IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 9, 2019, accepted May 27, 2019, date of publication May 31, 2019, date of current version June 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920237

Modeling the Local and Global Evolution Pattern of Community Structures for Dynamic Networks Analysis

WEI YU[1], WENJUN WANG[1], PENGFEI JIAO[2], HUAMING WU[3], YUEHENG SUN[1], AND MINGHU TANG[4]

[1]College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
[2]Center of Biosafety Research and Strategy, Tianjin University, Tianjin, 300072, China
[3]Center for Applied Mathematics, Tianjin University, Tianjin, 300072, China
[4]School of Computer Science and Technology, Qinghai Nationalities University, Qinghai 810007, China

Corresponding author: Yueheng Sun (yhs@tju.edu.cn)

This work was supported in part by the Project of National Social Science Fund under Grant 15BTQ056 and in part by the National Key R & D Program of China under Grant 2018YFC0809800.

ABSTRACT Exploring and understanding the temporal structure of dynamic networks attract extensive attention over the past few years. Most of these current research focuses on temporal community detection, evolution analysis or link prediction from a mission-oriented perspective. In fact, these three tasks should be not isolated but mutually reinforcing. Transforming these three tasks into a unified framework, it is crucial to extract the evolution pattern, which helps to understand the time-varying characteristics of temporal structure in essence. In addition, to the best of our knowledge, there is no work focusing on modeling and uncovering the local and global evolution pattern hidden in temporal community structure, simultaneously. In this paper, we propose a novel framework based on Orthogonal Nonnegative Matrix Factorization to Explore the Evolution Pattern (ONMF-EEP) for analyzing and predicting the time-varying structures in dynamic networks from local and global perspectives. The nature of this framework assumes that community structures are subject to a local evolution pattern (LEP) at each snapshot, and these LEPs are from a common global evolution pattern (GEP). The framework can synchronously detect temporal community structure, extract evolution pattern, and predict structure including communities and future snapshot links. The extensive experiments on real-world networks and artificial networks demonstrate that our proposed framework is highly effective on the tasks of dynamic network analysis.

INDEX TERMS Orthogonal non-negative matrix factorization (ONMF), temporal community detection, evolutionary pattern extraction, structure prediction.

I. INTRODUCTION

Dynamic networks [1] are usually used to model and explore temporal complex systems effectively in the real world, where the structure and the relationships between entities are time-varying. In recent years, dynamic network analysis [2] is gaining popularity rapidly in a wide variety of application domains, such as cyber-physical systems, social networks, biological networks and communication networks. It poses a great challenge for temporal structure analysis in dynamic networks as which are elusive and complicated. Usually, a dynamic network is represented as a series of snapshots or slices. For exploring and understanding the dynamic networks, researchers mainly focus on three problems: detecting what community structures are at different snapshots, analyzing how communities evolve over time, and predicting which links will appear or disappear in the near future. These three problems correspond to three main research tasks, i.e., temporal community detection, evolution analysis and link prediction in dynamic network analysis, respectively.

Firstly, the temporal community detection, which is different from static community detection [3]–[5], could discover time-varying meaningful structures and functions hidden in dynamic networks. Current research methods can be

The associate editor coordinating the review of this manuscript and approving it for publication was Donghyun Kim.

divided into two categories, i.e., two-step based approach and evolutionary clustering based approach. A two-step strategy based approach [6], which typically detects the community structure independently at each snapshot by using a method designed for static networks, and then partially adjust. However, dynamic networks usually evolve slowly and their community structures of adjacent snapshots are similar. Therefore, two-step based approach ignores the influence of historical information on current community structures so that it is sensitive to noise. For addressing this issue, the other type of methods based on evolutionary clustering [7], [8] are proposed to detect current community structure while considering previous snapshots. A detailed review of community detection in dynamic networks is in [1], [2]. Unfortunately, most of these methods do not capture evolution patterns of communities, which helps to understand the time-varying characteristics of temporal structure in essence.

Secondly, evolution analysis of community structure is usually used for understanding the time-varying characteristics and tracking the trend of temporal structure of complex systems. For example, in a scientific collaboration network, a node represents a researcher, a link represents a collaboration relation between two nodes, and a community usually represents a group of researchers with a same research interest respectively. In most cases, we not only need to know which research field a researcher belongs to, but also the changing trends in the future to follow the academic dynamic. Most of these existing methods are based on heuristic approaches, which first obtain the community structure of each snapshot and then analyze their evolution between adjacent snapshots. Some other existing approaches to model the dynamic evolution from tracking network measures [9]. In addition, someone would like to provide useful intuitions about the community structure changes happening in the underlying network by tracking the evolution [10]. However, these methods lack a mechanism to model the evolution pattern for predicting the future structure including community or link structure in dynamic networks.

Thirdly, the task of link prediction in dynamic networks is to predict near future snapshots according to the past observed snapshots. In the early years, some snapshot collapse based methods were proposed, which transform the snapshots into a single collapsed network and then predict the links [11]. These approaches usually have low accuracy because they predict links solely based on the frequency of links. Later, some methods improved the accuracy of link prediction by considering the topological structure of dynamic networks [12]. Recently, matrix decomposition was introduced to discover low dimensional representations of dynamic networks then dealing with the link prediction problem [13]. These algorithms tend to significantly outperform the others. A few other methods were just unilaterally designed for either predicting community structure or links [14], [15] from a mission-oriented perspective.

In fact, these three problems, community detection, evolution analysis and link prediction are not isolated, but mutually reinforcing. The reason is that modularity and temporal variability of dynamic networks always exist simultaneously. In addition, community structures usually evolve slowly, but their evolution (e.g., community structures changed dramatically) may be anomalous in some abnormal case (e.g., some major emergency broke out). However, most of the existing methods cannot discover the abnormal evolution of community structures hiding in dynamic networks as the evolution extent is hard to measure. In this paper, we propose a novel two-step framework based on Orthogonal Nonnegative Matrix Factorization [16] to Extract Evolution Pattern (ONMF-EEP) for exploring and predicting the time-varying structure in dynamic networks from local and global perspectives. In detail, we propose a hypothesis that community structures are subject to a local evolution pattern (LEP) at each snapshot respectively. At the same time, a common global evolution pattern (GEP) is hidden in the underlying network, which would be infinitely close to these LEPs in normal case. Under this hypothesis, we design ONMF-EEP as two steps: LEP extraction and GEP extraction. In the first step, the proposed framework discovers temporal community structure and detects the corresponding LEP with ONMF based on the core idea of evolution clustering. In the second step, to extract GEP, the framework maximizes the shared information among LEPs by optimizing the distance between the matrix representations of all the LEPs and the GEP on Grassmann manifold. We demonstrate the superior performance of our proposed algorithms over the baseline methods on both artificial and real-world dynamic networks. It is worthwhile to highlight several contributions of our proposed framework ONMF-EEP as follows:

- To our best knowledge, we first extract evolution pattern by modeling and uncovering the evolution characteristics of dynamic networks from local and global perspectives.
- We propose a novel unified framework ONMF-EEP, which is easy to be extended, for exploring and understanding the temporal structure including community detection, evolution analysis, and link structure prediction.
- The extensive experimental results demonstrate that our proposed method ONMF-EEP has good performance for temporal community detection, and has effective ability for evolution analysis and link prediction both on real-world networks and artificial networks.

II. RELATED WORK

Dynamic networks analysis [1] is widely concerned in many fields. Most of the current research about it mainly focus on community detection [2], [17], evolutionary analysis [18], [19], link prediction [13], [14], and anomaly detection [20], [21]. Here, we give a general overview on evolutionary analysis for dynamic network, which is the most related for this work. The methods of evolutionary analysis for dynamic network are mainly divided into three categories: heuristic approaches [22], [23], machine learning based methods [24], [25] and generative model based methods [10], [15], [27].

