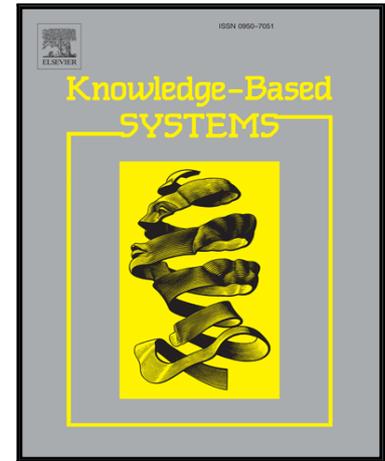


Accepted Manuscript

Autonomous Overlapping Community Detection in Temporal Networks: A Dynamic Bayesian Nonnegative Matrix Factorization Approach

Wenjun Wang, Pengfei Jiao, Dongxiao He, Di Jin, Lin Pan, Bogdan Gabrys

PII: S0950-7051(16)30235-0
DOI: [10.1016/j.knosys.2016.07.021](https://doi.org/10.1016/j.knosys.2016.07.021)
Reference: KNOSYS 3609



To appear in: *Knowledge-Based Systems*

Received date: 10 December 2015
Revised date: 11 July 2016
Accepted date: 15 July 2016

Please cite this article as: Wenjun Wang, Pengfei Jiao, Dongxiao He, Di Jin, Lin Pan, Bogdan Gabrys, Autonomous Overlapping Community Detection in Temporal Networks: A Dynamic Bayesian Nonnegative Matrix Factorization Approach, *Knowledge-Based Systems* (2016), doi: [10.1016/j.knosys.2016.07.021](https://doi.org/10.1016/j.knosys.2016.07.021)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Autonomous Overlapping Community Detection in Temporal Networks: A Dynamic Bayesian Nonnegative Matrix Factorization Approach

Wenjun Wang^a, Pengfei Jiao^a, Dongxiao He^{*,a}, Di Jin^a, Lin Pan^a, Bogdan Gabrys^b

^a*School of Computer Science and Technology, Tianjin University, Tianjin 300072, China*

^b*Data Science Institute, Bournemouth University, Poole House, Talbot Campus, Fern Barrow Poole, Dorset, BH12 5BB, UK*

Abstract

A wide variety of natural or artificial systems can be modeled as time-varying or temporal networks. To understand the structural and functional properties of these time-varying networked systems, it is desirable to detect and analyze the evolving community structure. In temporal networks, the identified communities should reflect the current snapshot network, and at the same time be similar to the communities identified in history or say the previous snapshot networks. Most of the existing approaches assume that the number of communities is known or can be obtained by some heuristic methods. This is unsuitable and complicated for most real world networks, especially temporal networks. In this paper, we propose a Bayesian probabilistic model, named Dynamic Bayesian Nonnegative Matrix Factorization (DBNMF), for automatic detection of overlapping communities in temporal networks. Our model can not only give the overlapping community structure based on the probabilistic memberships of nodes in each snapshot network but also automatically determines the number of communities in each snapshot network based on automatic relevance determination. Thereafter, a gradient descent algorithm is proposed to optimize the objective function of our DBNMF model. The experimental results using both synthetic datasets and real-world temporal networks demonstrate that the DBNMF model has superior performance compared with two widely used methods, especially when the number of communities is unknown and when the network is highly sparse.

Key words: community detection, temporal networks, Bayesian nonnegative matrix factorization, gradient descent, model selection.

1. Introduction

Complex networks, such as social networks, biological networks and information networks, are very common in real life. The analysis of complex networks has been becoming more and more

*Corresponding author.

Email address: hedongxiao@tju.edu.cn (Dongxiao He)

important and has drawn great attention in the last few years. As a heated topic in network
5 science, community detection [1] is aimed at identifying meaningful groups or clusters in complex
networks, which have a larger density in intra-groups compared to that in inter-groups. Intuitively,
there is a closer relationship among individuals in intra-communities while a weaker relationship for
inter-communities. Taking the protein-protein interaction networks [2] as an example, each protein
community may correspond to an analogous functional module. Similarly, individuals in the same
10 group may have common interests in social networks [3] such as Facebook, Twitter, LiveJournal and
so on.

