARI of our method are roughly equal to that of other methods in the first snapshot. The later the snapshots are, the better the performance of our method, which shows a general upward trend. This is because the transition has no impact on the first snapshot in the model, while from the second snapshot, the advantage of our model in the detailed description of node transition heterogeneity is reflected. Besides, DSBM also has competitive performance. To sum up, the statistical modes with dynamic evolution are more suitable for real-world temporal networks.

The experimental results on DBLP are shown in Fig. 4 (d). The performance of our method is significantly superior to that of other methods, which demonstrates the effectiveness of our method in community detection. Among other methods, DSBM has a better effect because it is a generative model, while the sparsity of DBLP data and the parameter sensitivity of PisCES lead to the poor effect of PisCES. It can be seen from this figure that our curve is smoother than that of other methods, because we put community detection and community evolution in a unified framework, making full use of the transition between nodes and communities.

D. Community Evolution Analysis

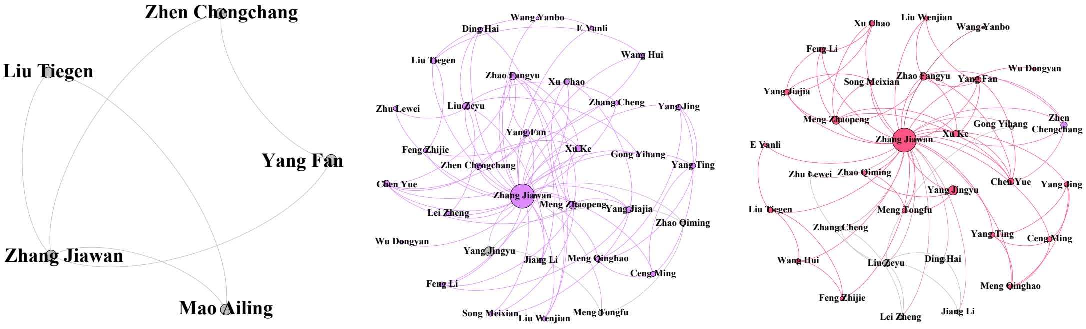
Previous models, such as DSBM, treat nodes in the same community indiscriminately, i.e., two nodes in the same community will have the same transform tendency in the next snapshot, which means that the method can only reveal community-level transform trends. In our model, the latent parameters C and A represent the node- and community-level transform trends, respectively. For the comparison of community evolution, if one method could not analyze the dynamic behavior, we will match the temporal communities across the snapshots and show its dynamic evolution. For the synthetic datasets, we select two representative networks for analyzing the dynamic behaviors. Figs. 3 and (APPENDIX D Fig.5) describe the Sankey diagram of evolution of the networks on Synthetic Dataset 1 and Synthetic Dataset 2 merge-split, respectively. In addition, Sankey diagram is a kind of flow diagram, where the width of every branch is proportional to the flow rate, which is $\frac{\textit{the number of nodes}}{N}$ in our diagram. So the Sankey diagram can visualize the community evolution in the dynamic network.

For the first temporal network, there are four communities and a stable dynamic evolution, Fig. 3 (a) is the ground



(a) Year 2008 (b) Year 2009 (c) Year 2010

Fig. 7: An anomaly case in DBLP dataset from 2008 to 2010, where *Shuicheng Yan* has the largest anomaly index.



(a) Year 2015 – 2016 (b) Year 2017 – 2018 (c) Year 2018 – 2020

Fig. 8: An anomaly case in Patent Dataset from 2015 to 2020. *Zhang Jiawan* belongs to the School of Software Engineering. Note that his ego network expanded rapidly from 2016 to 2018, and it contains many people belonging to other communities, such as the School of Electrical and Information Engineering during 2018 – 2020.

truth, and other subfigures are the results of different methods. Although all methods show good performance on each snapshot, few methods can detect the evolution perfectly. DECS, DYNMOGA and ESPRA present a chaotic evolutionary trajectory. On the contrary, DSBM, GenLouvain and our HB-DSBM could reveal the hidden behaviors of the network. However, DSBM fails to perform node transition across the snapshots because it is only designed for the communities, while HB-DSBM is much closer to the real-world evolution of the network.