Heuristic approaches for evolutionary analysis explore usually the community evolution based on some similar criterions after detecting community structure at snapshots. In the early day, Palla *et al.* [22] detect the temporal community structure based on clique percolation, and then uncover basic relationships characterizing community evolution. They define possible events in community evolution including growth, contraction, merging, splitting, birth and death. Later, Asur *et al.* [23] present a characterization of critical behavioral patterns for temporally varying interaction networks based on the events in community evolution. This type of approaches usually depend too much on defined rules and have high computation complexity.

Machine learning based methods for evolutionary analysis tend to identify the features of the detected temporal community structure based on feature learning method. Typically, Ilhan and Ögüdücü [24] proposed a framework that community features identification based on feature selection methods, and study the evolutionary characteristic to predict the community evolution. Liu *et al.* [25] use the existing Louvain method [26] to detect temporal community structure in the bibliographic coupling and co-citation networks, and then train a classifier to predict the evolution of physics research. This type of methods tend to ignore structural information of the historical snapshots and lack of model interpretation.

Generative model-based methods for evolutionary analysis, which is the most related with our proposed method, model the generative mechanism of dynamic networks and the evolution of temporal community structure. Differently from the first two types of approaches, generative model-based methods tend to model the community structure and the evolution characteristic hiding in dynamic networks synchronously. For example, Yu *et al.* [10] model the temporal status of the edges as a function of time-based on NMF in dynamic networks. Their approaches are used to predict links and detect node-centric anomaly. Tajeuna *et al.* [27] model the evolution of temporal community structure with an auto regressive model, and predict future changes with survival analysis techniques. Wu *et al.* [15] introduce the spectral graph theory to track the latent feature vectors of dynamic networks, then use the Finite Impulse Response (FIR) filter to model the evolution of the latent feature vector of each node. This type of approaches usually has reasonable model interpretation and model the community structure and the evolution characteristic uniformly. Unfortunately, this type of methods cannot discover the abnormal evolution of community structure hiding in dynamic networks as the evolution extent is hard to measure.

III. PROBLEM FORMALIZATION

Here, we represent a dynamic network as a series of network snapshots. Let $\{1, 2, ..., T\}$ be a finite set of time snapshots and $\mathcal{G} = (V, E^{(t)})$ be an undirected and unweighed dynamic network, where $V = \{1, ..., N\}$ is a set of entities or nodes, $E^{(t)}$ is a set of edges that connect two nodes of V at snapshot t. $\mathcal{G}$ is represented with a sequence of $N \times N$ adjacency matrix $A^{(t)}$, where the element at snapshot t

$$A_{ij}^{(t)} = \begin{cases} 1 & (i,j) \in E^{(t)} \\ 0 & (i,j) \notin E^{(t)}. \end{cases}$$

In addition, we summarize the main notations in table 1.

TABLE 1. Table of notations.

Symbol	Definition
t	the snapshot label, and t $\in [1, T]$;
V	the set of nodes, and $V = \{1, 2, \cdots, N\}$;
$E^{(t)}$	the set of edges at snapshot t;
$A^{(t)}$	the adjacent matrix at snapshot t;
$H^{(t)}$	the community membership matrix at t, and $H^{(t)} \in R_{+}^{N \times K_t}$;
Z_t	the evolution matrix that models LEP at snapshot t;
$\mathcal{Z}$	the evolution matrix that models GEP;
C_t	the community label matrix of snapshot t.
K	the number of communities.

In general, dynamic networks analysis mainly include three tasks: detecting community structure, extracting evolution pattern and predicting future structure (please see as Fig. 1).

Usually, a static network could be divided into groups of nodes with dense connections internally and sparse connections between groups. This modular structure is called community structure of complex networks, and a group is called a community (or cluster). Accordingly, a dynamic network can be seen as a sequence of static network snapshots. The community structures are represented as a temporal N dimension vector $C^{(t)}$, where the element $C_i^{(t)}$ is the community label of node i at snapshot t. The set of community labels is $\{1, 2, \cdots, K\}$, where K is the number of communities.

We model the evolution patterns of temporal community structure with evolution matrices. To model the LEPs, we introduce a temporal $N \times K$ dimension matrix $Z^{(t-1)}(t \geq 2)$, of which element $Z_{lk}^{(t-1)}$ represents the evolution probability from community l to community k at snapshot $t-1$. Similarly, we introduce a common $N \times K$ dimension matrix $\mathcal{Z}$ to model the GEP.

Link prediction is a common problem in static networks. Its core purpose is to predict missing links according to a static state of the observed network. Differently in dynamic networks, we need to predict a future snapshot network $\hat{A}^{(T+1)}$ according to the trends of the past snapshot networks $\mathcal{G}$: $(V, E^{(t)})$.

The tasks of the proposed framework are summarized as follows.

- *Input:* A dynamic network $\mathcal{G} : (V, E^{(t)})$.
- *Output:* The community structure $C^{(t)}(t = 1, 2, \cdots, T)$, LEPs, GEP, the future community structure $C^{(T+1)}$, the future link structure $\hat{A}^{(T+1)}$.

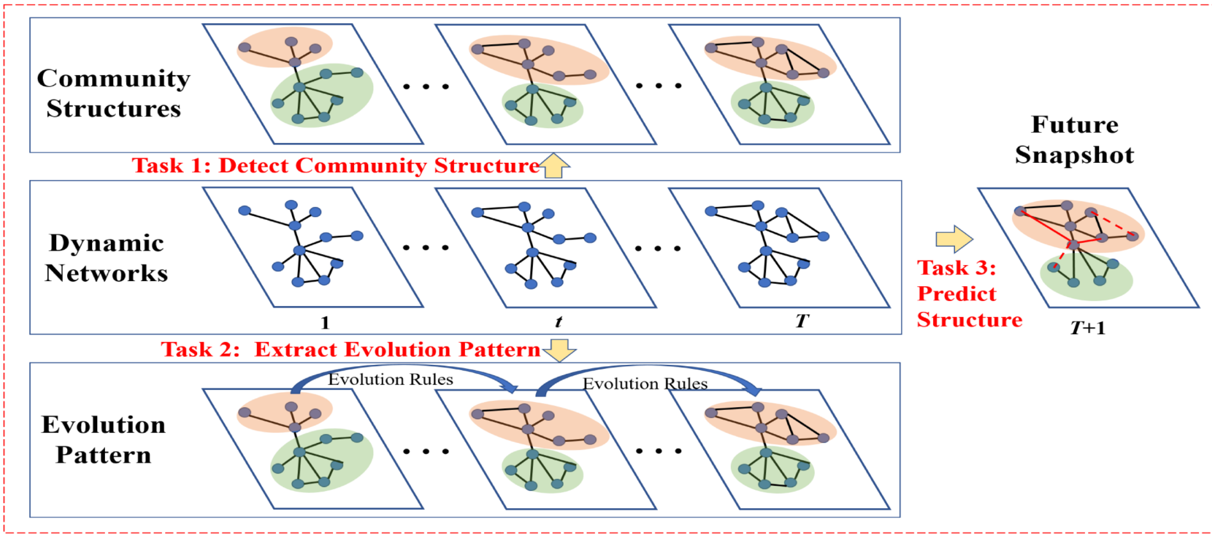


FIGURE 1. Illustration of three main tasks in dynamic networks analysis.

IV. METHODS

In this section, we present our proposed framework ONMF-EEP for exploring dynamic networks including community detection, evolution analysis and link prediction.

A. ONMF-EEP

When Nonnegative Matrix Factorization (NMF) is applied to community detection, the core idea is that leaning a low rank non-negative representation H of the underlying adjacency matrix. And H is just the community membership matrix, of which element H_{ik} represents the preference of node i belongs to community k. It has derived different versions of NMF models including SNMF, PNMF, SNMTF, Semi-NMTF [16]. These methods are all suitable for extending into our unified framework ONMF-EEP designing for dynamic networks.