One problem is that each individual may have multiple roles, which leads to an overlapping
community structure [4] [5]. For example, one person with different interests may often join multiple
groups in real life. Various methods for (overlapping) community detection have been proposed, such
15 as hierarchical clustering, spectral clustering, modularity-based methods [6], random walks [7], clique
percolation [8], and stochastic block model [9]. However, most of these methods are designed for
static networks, while many networks are temporal or dynamic [10] in real world. For instance,
communication networks or email networks are highly dynamic, and the inter-time distribution of
individual's interaction follows power-law property. Cooperation networks may change with the
20 physical moving of people from one institution to another or the changing of the academic interests
of the authors. Protein functions always evolve with aging. Thus, overlapping community detection
in temporal networks is increasingly important and challenging in practical applications.

So, what are temporal networks and how to model their dynamics? What is the relationship
between community detection in temporal networks and community evolution? How to detect
25 dynamic communities in temporal networks, and what are the difficulties of community detection
in temporal networks? In the following we will discuss and attempt to provide answers to these
questions.

A temporal network can be modeled as a series of static networks [11]. Here as the most published
research did, we assume that there are three types of changes in temporal networks including: 1)
30 new nodes adding or old nodes removing; 2) edges appearing or disappearing; and 3) the change of
the number of communities. Generally, network evolution or community evolution has been studied
for several years [12], and there are several types of dynamics, such as expansion and shrinking.
However, we argue that there have been some essential differences between community evolution
and community detection in temporal networks. The former mainly focuses on the temporal char-
35 acteristics of communities and their future trends with the network evolution [13]; while the later
is aimed at identifying dynamic communities with considering their evolution. In other words, a
part of community evolution can be considered as some constraints or essential conditions when
identifying communities in temporal networks.

Recently, several methods have been proposed for community detection in temporal networks,

40 which can be classified into three main types. The first type of approaches is the well-known two-steps strategy, which identifies community structure in each snapshot network by using methods for static networks, and then analyzes the community evolution according to some principles, such as the Jaccard index [13]. The second type of approaches is the generative model [14] which is based on the combination of stochastic block model (SBM) and the state space model or liner dynamic system, i.e., the whole temporal network is represented as a sample of a dynamic generative model. 45 Then the detection of communities and the analysis of community evolution are transformed into an unified problem of parameters estimation based on likelihood maximization. The last type of approaches is the so called evolutionary clustering [15] which considers the clustering results of the previous snapshot networks when analyzing the current snapshot network.

50 In general, there are at least three factors leading to the difficulties of community detection in temporal networks. The first is that one may encounter different types of dynamics in temporal networks. With the evolution of networks, the original links or nodes in the networks may disappear and new ones may occur, thus the communities in temporal networks will be created or disappear, and can also merge or split [16]. In this sense, community detection in temporal networks is no longer a static model or function to be computed or optimized, but needs to be analyzed as a network stream or a temporal process. The second is the detection of overlapping communities which has been extensively discussed in the case of static networks [4], but is much less well covered and more challenging in the case of dynamic overlapping community detection. The last factor is the model selection problem, i.e., the determination of the number of communities in temporal networks. In fact, this is a common problem to be solved in all community detection approaches, especially for large-scale temporal networks [17]. Therefore, one needs an effective model or method that can not only detect overlapping communities but also determine the number of communities in temporal networks. 60

As discussed above, there are mainly two significant and difficult problems for most current community detection methods in temporal networks. One is the detection of overlapping community structure, and the other is the automatic determination of the number of communities in each snapshot of a temporal network. 65

In this paper, we propose a dynamic Bayesian probability model, namely Dynamic Bayesian Non-negative Matrix Factorization (DBNMF), which belongs to the evolutionary clustering approaches with a probability interpretation. The merit of this new model are twofold: it can identify the dynamic overlapping communities and also automatically determine the number of communities in temporal networks, both of which are often ignored by most other methods for dynamic community detection. To be specific, our obtained overlapping community structure has a theoretical interpretation based on nonnegative matrix factorization. The number of communities in the temporal network is automatically detected based on automatic relevance determination [18], which 75

is closely-related with sparse Bayesian learning framework and can be effectively learned in most situations.