For more complex situations of merge-split and DBLP networks, we present some better baselines, as shown in (APPENDIX D Fig.5) and (APPENDIX D Fig.6). The behaviors revealed by DYNMOGA are extremely chaotic, because it could not directly model the evolution. Among other methods that could model the evolution of temporal networks, DSBM adapts to the network due to its focus on the community level. Our HB-DSBM is more realistic than GenLouvain and PisCES because it can model the dynamic evolution at both community and node levels.

Furthermore, we analyze the community-level interaction based on the learned parameters. As depicted in Fig. 5, DSBM, PisCES and HB-DSBM can learn the block matrix. It presents the community interaction for two different temporal networks, PisCES has no clear block structure. Compared with DSBM, our HB-DSBM not only learns the link probability within the community, but also preserves interactions between the communities. This result also demonstrates the advantage of our method in discovering the dynamic behavior of nodes.

E. Abnormal Behaviors

Based on the node-level transition matrix C over time in HB-DSBM, we can capture node transition behaviors between consecutive slices, so it can detect node's abnormal behaviors by calculating the entropy of C. More specifically, the anomaly index of node i can be calculated as follows:

$$E_i = -\sum_{t=1}^{T-1} \sum_{k} c_{ik}^{t} \log c_{ik}^{(t+1)}, \tag{20}$$

where a large E_i indicates that node i transfers its community membership across the snapshots frequently, which are usually the abnormal behaviors in many cases of temporal networks.

We calculate the anomaly index on three temporal networks, namely, DBLP, Patent and Synthetic Dataset 2 merge-split event. It can be seen from Fig. 6 that the DBLP network only has the four largest anomaly indices, which indicates that DM, DB and AI have relatively stable community structures during nine years (2001 – 2009). In contrast, most nodes in Patent have the largest anomaly index, we believe Patent data has significant community evolution events from 2014 to 2020. To prove it, we also calculate the anomaly index of

Synthetic Dataset 2 merge-split event (based on the generation and evolution mechanism, there are indeed a large number of evolutionary anomalous nodes). It can be observed from Fig. 6 (b) and (c) that Patent and Synthetic networks have the same patterns, which confirms our inference.

In addition, we find some anomaly and interesting cases in DBLP and Patent networks. Fig. 7 shows some ego networks of the most abnormal authors in DBLP, we can see that *Shuicheng Yan* and his partners *Thomas S. Huang et al.* always belong to the same community, which indicates that they might be in the same team. We emphasize that DBLP Dataset only represents someone who publishes a paper with someone else as an edge. This kind of relationship is a weak relationship. Therefore, *Shuicheng Yan* switches his community frequently only represents that his research involves all three domains. As depicted in Fig. 8, *Zhang Jiawan*'s patent ego network expanded rapidly from 2016 to 2018, which may indicate that he has begun to actively carry out the implementation of scientific research results. However, from 2018 to 2020, he not only switched community, but also contacted other communities like *Liu Zeyu*. According to our investigation, this was caused by a community merger event. The School of Software Engineering and the School of Computer Science are merged into the College of Intelligence and Computing. This further proves that our anomaly index can discover anomalous nodes and events.

F. Complexity Comparison

Though the VEM algorithm of HB-DSBM is an iterative method, it only requires a small number of iterations to reach the local optimum (APPENDIX B). Furthermore, its computational complexity depends on the total number of edges in the whole network, while the real-world networks are always sparse. Thus, the running time of HB-DSBM is acceptable. Fig. 9 shows the execution time of our method and baselines in synthetic dataset 1 with multiplied number of nodes. And all the methods are tested on a PC with 16 Gb memory and Intel Core $i5-7400$ CPU. The average degree of each dataset is fixed at 16, thus with the increase of the number of nodes, the network will become increasingly more sparse. As we can see, DYNMOGA is very fast on each dataset due to multiobjective optimization. GenLouvain is faster than DYNMOGA on the datasets with the number of nodes below 8192, however, when the number of nodes increases to 8192, GenLouvain will receive an OOM (our of memory) error. When the number of nodes is below 2048, ESPRA is faster than HB-DSBM, DSBM, DECS and ECD, but when the number of nodes increases, its run time is not acceptable (OOT). Both DECS and ECD receive errors (OOT and OOM, respectively) when the number of nodes is larger than 1024. Though DSBM has less parameters than HB-DSBM, the MCMC algorithm makes DSBM rather slow in terms of its parameter scale. Besides, if we use the VEM method for the DSBM, it will be faster than the MCMC, but the accuracy will be lost. Finally, although the execution time of HB-DSBM is not faster than GenLouvain and DYNMOGA, it is sufficient to handle a network with 8192 nodes due to good convergence performance. Furthermore, HB-DSBM is still not fast enough to calculate a network with millions of nodes, but this can be solved with parallel technology or the utilize of neural network, this is also what we will do in the future.