Our proposed framework ONMF-EEP assumes that community structures are subject to a LEP of community structure at each snapshot respectively, and a common GEP is hidden in them. According to the two assumptions, ONMF-EEP consists of two steps (Fig.2):

- *Step 1 (LEP Extraction):* For each snapshot network, we obtain its non-negative, low-dimensioanl representation, $H^{(t)}$, under column orthonormality constraints i,e., $H^{(t)T}H^{(t)T} = I$, by using extensional Orthogonal NMF (ONMF). In addition, we integrate LEP with a temporal $K \times K$ dimension matrix $Z^{(t-1)}(t \geq 2)$ using the idea of evolutionary clustering [7].
- *Step 2 (GEP Extraction):* To obtain the GEP hidden in a dynamic network, we fuse the LEP at each snapshot into a common GEP, which is introduced with a common $N \times K$ dimension matrix $\mathcal{Z}$, by an ONMF model. The main idea is that minimizing the distance between $\mathcal{Z}$ and each $Z^{(t-1)}$.

Below we just take Orthogonal SNMF as an example to extend to ONMF-EEP for exploring dynamic networks.

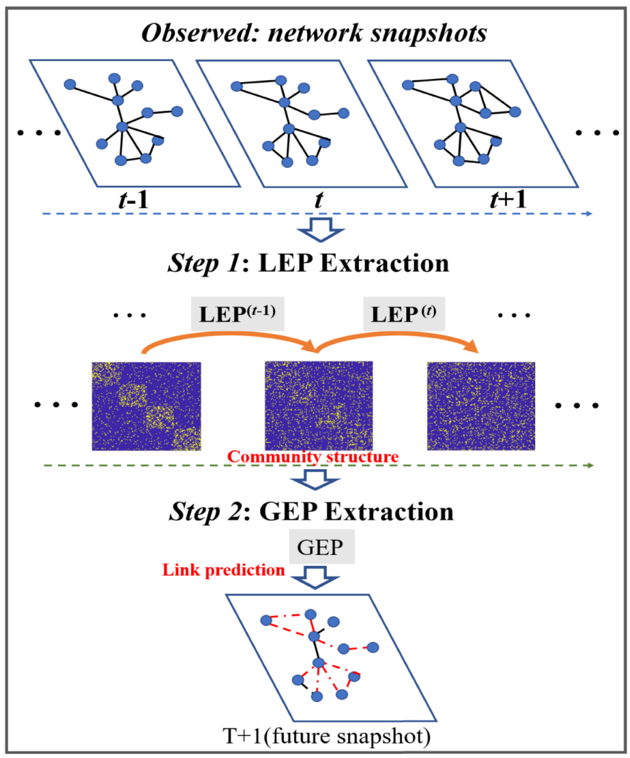


FIGURE 2. Illustration of the proposed ONMF-EEP.

B. STEP 1: LEP EXTRACTION

In a dynamic network, similarly, the element $H_{ik}^{(t)}$ represents the propensity that node i belongs to community k at snapshot t. Differently, we set $H^{(t)}$ under column orthonormality constraints, $H^{(t)T}H^{(t)} = I$, since we focus on non-overlapping community detection here a node just may belong to only one community. Then $H_{i.}^{(t)} \times H_{j.}^{(t)T}$ represents the expected value of the link that exists between node i and j. In fact, to obtain $H^{(t)}$ from the observed adjacency matrix $A^{(t)}$, we just need to make each entry of matrix $H^{(t)}H^{(t)T}$ as close to $A^{(t)}$ as possible. Here, we assume that the difference between $A^{(t)}$ and $H^{(t)}H^{(t)T}$ obeys Gaussian distribution with zero mean. Then the loss can be constructed by Euclidean distance, $||A^{(t)} - H^{(t)}H^{(t)T}||_F^2$, which is the square of the Frobenius norm of two matrices difference [8].

In addition, we assume that community structures are subject to an independent LEP of community structure at each snapshot, respectively. To model the LEPs, we introduce a temporal $N \times K$ dimension matrix $Z^{(t-1)}(t \geq 2)$, where element $Z_{lk}^{(t-1)}$ represents the evolution probability from community l to community k at snapshot $t-1$. Similarly, we need to make each entry of matrix $H^{(t-1)}Z^{(t-1)}$ as close to $H^{(t)}$ as possible. Then we introduce the loss $||H^{(t)} - H^{(t-1)}Z^{(t-1)}||_F^2$ as a penalty term. Then the objective function is constructed as follow:

$$\min_{H^t, Z^{(t-1)} \geq 0} Q^{(t)}(H^{(t)}, Z^{(t-1)}) = \begin{cases} ||A^{(t)} - H^{(t)}H^{(t)T}||_F^2 & t = 1, \\ ||A^{(t)} - H^{(t)}H^{(t)T}||_F^2 \\ + 2\alpha||H^{(t)} - H^{(t-1)}Z^{(t-1)}||_F^2 & t \geq 2, \end{cases} \tag{1}$$

where α is a balance parameter, $H^{(t)T}H^{(t)} = I$. When $t = 1$, the objective function corresponds to the stand ONMF [16], and the update rules are as follows:

$$H^{(1)} \leftarrow H^{(1)} \odot \frac{A^{(1)}H^{(1)}}{H^{(1)}H^{(1)T}A^{(1)}H^{(1)}} \tag{2}$$

where $\cdot \odot \cdot$ represents hadamard product. When $t \geq 2$, the derivative of $Q^{(t)}$ with respect to $H^{(t)}$ is as follows:

$$\frac{\partial Q^{(t)}}{\partial H^{(t)}} = -4A^{(t)}H^{(t)} + 4(1+\alpha)H^{(t)} - 4\alpha H^{(t-1)}Z^{(t-1)} \tag{3}$$

To incorporate the orthonormality constraint into the update rule, the concept of natural gradient needs to be introduced [28]. As the natural gradient $\frac{\partial Q^{(t)}}{\partial H^{(t)}} = \frac{\partial Q^{(t)}}{\partial H^{(t)}} - H^{(t)}H^{(t)T}\frac{\partial Q^{(t)}}{\partial H^{(t)}}$, we have that

$$\begin{aligned}\frac{\partial Q^{(t)}}{\partial H^{(t)}} &= -4A^{(t)}H^{(t)} - 4\alpha H^{(t-1)}Z^{(t-1)} \\ &\quad + 4H^{(t)}H^{(t)T}A^{(t)}H^{(t)} \\ &\quad + 4\alpha H^{(t)}H^{(t)T}H^{(t-1)}Z^{(t-1)}\end{aligned} \tag{4}$$

Following the Karush-Kuhn-Tucker (KKT) condition, the update rule of $H^{(t)}$ is as follows:

$$H^{(t)} \leftarrow H^{(t)} \odot \frac{A^{(t)}H^{(t)} + \alpha H^{(t-1)}Z^{(t-1)}}{H^{(t)}H^{(t)T}A^{(t)}H^{(t)} + \alpha H^{(t)}H^{(t)T}H^{(t-1)}Z^{(t-1)}} \tag{5}$$

For the evolutionary matrix $Z^{(t-1)}$, the derivative of $Q^{(t)}$ with respect to $Z^{(t-1)}$ is as follows:

$$\frac{\partial Q^{(t)}}{\partial Z^{(t-1)}} = -4\alpha H^{(t)}H^{(t-1)T} + 4\alpha Z^{(t-1)} \tag{6}$$

Correspondly, the update rule of $Z^{(t-1)}$ is as follows:

$$Z^{(t-1)} \longleftarrow H^{(t-1)T}H^{(t)} \tag{7}$$

Updating iteratively $H^{(t)}$ and $Z^{(t-1)}$ according to the rule 5 and 7 until the objective function 1 converges. Then the community label of each node at snapshot t is derived by the equation as follows:

$$C_i^{(t)} = \arg\max_k \ (H_{ik}^{(t)}).$$

In addition, $Z^{(t-1)}(t = 2, \cdots, T)$ can represent the LEP of dynamic networks, which model the evolution rules between consecutive snapshot pairs, independently.