In the DBNMF model, the community structure at snapshot t is influenced by the community structure of the snapshot $t-1$ and independent of the previous snapshot networks. For each snapshot network, we propose to tie the columns of the membership matrix through scale parameters that are drawn from Half-Normal distribution, the smaller the parameter the less weight of the column. After finishing all the calculations of the temporal network, we remove the columns whose weights are close to zero in the membership matrix [19]. Then, the overlapping community results and the number of communities are derived simultaneously. Besides, the proposed model can be applied to large and sparse networks because of its stability and effectiveness, which is partly validated in the experiments.

The contributions of this work are twofold:

- We give a well theoretically interpretable model namely DBNMF (Dynamic Bayesian Nonnegative Matrix Factorization) to detect overlapping community structure from temporal networks, which is optimized by a gradient descent algorithm.
- Using the automatic relevance determination, in which we assume all the scale parameters of all columns for every snapshot network are independent and identically distributed, the proposed model can automatically detect the number of communities in temporal networks. This is especially important to deal with large and unexplored real networks in temporal situations.

The remainder of this paper is organized as follows. A literature review about community detection in temporal networks is presented in Section 2. The notation and model is described in Section 3. Section 4 gives the gradient descent algorithm and its computational complexity analysis. Section 5 offers the comprehensive experiments and detailed analysis on both synthetic and some real-world networks. We conclude this work and discuss some existing problems in Section 6.

2. Related Work

In this section, we provide a general overview of the community detection methods in temporal networks.

Most methods for community detection in temporal networks can be divided into three categories. In the first type of approaches, the community structure and its evolution are analyzed in two separated stages. For example, GRAPHSCOPE [17] is the most often used method for community detection in bipartite networks. It encodes every snapshot network based on Minimum Description Length (MDL), and the snapshot networks with similar descriptions will be grouped together into a time segment network. If a new snapshot network cannot fit well into the old segment network, the GRAPHSCOPE gives a change point. This model is optimized by a greedy optimization algorithm.

110 Modularity optimization in [16] represents a series methods. They often first cluster the snapshot network at $t = 1$ based on static community detection methods. And then, at each snapshot network at $t > 1$, the methods deal with a series of events based on the change of the succession of two snapshot networks, that allots which community nodes should join in by computing the maximum probability or generates a new community based on modularity gain. These approaches
115 usually ignore the history information and are sensitive to the noise.

In the second type of approaches, the temporal networks can be regarded as a sample of a dynamic generative model or a liner dynamic network model, and the detection of communities can be taken as a parameters estimation problem based on the maximization of a posteriori probability. The basic idea of this type of models is that the temporal network can be generated conditional on a dynamic
120 mechanism [20]. In general, the results of the community structure and its evolution can be obtained by parameters estimation or the maximization of posterior probability. For example, the dynamic stochastic block model (DSBM) in [14] generates the temporal network based on classic stochastic block model. The DSBM presents the dynamics by adding a block probability transition matrix to model nodes transition in different communities. They also proposed a probabilistic simulated
125 annealing algorithm combined with the Gibbs sampling to estimate all parameters of the model though this model assumes that the number of communities in each snapshot network are the same and known.