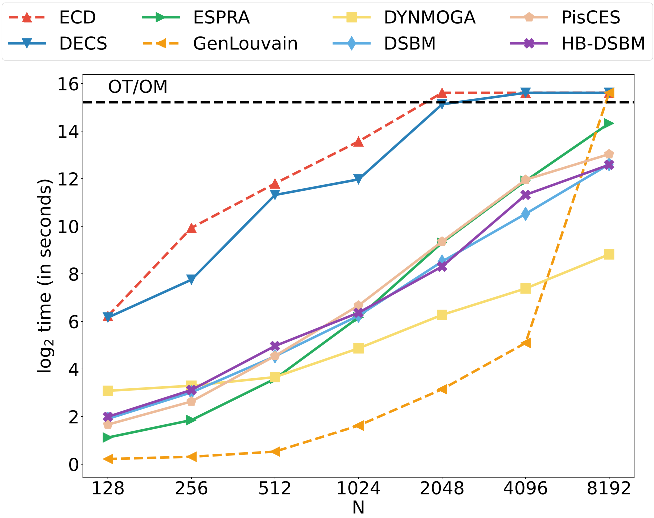


Fig. 9: Execution time comparison on synthetic dataset 1 ($\sigma = 3, nC = 3, aD = 16$) with different numbers of nodes (due to low performance, we omitted AFFECT). The dotted line indicates that the corresponding method receives an OOM (out of memory) error on the data point beyond the OT/OM line. When the data point belongs to a solid line beyond the OT/OM line, it means that the corresponding method receives an OOT (out of time) error.

VI. Conclusion and Future Work

Temporal community detection and its evolution analysis have been widely applied in a variety of applications of network science. Constructing the generated model for dynamic networks can help predict the varying of its structure and function. In this paper, we propose a full Bayesian generative model named HB-DSBM, which models dynamic networks, detects community structure, analyzes node-level and community-level evolution and identifies abnormal behaviors. It describes the generation and evolution of the network in detail from the perspective of joining nodes and communities. Furthermore, we propose an effective optimization algorithm for the model based on the variational inference, in which we design approximate posterior distributions for dynamic transition behavior at the node and community levels. Experiments on community detection, dynamic evolution and abnormal behavior show that this model achieves better performance on both simulated and real-world datasets. On the whole, our model can also reveal the dynamic behavior of communities and nodes based on the well-designed generative mechanism and optimization.

There are some interesting points on our model that can be expanded. The first and most important problem is the model selection of dynamic networks, i.e., how to automatically determine the number of communities. Although a lot of methods have been developed for static networks and some for dynamic networks, they are not enough to meet the need of generative models, such as our HB-DSBM. Another problem is how to predict community change points based on dynamic behaviors and enable them to achieve better prediction tasks. Besides, the algorithm for our model is not suitable for very large-scale dynamic networks, using Stochastic Gradient Descent (SGD) to optimize HB-DSBM will be the focus of future research.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (61902278, 62071327, 61876193 and 62102262),Shenzhen Sustainable Development Project (KCXFZ20201221173013036) and National Key R&D Program of Jiangxi, China (20212ABCO3W12).

REFERENCES

[1] D. He, X. You, Z. Feng, D. Jin, X. Yang, and W. Zhang, "A network-specific markov random field approach to community detection," in *AAAI*, 2018.

[2] X. Luo and M. Shang, "Symmetric non-negative latent factor models for undirected large networks.," in *IJCAI*, pp. 2435–2442, 2017.

[3] G. Rossetti and R. Cazabet, "Community discovery in dynamic networks: a survey," *ACM Computing Surveys (CSUR)*, vol. 51, no. 2, pp. 1–37, 2018.

[4] F. Ye, C. Chen, Z. Wen, Z. Zheng, W. Chen, and Y. Zhou, "Homophily preserving community detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 8, pp. 2903–2915, 2020.

[5] L. Yang, W. Fan, and N. Bouguila, "Clustering analysis via deep generative models with mixture models," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–11, 2020.

[6] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics reports*, vol. 659, pp. 1–44, 2016.