C. STEP 2: GEP EXTRACTION

We assume that a common GEP exists behind the entire dynamic network. In most cases, dynamic network evolves slowly and LEPs approach to GEP. In fact, GEP extraction could translate to minimizing the total geodesic distance between a common evolution matrix $\mathcal{Z}$ and the temporal evolution matrices $Z^{(t-1)}(t \geq 2)$ at each snapshot. Then we merge the temporal evolution matrices $Z^{(t-1)}$ of snapshot networks to $\mathcal{Z}$ on Grassmann Manifolds. According to the work of the literature [16], the square distance between $\mathcal{Z}$ and $Z^{(t-1)}$ on Grassmann Manifolds can be computed as $Dis(\mathcal{Z}, Z^{(t-1)}) = k - tr(\mathcal{Z}\mathcal{Z}^T Z^{(t-1)}Z^{(t-1)T})$, where k is the number of dimension of subspaces, $tr(\cdot)$ represents the trace of matrix. Then the objective function of Step 2 is as follow:

$$\begin{aligned}\min_{\mathcal{Z} \geq 0} J(\mathcal{Z}) = \sum_{t=2}^{T} ||H^{(t)} - H^{(t-1)}\mathcal{Z}||_F^2 \\ + \beta \sum_{t=1}^{T-1} (k - tr(\mathcal{Z}\mathcal{Z}^T Z^{(t)}Z^{(t)T})).\end{aligned} \tag{8}$$

The derivative of J with respect to $\mathcal{Z}$ is as follow:

$$\frac{\partial J}{\partial \mathcal{Z}} = \sum_{t=2}^{T}(2\mathcal{Z} - 2H^{(t-1)T}H^{(t)}) - 2\beta \sum_{t=1}^{T-1} Z^{(t)}Z^{(t)T}\mathcal{Z} \tag{9}$$

Similarly, the update rule of $\mathcal{Z}$ is as follow:

$$\mathcal{Z} \longleftarrow \mathcal{Z} \odot \frac{\sum_{t=2}^{T}(H^{(t-1)T}H^{(t)}) + \beta \sum_{t=1}^{T-1} Z^{(t)}Z^{(t)T}\mathcal{Z}}{\sum_{t=2}^{T} \mathcal{Z}}, \tag{10}$$

where β is a balance parameter for controlling the weight of the geodesic distance. $\mathcal{Z}$ represents the GEP, which represents the global evolution characteristic of dynamic networks.

Then based on the GEP, we can predict the community membership matrix at snapshot $T+1$ according to $\hat{H}^{(T+1)} \approx H^{(T)}\mathcal{Z}$. Naturally, we can predict the community structure at snapshot $T+1$ with $\hat{C}_i^{(T+1)} = \arg\max_k \ (H_{ik}^{(T+1)})$. At the same time, we can predict the link structures by reconstructing the adjacency matrix at snapshot t with $\hat{A}^{(T+1)} \approx H^{(T+1)}H^{(T+1)T}$ according to the model hypothesis.

Algorithm 1 ONMF-EEP

Input: $A^{(t)}(t = 1, 2, \cdots, T)$, α and β
Output: $C^{(t)}$, $H^{(t)}$, $Z^{(t-1)}$, $\mathcal{Z}$
Initialize $H^{(t)}$, $Z^{(t-1)}$, $\mathcal{Z}$;
while not converge **do**
    Update $H^{(1)}$ according to Eq.2;
for $t \in [2, T]$ **do**
    while not converge **do**
        Update $H^{(t)}$ according to Eq.5;
        Update $Z^{(t-1)}$ according to Eq.7;
for $t \in [1, T]$ **do**
    $C_i^{(t)} = \arg\max_k \ (H_{ik}^{(t)})$
while not converge **do**
    Update $\mathcal{Z}$ according to Eq.10;
$\hat{H}^{(T+1)} \approx H^{(T)}\mathcal{Z}$;
$\hat{C}_i^{(T+1)} = \arg\max_k \ (H_{ik}^{(T+1)})$;
$\hat{A}^{(T+1)} \approx H^{(T+1)}H^{(T+1)T}$;
return $C^{(t)}$, $H^{(t)}$, $Z^{(t-1)}$, $\mathcal{Z}$.

D. COMPLEXITY ANALYSIS

Here, we give the optimization algorithm of ONMF-EEP in Algorithm 1 according to the update rules above. We assume that the average number of edges at snapshots is $\hat{m}$, and the average number of iterations is $\hat{n}_{iter}$. From Algorithm 1, the algorithm mainly consists of two steps, and the most time-consuming part is the optimization of $H^{(t)}$, which is $O(\hat{n}_{iter}T(4N^2K + 4NK^2 + 2NK))$. The N^2 can be approximately equal to $\hat{m}$ because the real dynamic networks are usually very sparse. Considering K is much less than N and $\hat{m}$, it could be ignored for time complexity. Therefore, the time complexity degrades to $O(\hat{n}_{iter}T(\hat{m} + N))$, which is approximately linear for the number of nodes and edges in dynamic networks.

V. RESULTS AND DISCUSSION

In this section, we demonstrate the performance of ONMF-EEP for analyzing the time-varying structure including community detection, LEP and GEP extraction and link prediction at future snapshot on synthetic and real-world dynamic networks. We set the balance parameters $\alpha = 10$ and $\beta = 0.4$ on all experiments according our parameter analysis.

A. MEASUREMENTS

To measure the performance of the algorithms, we introduce several evaluation measures for community detection and link prediction of dynamic networks in this subsection.

For community detection, normalized mutual information (NMI) and error rate (ER) are good choices when the ground truth is available [29], [30]. They are defined as follows:

$$\mathrm{NMI} = \frac{2I(\hat{G}, G)}{H(\hat{G}) + H(G)}, \tag{11}$$

$$\mathrm{ER} = ||\hat{G}\hat{G}^T - GG^T||_F^2, \tag{12}$$

where $\hat{G}$ denotes the community structures detected from the algorithm and G denotes the ground truth. $H(\hat{G})$ and $H(G)$ denote the entropies of $\hat{G}$ and G respectively, and $I(\hat{G}, G)$ denote the mutual information between $\hat{G}$ and G respectively. NMI is used to measure the consistency between two partitions as an entropy metric, which is restrained in [0, 1]. The higher the value of NMI, the more similar the two partitions are. In detail, NMI $= 1$ indicates that G and $\hat{G}$ are identical, and NMI $= 0$ indicates that the two partitions are entirely different. ER measures the difference between two different partitions, and the smaller it is, the better the performance is. In general, ER tends to increase with the scale of networks.

For link prediction, we choose the widely-used evaluation metrics, Root Mean Square Error (RMSE), which is defined as

$$\mathrm{RMSE} = \sqrt{||\hat{\mathcal{A}} - \mathcal{A}||}, \tag{13}$$

where $\hat{\mathcal{A}}$ and $\mathcal{A}$ are the predicted and true matrices for the future snaphshot, respectively.

B. TEMPORAL COMMUNITY DETECTION

To demonstrate the performance of community detection, we compare our proposed ONMF-CCP with four benchmark methods: SNMF, Multislice [26], FaceNet [31], DYNMOGA [32]. Here, we set the penalty co-efficient of smoothness as 0.8 for FaceNet, the resolution parameter $\gamma = 1$ and couple parameter $\omega = 0.2$ for Multislice. These are the common parameter settings in the related researches. In addition, all the algorithms are repeated 10 times, and the average results and the corresponding variance bars are given in figures.