The last and the most popular type of approaches is based on evolutionary clustering, the idea of which is to cluster the current snapshot data by adding a history regularization. To be specific,
130 Chakrabarti, et al. [15] first provided the evolutionary clustering framework and used the classic k -means and hierarchical clustering methods to cluster dynamic data by adding a temporal smoothness constraint. Chi, et al. [21] further modified the spectral clustering based on evolutionary clustering to analyze the dynamic data. They proposed two frameworks: one is the preserving cluster quality (PCQ) framework which is based on the similarity between the current similarity matrix and the
135 history similarity matrix, and the other is the preserving cluster membership (PCM) framework which is based on the difference between the current partition and the historic partition. To sum up, the goal of evolutionary clustering is to find a good trade-off between clustering accuracy of the current clustering and the deviation from the history. Although these methods are originally used for data clustering with known number of clusters, they can also be used for community detection
140 in temporal networks based on some quality definition or feature extraction of snapshot networks. Besides, there are also some works on community detection in the temporal networks based on evolutionary clustering. Kim and Han [22] proposed a particle and density method for community detection in temporal networks by defining the nano community. Tang, et al. [23] proposed a modified spectrum method for community detection on weighted temporal networks. Xu, et al.
145 [24] presented a generative model based on Hidden Markov Model (HMM) with a prior Dirichlet

process under the framework of evolutionary clustering, and the model assumed a fixed number of communities for the temporal networks. Lin, et al. [25] proposed the FacetNet method that is based on low-rank matrix recovery techniques with a temporal smoothness. This method is an improvement of SNMF (Symmetric Nonnegative Matrix Factorization) by defining the snapshot cost and the temporal cost with Kullback-Leibler divergence. Francesco and Clara [19] regarded the snapshot cost and the temporal cost as a multi-objective function and proposed a genetic algorithm for optimization. However, most of this type of approaches assume a fixed number of communities in all the snapshot networks and cannot detect the overlapping community structure.

The determination of the number of communities K , which is the so called model selection problem, is often solved by running the algorithm with different k and picking the best one corresponding to the optimal modularity [25]. Besides, Bayesian model selection is a general method for the determination of the number of the clusters in data clustering problem. Specifically, Bayesian Nonnegative Matrix Factorization (BNMF) [26] has also been used to detect community structure with the automatic determination of the number of communities in static networks. To the best of our knowledge, there are a few model selection methods proposed for dynamic community detection.

Similarly to community detection, the analysis and mining over temporal, dynamic, uncertain and stochastic networks have also drew many attentions. Some assumptions and evidences for our proposed model are also derived from the ideas of these works. For instance, Ahmed and Chen [27] proposed an efficient algorithm for link prediction in temporal uncertain social networks, in which each edge is associated with a probability value indicating its existence in the network. Yuan, et al. [28] employ a filtering-and-verification framework for retrieve all qualified matches of a query pattern in the uncertain graph, in which a probabilistic matching tree (PM-tree) is built from match cuts obtained by a cut selection process and based on the PM-tree, and a collective pruning strategy is devised to prune a large number of unqualified matches. Rezvanian and Meybodi [29] first define minimum vertex covering in stochastic graphs and give four learning automata-based algorithms for solving minimum vertex covering problem in stochastic graphs, in which the probability distribution functions of the weights associated with the vertices of the graph are unknown and can be parameterized a proper choice of the parameter. Du, et al [30] investigates the problem of node similarity computation on large uncertain graphs.

As we have discussed, the detection of overlapping community structure and the determination of the number of communities are two key problems for community detection in temporal networks. However, both problems have not been analyzed simultaneously in the existing work. In this paper, we proposed a Dynamic Bayesian Nonnegative Matrix Factorization (DBNMF) model for the detection of overlapping community structure and the automatic determination of the number of communities in temporal networks. A gradient descent algorithm to learn the parameters of the model is also proposed.

3. Notations and Model

3.1. Notations

In this subsection we give the notations of the proposed DBNMF model and its corresponding optimization algorithm. A temporal network is defined as $G = \{V_1, V_2, \dots, V_T\}$, where T is the number of snapshot networks, V_t denotes each snapshot network of the temporal network. If a symbol has one subscript, this subscript represents time snapshot, such as K_t represents the number of communities in the t -th snapshot network. If the symbol has more than one subscript, we will separate them by commas, such as $V_{ij,t}$ denotes the element of the i -th row and the j -th column at the t -th snapshot network.