[7] M. Qiao, J. Yu, W. Bian, Q. Li, and D. Tao, "Improving stochastic block models by incorporating power-law degree characteristic.," in *IJCAI*, pp. 2620–2626, 2017.

[8] D. Greene, D. Doyle, and P. Cunningham, "Tracking the evolution of communities in dynamic social networks," in *ASONAM*, pp. 176–183, IEEE, 2010.

[9] G. Palla, A.-L. Barabási, and T. Vicsek, "Quantifying social group evolution," *Nature*, vol. 446, no. 7136, pp. 664–667, 2007.

[10] W. Wang and X. Li, "Temporal stable community in time-varying networks," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 3, pp. 1508–1520, 2020.

[11] E. G. Tajeuna, M. Bouguessa, and S. Wang, "Modeling and predicting community structure changes in time-evolving social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 6, pp. 1166–1180, 2018.

[12] G. Song, Y. Li, X. Chen, X. He, and J. Tang, "Influential node tracking on dynamic social network: An interchange greedy approach," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 2, pp. 359–372, 2016.

[13] Y. Yang, Z. Wang, J. Pei, and E. Chen, "Tracking influential individuals in dynamic networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 11, pp. 2615–2628, 2017.

[14] H. Lin, G. Liu, J. Wu, Y. Zuo, X. Wan, and H. Li, "Fraud detection in dynamic interaction network," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 10, pp. 1936–1950, 2020.

[15] B. Kim, K. H. Lee, L. Xue, and X. Niu, "A review of dynamic network models with latent variables," *Statistics surveys*, vol. 12, p. 105, 2018.

[16] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. L. Tseng, "Facetnet: a framework for analyzing communities and their evolutions in dynamic networks," in *WWW*, pp. 685–694, ACM, 2008.

[17] T. Zhu, P. Li, L. Yu, K. Chen, and Y. Chen, "Change point detection in dynamic networks based on community identification," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 3, pp. 2067–2077, 2020.

[18] T. Yang, Y. Chi, S. Zhu, Y. Gong, and R. Jin, "Detecting communities and their evolutions in dynamic social networks-a bayesian approach," *Machine learning*, vol. 82, no. 2, pp. 157–189, 2011.

[19] K. S. Xu and A. O. Hero, "Dynamic stochastic blockmodels for time-evolving social networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 552–562, 2014.

[20] C. Matias, V. Miele, *et al.*, "Statistical clustering of temporal networks through a dynamic stochastic block model," *Journal of the Royal Statistical Society Series B*, vol. 79, no. 4, pp. 1119–1141, 2017.

[21] F. Liu, J. Wu, S. Xue, C. Zhou, J. Yang, and Q. Sheng, "Detecting the evolving community structure in dynamic social networks," *World Wide Web*, vol. 23, no. 2, pp. 715–733, 2020.

[22] R. C. Cheung, A. Aue, S. Hwang, and T. C. Lee, "Simultaneous detection of multiple change points and community structures in time series of networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 6, pp. 580–591, 2020.

[23] T. Li, W. Wang, X. Wu, H. Wu, P. Jiao, and Y. Yu, "Exploring the transition behavior of nodes in temporal networks based on dynamic community detection," *Future Generation Computer Systems*, vol. 107, pp. 458–468, 2020.

[24] J. Sun, C. Faloutsos, S. Papadimitriou, and P. S. Yu, "Graphscope: Parameter-free mining of large time-evolving graphs," in *SIGKDD*, p. 687–696, 2007.

[25] D. Xu, J. Liang, W. Cheng, H. Wei, H. Chen, and X. Zhang, "Transformer-style relational reasoning with dynamic memory updating for temporal network modeling," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 4546–4554, 2021.

[26] D. Zhuang, M. J. Chang, and M. Li, "Dynamo: Dynamic community detection by incrementally maximizing modularity," *IEEE Transactions on Knowledge and Data Engineering*, 2019.

[27] G. Rossetti, L. Pappalardo, D. Pedreschi, and F. Giannotti, "Tiles: an online algorithm for community discovery in dynamic social networks," *Machine Learning*, vol. 106, no. 8, pp. 1213–1241, 2017.

[28] F. Folino and C. Pizzuti, "An evolutionary multiobjective approach for community discovery in dynamic networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1838–1852, 2013.

[29] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," *Science*, vol. 328, no. 5980, pp. 876–878, 2010.