We design the comparison experiments over NMI and ER on a synthetic data, dynamic Grow-shrink benchmark [33], which uses a triangular waveform function to model the dynamic of network and generates each snapshot network with the classic Stochastic Block Model [34]. Here, we set the number of snapshots $T = 12$, the number of noes $N = 256$, the number of communities $K = 4$ for generating dynamic networks. Fig.3 shows the performance over NMI and ER on four dynamic Grow-shrink networks with different fuzzy parameters $(a, b)q = 31$, $(c, d)q = 33$, $(e, f)q = 35$, and $(g, h)q = 37$. It should be noted that the larger the fuzzy parameters, the fuzzier the community structure of generated networks. Most methods can easily achieve high accuracy on community detection of generated network if the parameter q is too small. Then the comparative results of methods are not comparable. On the contray, the comparative results of methods are also not comparable if the parameter q is too large. In the figure, the x-axes are snapshot labels t and the y-axes are NMI or ER values. From the Fig.3, we can conclude that

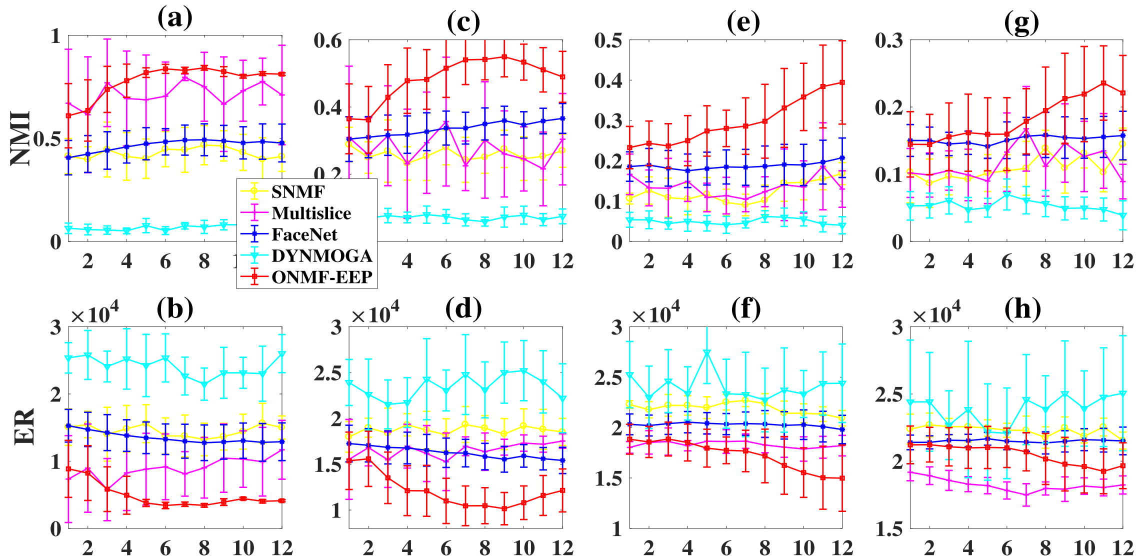


FIGURE 3. The performance of community detection over NMI and ER on the dynamic Grow-shrink networks with different fuzzy parameters $(a, b)q = 31$, $(c, d)q = 33$, $(e, f)q = 35$, and $(g, h)q = 37$.

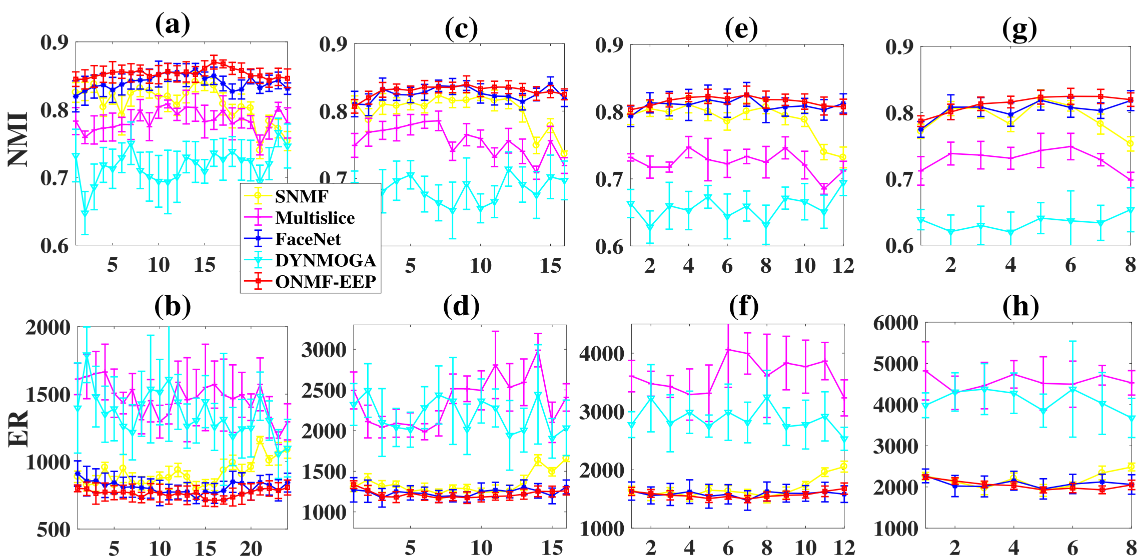


FIGURE 4. The performance of community detection over NMI and ER on the MIT email networks with different resolutions $(a, b)T = 24$, $(c, d)T = 16$, $(e, f)T = 12$, and $(g, h)T = 8$.

the proposed methods always achieves better performance than other four methods for different fuzzy parameters on both NMI and ER. It's worth noting that the ER values of Multislice is the lowest in Fig.3 (h) corresponding to $q = 37$. The reason may be that Multislice optimizes modularity with a greedy heuristic method that it could achieve good performance on ER, even though it tends to have high time complexity. Looking at the subfigures from left to right, the whole trend of the accuracy on both NMI and ER decreases gradually as the fuzzy parameters q increases. Obviously, this is because the community structures of networks are unsharp when the parameter q is large.

In addition, we design the comparison experiments over NMI and ER on a real-world data, e-mail communication network [35]. The data includes 48 consecutive months from September 2006 to August 2010. Here, we cut the network into multi-time slices with different resolutions. As shown in Fig.4, four dynamic networks from the e-mail communication correspond to $(a, b)T = 24$, 2 consecutive months as a snapshot, $(c, d)T = 16$, 3 consecutive months as a snapshot, $(e, f)T = 12$, 4 consecutive months as a snapshot, and $(g, h)T = 8$, 6 consecutive months as a snapshot. Similarly, the x-axes are snapshot labels t and the y-axes are NMI or ER values. From Fig.4, in most cases, the results demonstrate our proposed ONMF-EEP has the best performance over NMI and ER on the real-world email networks with different resolutions. On closer inspection, we find the performance of FaceNet is just slightly inferior to ONMF-EEP on real-world