Without loss of generality, we assume that the networks are undirected and unweighed, so each element of V_t , $t = 1, 2, \dots, T$, represents the interaction between nodes i and j . N_1, N_2, \dots, N_T denote the number of nodes in each snapshot network, respectively. Thus we have $V_{ij,t} \in \{0, 1\}^{N_t \times N_t}$, $i, j = 1, 2, \dots, N_t$, $t = 1, 2, \dots, T$. If $V_{ij,t} = 1$ there is an edge between nodes i and j in the snapshot network t , and 0 otherwise. It is easy to extend the model to analyze the weighed temporal networks although in this work we mainly focus on undirected and unweighed temporal networks.

3.2. Model Formulation

As represented in Figure 1, for each snapshot t , $V_{ij,t}$ represents the interaction between nodes i and j , which is generated by $H_{ik,t}$ and $H_{jk,t}$, $k = 1, 2, \dots, K_t$, where K_t denotes the number of communities of snapshot network t , and $H_{ik,t}$ denotes the fraction of node i belonging to the k -th community at snapshot network t . Here, each $H_{ik,t}$ is drawn from a half normal distribution with parameter β_t which denotes the scale parameter on the latent variable $H_{ik,t}$ in terms of the shrinkage of each community and is parameterized by a Gamma distribution with two hyper-parameters a_t and b_t . Gamma distribution is the conjugate prior of half normal distribution, and each β_t evaluates the parameterized community and shrinks irrelevant communities, and thus we solve the model selection problem in the community detection in the temporal network. Some similar works can be also seen in [26][31][32]. Thereafter we have $\sum_{k=1}^{K_t} H_{ik,t} = 1$. We then take $\sum_{k=1}^{K_t} H_{ik,t} H_{jk,t}$ as the expected number of links between nodes i and j at snapshot t . The $V_{ij,t}$ is drawn from a Poisson distribution with the mean $\sum_{k=1}^{K_t} H_{ik,t} H_{jk,t}$. It means that the probability of an edge and the expected number of edges are equal in the limitation of a large network, which has been also used in [33][34].

In fact, our model for each snapshot t of temporal networks can be also regarded as a hierarchical Bayesian approach to model selection. On one hand, $\beta_{k,t}$ is drawn from a Gamma distribution, which can be regarded as the weight of each column of $H_{ik,t}$ and solves the model selection problem in the community detection in the temporal network. On the other hand, $H_{ik,t}$ is drawn from the half normal distribution, where $\log P(H_t | \beta_t)$ corresponds to the sparse regularization of the objective

function in our model, and β_t is the coefficient. Then we can get the number of communities and the community structure at each snapshots in the temporal networks.

The number of communities or the inter rank at each snapshot network will be decided by the $\beta_{k,t}$, where $t = 1, \dots, T$, $k = 1, \dots, K_t$. Here, based on automatic relevance determination (ARD), we give a Gamma distribution prior with parameters a_t and b_t for β_t that are scaled to every column of H_t . After the calculation of all H_t and β_t , we remove the columns of H_t where the values in $\beta_k < \epsilon$ and we set $\epsilon = 0.01$, and then we get the overlapping community structure and the number of communities together at every snapshot network t . We set $a_t = a$ and $b_t = b$ for all of the snapshot networks for convenience, and the detailed description of β_k can be found in [26] [31]. Besides, we introduce a parameter α to balance the clustering result H_t on the current snapshot network and the previous clustering result H_{t-1} . The detailed analysis of the α will be discussed later.

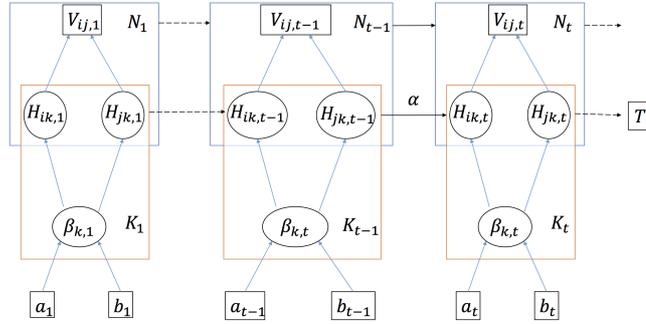


Figure 1: Graph model of our DBNMF. Here $V_{ij,t}$ represent the element at the i -th rows j -th column on the temporal network at snapshot t , and $H_{ik,t}$ denotes the expected probability that node i belongs to community k at snapshot t .