[30] A. Lancichinetti and S. Fortunato, "Community detection algorithms: A comparative analysis," *Phys. Rev. E*, vol. 80, p. 056117, Nov 2009.

[31] M. Seifikar, S. Farzi, and M. Barati, "C-blondel: An efficient louvain-based dynamic community detection algorithm," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 308–318, 2020.

[32] F. Liu, D. Choi, L. Xie, and K. Roeder, "Global spectral clustering in dynamic networks," *Proceedings of the National Academy of Sciences*, vol. 115, no. 5, pp. 927–932, 2018.

[33] S. Kumar, X. Zhang, and J. Leskovec, "Predicting dynamic embedding trajectory in temporal interaction networks," in *SIGKDD*, pp. 1269–1278, ACM, 2019.

[34] D. K. Sewell, Y. Chen, *et al.*, "Latent space approaches to community detection in dynamic networks," *Bayesian Analysis*, vol. 12, no. 2, pp. 351–377, 2017.

[35] D. K. Sewell and Y. Chen, "Latent space models for dynamic networks," *Journal of the American Statistical Association*, vol. 110, no. 512, pp. 1646–1657, 2015.

[36] K. Xu, "Stochastic block transition models for dynamic networks," in *AISTATS*, pp. 1079–1087, 2015.

[37] A.-K. Becker and H. Holzmann, "Nonparametric identification in the dynamic stochastic block model," *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4335–4344, 2019.

[38] X. Tang and C. C. Yang, "Detecting social media hidden communities using dynamic stochastic blockmodel with temporal dirichlet process," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 2, pp. 1–21, 2014.

[39] L. Yu, W. H. Woodall, and K.-L. Tsui, "Detecting node propensity changes in the dynamic degree corrected stochastic block model," *Social Networks*, vol. 54, pp. 209–227, 2018.

[40] A. Becker and H. Holzmann, "Nonparametric identification in the dynamic stochastic block model," *IEEE Transactions on Information Theory*, vol. 65, pp. 4335–4344, July 2019.

[41] S. Yang and H. Koeppl, "A poisson gamma probabilistic model for latent node-group memberships in dynamic networks," in *AAAI*, 2018.

[42] X. Ma and D. Dong, "Evolutionary nonnegative matrix factorization algorithms for community detection in dynamic networks," *IEEE transactions on knowledge and data engineering*, vol. 29, no. 5, pp. 1045–1058, 2017.

[43] L. Wang, M. Rege, M. Dong, and Y. Ding, "Low-rank kernel matrix factorization for large-scale evolutionary clustering," *IEEE Transactions*

on *Knowledge and Data Engineering*, vol. 24, no. 6, pp. 1036–1050, 2010.

[44] P. Jiao, W. Wang, and D. Jin, “Constrained common cluster based model for community detection in temporal and multiplex networks,” *Neurocomputing*, vol. 275, pp. 768–780, 2018.

[45] D. Koutra, N. Shah, J. T. Vogelstein, B. Gallagher, and C. Faloutsos, “Deltacon: Principled massive-graph similarity function with attribution,” *ACM Transactions on Knowledge Discovery from Data*, vol. 10, Feb. 2016.

[46] D. Koutra, T.-Y. Ke, U. Kang, D. H. P. Chau, H.-K. K. Pao, and C. Faloutsos, “Unifying guilt-by-association approaches: Theorems and fast algorithms,” in *ECML*, pp. 245–260, Springer, 2011.

[47] W. Yu, W. Cheng, C. C. Aggarwal, K. Zhang, H. Chen, and W. Wang, “Netwalk: A flexible deep embedding approach for anomaly detection in dynamic networks,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2672–2681, 2018.

[48] Z. Zhang, D. Chen, L. Bai, J. Wang, and E. R. Hancock, “Graph motif entropy for understanding time-evolving networks,” *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[49] R. A. Rossi, R. Zhou, and N. K. Ahmed, “Estimation of graphlet counts in massive networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 44–57, 2019.

[50] L. Peel and A. Clauset, “Detecting change points in the large-scale structure of evolving networks,” in *AAAI*, pp. 2914–2920, 2015.

[51] D. Jin, K. Wang, G. Zhang, P. Jiao, D. He, F. Fogelman-Soulie, and X. Huang, “Detecting communities with multiplex semantics by distinguishing background, general, and specialized topics,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, pp. 2144–2158, nov 2020.