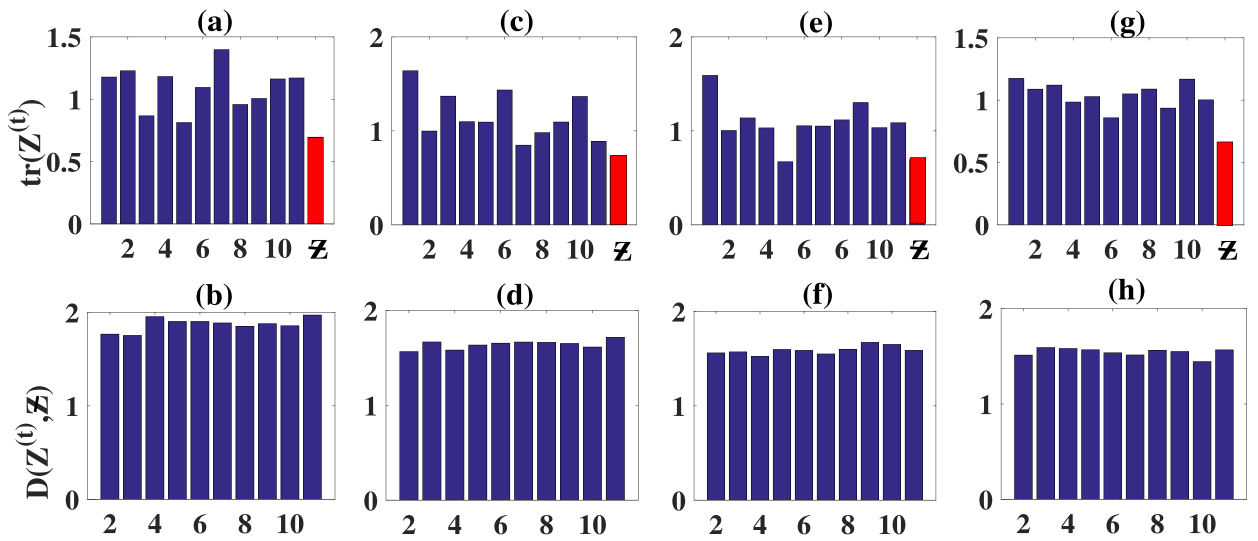


FIGURE 5. The intensity of evolution over time on the dynamic Grow-shrink networks with different fuzzy parameters. (a, c, e, d) are the trace of $Z^{(t)}$ and $\mathcal{Z}$, and (b, d, f, h) are the square distance between $Z^{(t)}$ and $\mathcal{Z}$.

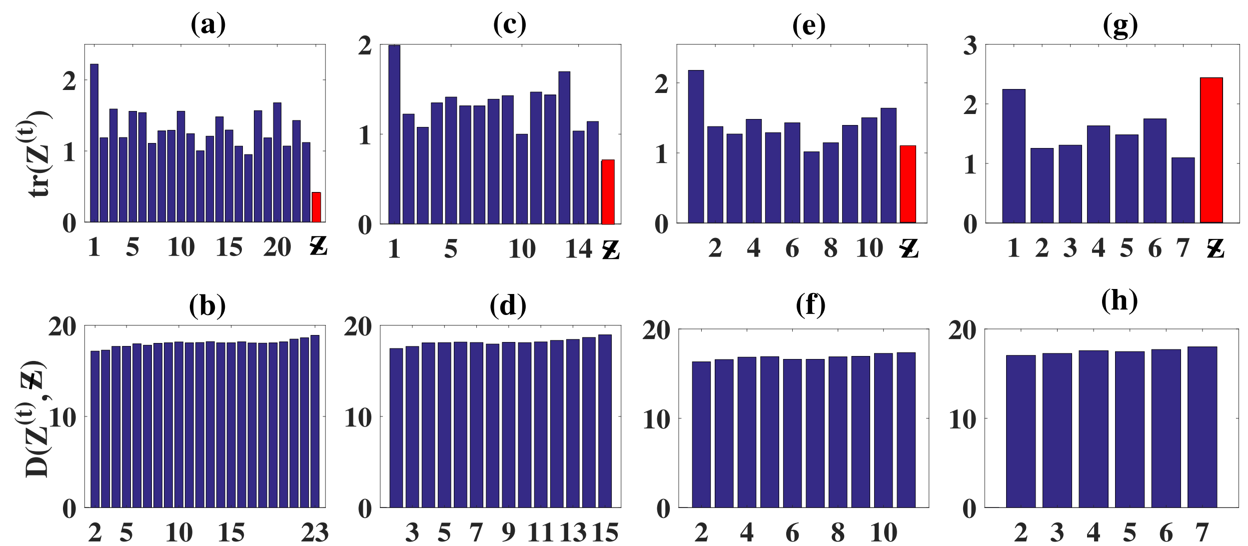


FIGURE 6. The intensity of evolution over time on the MIT e-mail networks with different resolutions. (a, c, e, d) are the trace of $Z^{(t)}$ and $\mathcal{Z}$, and (b, d, f, h) are the square distance between $Z^{(t)}$ and $\mathcal{Z}$.

email networks. It is suitable for the real-world network which has a natural evolution, because FaceNet also considers the evolutions of communities and the temporal smoothness of evolutions.

From the above conclusions, our proposed ONMF-EEP always achieves better performance than the four benchmark methods over both NMI and ER on 8 dynamic networks from the Grow-shrink benchmark and the MIT e-mail dataset. At the same time, the variance rods are stable and small in most cases. This shows that our algorithm has good robustness for temporal community detection task in dynamic networks.

C. EVOLUTION ANALYSIS AND LINK PREDICTION

Evolution analysis of community structure is very important for tracking the time-varying trend of dynamic networks. According to the obtained evolution matrices of IPE and GEP, we demonstrate the intensity of evolution over time on the dynamic Grow-shrink networks and the MIT e-mail networks. Here, the trace of evolution matrix $Z^{(t)}$, i.e., $tr(Z^{(t)})$ ($t = 1, 2, \cdots, T-1$), is used for quantizing the propensity of keeping in current community. In other words, the smaller the trace of $Z^{(t)}$, the more dramatic the evolution. In addition, we compute the distance between the GEP and the LEPs with $D(Z^{(t)}, \mathcal{Z}) = ||Z^{(t)} - \mathcal{Z}||_F^2$. The big distance can indicate the LPE is abnormal, which can be used for evolution anomaly detection of dynamic networks.

Fig.5 shows the intensity of evolution over time on the dynamic Grow-shrink networks with different fuzzy parameters $q = 31, 33, 35, 37$ from left to right. In the figure, the traces of $Z^{(t)}$ and $\mathcal{Z}$ are demonstrated in Fig.5 (a, c, e, g), and the distances between $Z^{(t)}$ and $\mathcal{Z}$ in Fig.5 (b, d, f, h),

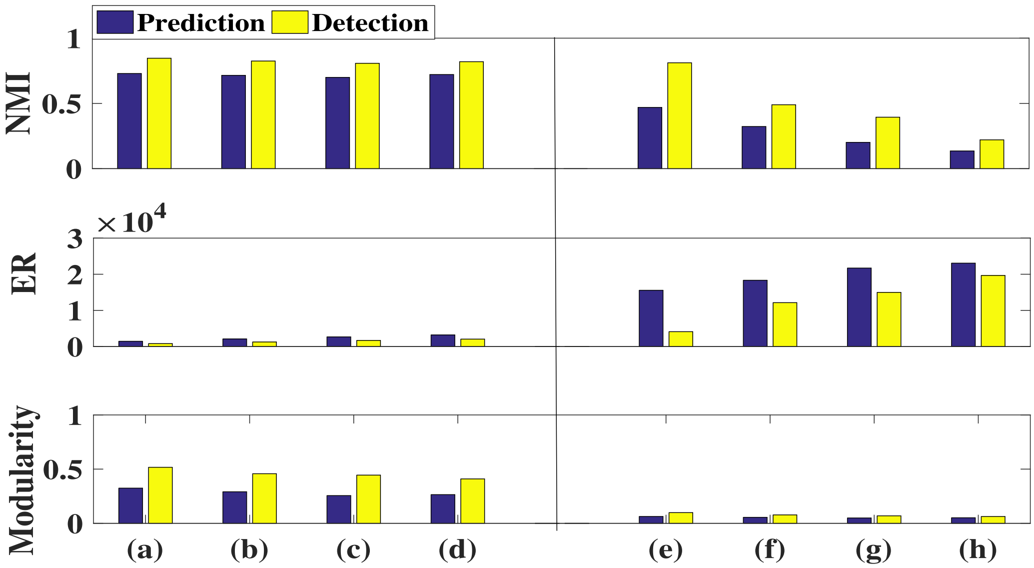


FIGURE 7. The performance of community prediction and detection over NMI, ER and modularity on the MIT e-mail networks and the dynamic Grow-shrink networks.

respectively. The x-axes are snapshot labels t. From the first row of Fig.5, we find the trace of $\mathcal{Z}$ (see as red bars) are lower than the $Z^{(t)}$ on the four networks of Grow-shrink benchmark. This indicates that GEP has a stronger tendency of community transfer than IPEs. From the secongd row of Fig.5, we can discover evolution anomaly at a glance. Because the higher the $D(Z^{(t)}, \mathcal{Z})$ value, the more abnormal the evolution. For example, the IPE at snapshot 4 and 11 are more abnormal on the networks with $q = 31$.

Similarly, we show the intensity of evolution over time on the MIT e-mail networks with different resolutions $T = 24, 16, 12, 8$ from left to right in Fig.6. On the contrary, the $tr(\mathcal{Z})$ (see as red bars) is getting higher and higher from left to right in the first row of Fig.6. This indicates that the resolution of snapshots has a significant influence on the GEP of the MIT e-mail networks, eg., there is a larger propensity of keeping in current community for the networks with longer snapshot windows. It's worth noting that $tr(Z^{(1)})$ is relatively bigger than others on each of the four networks in thefirst row of Fig.6. The reason for this phenomenon is that there is no historical information used for IPE extraction at the first snapshot. From the secongd row of Fig.6, we can't discover any evolution anomalies for the four networks of the MIT e-mail networks with different resolutions.

For the community structure prediction, we compute the NMI, ER and Modularity [26] at the future snapshot with $\hat{H}^{(T)}$, which is obtained according to $\hat{H}^{(T)} = H^{(T-1)}\mathcal{Z}$. We show the performance of community prediction and detection for snapshot T on the MIT e-mail networks ((a),(b),(c), and (d)) and the dynamic Grow-shrink networks ((e),(f),(g), and (h)) in Fig.7. From the comparison of results, we discover the accuracy of community prediction keep usually close to community detection over NMI, ER and Modularity, especially for the MIT e-mail networks ((a),(b),(c), and (d)). The results indicate that our method has effective predictive ability for community structure.

For temporal link prediction task, we compare our proposed ONMF-CCP with four existing methods: weighted collapsed tensor (WCT), Katz index-based algorithm (Katz), singular value decomposition (SVD), symmetric NMF-based algorithm (SNMF-FC) [13], which are widely used for temporal link prediction problem. WCT and Katz are topology based methods, while SVD, SNMF-FC and ONMF-EEP are matrix decomposition based methods. Similarly, all the algorithms are repeated 10 times over the evaluation metrics RMSE, and the average results are demonstrated in bar charts.

We demonstrate the results of link prediction over RMSE on the dynamic Grow-shrink networks and MIT e-mail networks in table 2. The dynamic Grow-shrink networks are generated with different fuzzy parameters: $q = 31, 33, 35, 37$. From the table, we conclude that WCT and SVD are less accurate than the others, and the proposed ONMF-EEP is a little bit higher more accurate than SNMF-FC on the four artificial dynamic networks. The reason is that WCT, Katz and SVD collapse the snapshot networks into a static network so that eliminate topological information. Similarly, the MIT e-mail networks are cut into multi-time slices with different resolutions: 2, 3, 4, 6 months as a snapshot, of which the numbers of snapshots are correspongding to $T = 24$, $T = 16$, $T = 12$, $T = 8$ respectively. Fortunately, the proposed ONMF-EEP is always the most accurate on the four real-world dynamic networks. Undeniably, we find the performance of SNMF-FC almost keeps pace with ours. The reason is that the SNMF-FC, which is similar to the proposed ONMF-EEP, collapses the feature matrices obtained at various time points with an immediate purpose to avoid eliminating topological information. The results in table 2 show that our proposed ONMF-EEP outperforms these benchmark

TABLE 2. The results of link prediction over RMSE on two dynamic networks.

Methods	Grow-shrink networks				MIT e-mail networks			
	$q = 31$	$q = 33$	$q = 35$	$q = 37$	$T = 24$	$T = 16$	$T = 12$	$T = 8$
WCT	18.55	18.77	18.92	19.08	8.44	9.71	10.47	11.23
Katz	12.98	12.78	12.75	13.06	9.87	14.11	8.89	13.95
SVD	18.55	18.77	18.92	19.08	8.44	9.71	10.47	11.23
SNMF-FC	12.34	12.49	12.60	12.69	5.74	6.20	6.41	**6.77**
ONMF-EEP	**11.72**	**11.84**	**11.93**	**12.00**	**5.13**	**5.75**	**6.20**	6.79

methods on the dynamic Grow-shrink networks and the MIT e-mail networks.

In summary, our proposed method ONMF-EEP has good performance for temporal community detection, and has effective ability for community and link structure prediction both on real-world network and artificial network.

VI. CONCLUSIONS

Extracting evolution pattern of temporal community structures helps to understand the time-varying and complicated dynamic networks. We first extract evolution pattern by modelling and uncovering the evolution characteristics of dynamic networks from local and global perspectives. Under this assumption, we propose a novel framework ONMF-EEP for community detection, evolution analysis, and community and link structure prediction. Fortunately, our proposed framework has excellent performance for temporal community detection both on real-world networks and artificial networks. Simultaneously, our approach has an effective ability for community and link structure prediction in dynamic networks.

There are still several problems with our proposed approach to be considered. Firstly, our method is a two-step strategy, and this is a question worth thinking about how to skillfully design our framework into a unified model for extracting LEP and GEP. Secondly, the balance parameters α and β are set according to experiments, and it is necessary to study a strategy for determining the balance parameters based on dynamic networks automatically. Thirdly, we are interested in exploring the correlation between LEP (or GEP) and topological features for dynamic networks.

REFERENCES

[1] C. Aggarwal and K. Subbian, "Evolutionary network analysis: A survey," *Comput. Surv.*, vol. 47, no. 1, p. 10, Jul. 2014.

[2] G. Rossetti, and R. Cazabet, "Community discovery in dynamic networks: A survey," *Comput. Surv.*, vol. 51, no. 2, p. 35, Jun. 2018.

[3] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Phys. Rep.*, vol. 659, pp. 1–44, Nov. 2016.

[4] D. Liu, D. Duan, S. Sui, and G. Song, "Effective semisupervised community detection using negative information," *Math. Problems Eng.*, vol. 2015, Jun. 2015, Art. no. 109671.

[5] L. Fan, S. Xu, D. Liu, and Y. Ru, "Semi-supervised community detection based on distance dynamics," *IEEE Access*, vol. 6, pp. 37261–37271, 2018.

[6] P. Lee, L. V. S. Lakshmanan, and E. E. Milios, "Incremental cluster evolution tracking from highly dynamic network data," in *Proc. IEEE 30th Int. Conf. Data Eng.*, Apr. 2014, pp. 3–14.

[7] D. Chakrabarti, R. Kumar, and A. Tomkins, "Evolutionary clustering," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2006, pp. 554–560.

[8] P. Jiao, H. Lyu, X. Li, W. Yu, and W. Wang, "Temporal community detection based on symmetric nonnegative matrix factorization," *Int. J. Mod. Phys. B,*, vol. 31, no. 13, May 2017, Art. no. 1750102.

[9] R. Michalski, P. Kazienko, and D. Krol. "Predicting social network measures using machine learning approach," in *Proc. Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2012, pp. 1056–1059.