[52] X. Wu, P. Jiao, Y. Wang, T. Li, W. Wang, and B. Wang, “Dynamic stochastic block model with scale-free characteristic for temporal complex networks,” in *Database Systems for Advanced Applications*, pp. 502–518, Springer International Publishing, 2019.

[53] C. Zhang, J. Butepage, H. Kjellstrom, and S. Mandt, “Advances in variational inference,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, pp. 2008–2026, aug 2019.

[54] F. Liu, J. Wu, C. Zhou, and J. Yang, “Evolutionary community detection in dynamic social networks,” in *IJCNN*, pp. 1–7, IEEE, 2019.

[55] P. Wang, L. Gao, and X. Ma, “Dynamic community detection based on network structural perturbation and topological similarity,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2017, no. 1, p. 013401, 2017.

[56] L. Lü, L. Pan, T. Zhou, Y.-C. Zhang, and H. E. Stanley, “Toward link predictability of complex networks,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 8, pp. 2325–2330, 2015.

[57] K. S. Xu, M. Kliger, and A. O. Hero Iii, “Adaptive evolutionary clustering,” *Data Mining and Knowledge Discovery*, vol. 28, no. 2, pp. 304–336, 2014.

[58] D. Chakrabarti, R. Kumar, and A. Tomkins, “Evolutionary clustering,” in *SIGKDD*, pp. 554–560, 2006.

[59] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. L. Tseng, “Analyzing communities and their evolutions in dynamic social networks,” *ACM Transactions on Knowledge Discovery from Data*, vol. 3, no. 2, pp. 1–31, 2009.

[60] T. Hartmann, A. Kappes, and D. Wagner, “Clustering evolving networks,” in *Algorithm Engineering*, pp. 280–329, Springer, 2016.

[61] R. Görke, M. Holzer, O. Hopp, J. Theuerkorn, and K. Scheibenberger, “Dynamic network of email communication at the department of informatics at karlsruhe institute of technology,” 2011.

Pengfei Jiao received the Ph.D. degrees in computer science from Tianjin University, Tianjin, China, in 2018. From 2018 to 2021, he was a lecture with the Center of Biosafety Research and Strategy of Tianjin University. He is currently a Professor with the School of Cyberspace, Hangzhou Dianzi University, Hangzhou, China. His current research interests include complex network analysis and its applications.

Tianpeng Li the B.E. and M.S. degrees from Tianjin University, China in 2017 and 2020, respectively. He is currently pursuing a Ph.D degree at the College of Intelligence and Computing, Tianjin University. His current research interests include complex network analysis, dynamic community detection and network embedding.

Huaming Wu received the B.E. and M.S. degrees from Harbin Institute of Technology, China in 2009 and 2011, respectively, both in electrical engineering. He received the Ph.D. degree with the highest honor in computer science at Freie Universität Berlin, Germany in 2015. He is currently an Associate Professor in the Center for Applied Mathematics, Tianjin University, China. His research interests include wireless networks, mobile edge computing, internet of things and complex networks.

Chang-Dong Wang received the Ph.D. degree in computer science in 2013 from Sun Yat-sen University, Guangzhou, China. He is a visiting student at University of Illinois at Chicago from Jan. 2012 to Nov. 2012. He joined Sun Yat-sen University in 2013 as an assistant professor with School of Mobile Information Engineering and now he is currently an associate professor with School of Computer Science and Engineering. His current research interests include machine learning and data mining. He has published over 120 scientific papers in international journals and conferences such as IEEE TPAMI, IEEE TKDE, IEEE TCYB, IEEE TNNLS, ACM TKDD, IEEE TII, IEEE TSMC-Systems, IEEE TSMC-C, KDD, AAAI, IJCAI, CVPR, ICDM, CIKM and SDM.

Dongxiao He received the PhD degrees in computer science from Jilin University, Changchun, China, in 2014. She is currently an associate professor with the College of Intelligence and Computing, Tianjin University, Tianjin, China. She has published more than 50 international journal and conference papers. Her current research interests include data mining and analysis of complex networks.

Wenjun Wang is currently a Professor at the School of College of Intelligence and Computing, Tianjin University. His research interests include computational social science, large-scale data mining, intelligence analysis and multi-layer complex network modeling. He was the principal investigator or was responsible for more than 50 research projects. He has published more than 50 papers on main international journals and conferences.