[10] W. Yu, C. C. Aggarwal, and W. Wang, "Temporally factorized network modeling for evolutionary network analysis," in *Proc. 10th ACM Int. Conf. Web Search Data Mining*, Feb. 2017, pp. 455–464.

[11] U. Sharan and J. Neville, "Temporal-relational classifiers for prediction in evolving domains," in *Proc. 8th IEEE Int. Conf. Data Mining*, Dec. 2008, pp. 540–549.

[12] C. A. Bliss, M. R. Frank, C. M. Danforth, and P. S. Dodds, "An evolutionary algorithm approach to link prediction in dynamic social networks," *J. Comput. Sci.*, vol. 5, no. 5, pp. 750–764, 2014.

[13] X. Ma, P. Sun, and G. Qin, "Nonnegative matrix factorization algorithms for link prediction in temporal networks using graph communicability," *Pattern Recognit.*, vol. 71, pp. 361–374, Nov. 2017.

[14] W. Yu, W. Cheng, C. C. Aggarwal, H. Chen, and W. Wang, "Link prediction with spatial and temporal consistency in dynamic networks," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 3343–3349.

[15] T. Wu, C.-S. Chang, and W. Liao, "Tracking network evolution and their applications in structural network analysis," *IEEE Trans. Netw. Sci. Eng.*, to be published.

[16] V. Gligorijevic, Y. Panagakis, and S. P. Zafeiriou, "Non-negative matrix factorizations for multiplex network analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 4, pp. 928–940, Apr. 2019.

[17] T. Hartmann, A. Kappes, and D. Wagner, "Clustering evolving networks," in *Algorithm Engineering* (Lecture Notes in Computer Science), vol. 9220, L. Kliemann and P. Sanders, Eds. Cham, Switzerland: Springer, 2016.

[18] C. Pizzuti, "Evolutionary computation for community detection in networks: A review," *IEEE Trans. Evol. Comput.*, vol. 22, no. 3, pp. 464–483, Jun. 2018.

[19] N. Dakiche, F. B.-S. Tayeb, Y. Slimani, and K. Benatchba, "Tracking community evolution in social networks: A survey," *Inf. Process. Manage.*, vol. 56, no. 3, pp. 1084–1102, May 2019.

[20] M. Turcotte, "Anomaly detection in dynamic networks," Los Alamos Nat. Lab., Los Alamos, NM, USA, Tech. Rep. LA-UR-14-28026, 2014.

[21] S. Ranshous, S. Shen, D. Koutra, S. Harenberg, C. Faloutsos, and N. F. Samatova, "Anomaly detection in dynamic networks: A survey," *J. Comput. Statist.*, vol. 7, no. 3, pp. 223–247, 2015.

[22] G. Palla, A.-L. Barabási, and T. Vicsek, "Quantifying social group evolution," *Nature*, vol. 446, nos. 71–36, pp. 664–667, Apr. 2007.

[23] S. Asur, S. Parthasarathy, and D. Ucar, "An event-based framework for characterizing the evolutionary behavior of interaction graphs," *ACM Trans. Knowl. Discovery from Data*, vol. 3, no. 4, p. 16, Nov. 2009.

[24] N. Ilhan and S. G. Ögüdücü, "Feature identification for predicting community evolution in dynamic social networks," *Eng. Appl. Artif. Intell.*, vol. 55, pp. 202–218, Oct. 2016.

[25] W. Liu, S. Saganowski, P. Kazienko, and S. A. Cheong, "Using machine learning to predict the evolution of physics research," Oct. 2018, *arXiv:1810.12116*. [Online]. Available: https://arxiv.org/abs/1810.12116

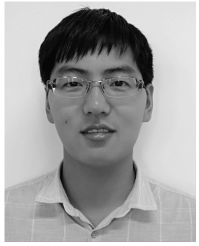
[26] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, "Community structure in time-dependent, multiscale, and multiplex networks," *Science*, vol. 328, nos. 59–80, pp. 876–878, May 2010.

[27] E. G. Tajeuna, M. Bouguessa, and S. Wang, "Modeling and predicting community structure changes in time evolving social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 6, pp. 1166–1180, Jun. 2018.

[28] Y. Panagakis, C. Kotropoulos, and G. R. Arce, "Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 3, pp. 576–588, Mar. 2010.
[29] P. Jiao, W. Yu, W. Wang, X. Li, and Y. Sun, "Exploring temporal community structure and constant evolutionary pattern hiding in dynamic networks," *Neurocomputing*, vol. 314, no. 7, pp. 224–233, Nov. 2018.
[30] 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, Jan. 2018.
[31] 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 *Proc. 17th Int. Conf. World Wide Web*, Apr. 2008, pp. 685–694.
[32] F. Folino and C. Pizzuti, "An evolutionary multiobjective approach for community discovery in dynamic networks," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1838–1852, Aug. 2014.
[33] C. Granell, R. K. Darst, A. Arenas, S. Fortunato, and S. Gómez, "Benchmark model to assess community structure in evolving networks," *Phys. Rev. E*, vol. 92, no. 1, Jul. 2015, Art. no. 012805.
[34] K. Nowicki and T. A. B. Snijders, "Estimation and prediction for stochastic blockstructures," *J. Amer. Stat. Assoc.*, vol. 96, pp. 1077–1087, Sep. 2001.
[35] (2011). *Karlsruhe Institue of Technology Dynamic Network of Email Communication at Department of Informatics at Karlsruhe Institue of Technology (kit)*. [Online]. Available: http://i11www.iti.uni-karlsruhe.de/en/projects/spp1307/emaildata

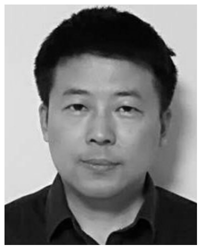
WEI YU received the B.S. degree from Jinggangshan University, Ji'an, China, in 2012, and the M.S. degree from Henan Normal University, Xinxiang, China, in 2015. He is currently pursuing the Ph.D. degree with the School of College of Intelligence and Computing, Tianjin University, Tianjin, China. His research interests include dynamic complex network analysis, large-scale data mining, and machine learning.

WENJUN WANG received the Ph.D. degree from Peking University, Beijing, China, in 2004. He is currently a Professor with the School of College of Intelligence and Computing, Tianjin University, Tianjin, China. His research interests include computational social science, emergency management, large-scale data mining, and network science.



PENGFEI JIAO received the B.S. degree from Hainan University, Haikou, China, in 2012, and Ph.D. degrees in computer science from Tianjin University, Tianjin, China, in 2018, where he is currently a Lecture with the Center of Biosafety Research and Strategy. He has published nearly 20 international journal articles and conference papers. His current research interests include data mining, complex network analysis, and machine learning.

HUAMING WU received the B.E. and M.S. degrees in electrical engineering from the Harbin Institute of Technology, China, in 2009 and 2011, respectively, and the Ph.D. degree (Hons.) in computer science from Freie Universität Berlin, Germany, in 2015. He is currently an Associate Professor with the Center for Applied Mathematics, Tianjin University. His research interests include model-based evaluation, wireless and mobile network systems, mobile cloud computing, and deep learning.



YUEHENG SUN received the Ph.D. degree from Tianjin University, Tianjin, China, in 2005, where he is currently a Lecturer with the School of Computer Science and Technology. His research interests include social computing, data mining, and machine learning.

MINGHU TANG received the M.S. degree from Yunnan University, Kunming, China, in 2008. He is currently pursuing the Ph.D. degree with the School of Computer Science and Technology, Tianjin University, Tianjin, China. He is currently an Associate Professor with the School of Computer Science and Technology, Qinghai Nationalities University, Xining, China. His current research interests include complex network analysis, data mining, and machine learning.