We add the following three paragraphs here, there are also two detailed issues to be resolved in our model. One is how to deal with the varying number of nodes in the temporal network, and the other is how to automatically determine the number of communities in each snapshot.

Considering the first issue, we assume that there are maximal $K_{initial}$ ($K_{initial} \geq \max\{K_t\}, t = 1, \dots, T$) communities in each snapshot of the temporal network. In the following, we explain how we deal with the varying number of nodes in consecutive snapshots in our algorithm. As shown in Figure 2, there are 9 nodes in snapshot $t-1$ and 10 nodes in snapshot t , with 12 unique nodes in total. There are 2 real communities for each of the networks, and we set $K_{initial}$ to be a much larger value ($K_{initial} \gg 2$). In our example, nodes 3 and 8 from snapshot $t-1$ disappear, and nodes 10, 11 and 12 are newly added at time t . After H_{t-1} is calculated, we first delete the rows corresponding to the disappeared nodes 3 and 8 in H_{t-1} , and thus get H'_{t-1} . Then, we add rows which correspond to the newly added nodes 10, 11 and 12 in snapshot t using some small random values, and thus get H''_{t-1} (which has the same size as H_t). And finally, we replace H_{t-1} with H''_{t-1} ,

which ensure that we now can calculate H_t using the result from snapshot $t - 1$ in our model.

Considering the automatic determination of the number of communities in each snapshot, we used the automatic relevance determination method as mentioned above, but still kept the number of columns in each H_t to be $K_{initial}$. After that when we have got all of the $H_t, t = 1, 2, \dots, T$, for each H_t we remove the columns whose summation of values closing to 0. And then, we get the expected number of communities for each snapshot t , which is much smaller than $K_{initial}$ in general.

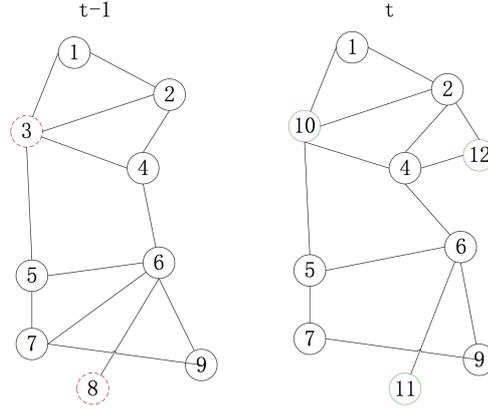


Figure 2: An illustrative example to explain how we deal with varying number of nodes in consecutive snapshots in order to ensure the consistency of our model formulation which use information from snapshots t and $t - 1$ together. There are 2 real communities in each of the networks. Nodes 3 and 8 in snapshot $t - 1$ disappear at time t , and nodes 10, 11 and 12 are newly added.

As we have described above, we first introduce the model for the temporal network G at snapshot $t = 1$, as shown in Figure 1, which is based on the Bayesian Non-negative Matrix Factorization (BNMF) model proposed in [26]. The joint distribution over all variables at snapshot network $t = 1$ with fixed hyper-parameters a_t and b_t can be written as

$$P(V_1, H_1, \beta_1) = P(V_1|H_1)P(H_1|\beta_1)P(\beta_1), \quad (1)$$

and the posterior probability of the model for the snapshot network at $t = 1$ is

$$P(H_1, \beta_1|V_1) = \frac{P(V_1|H_1)P(H_1|\beta_1)P(\beta_1)}{P(V_1)}. \quad (2)$$

It is commonly known that maximizing the posterior probability in (2) is equivalent to minimizing the negative log posterior by ignoring the term $P(V_1)$, which is defined as:

$$\begin{aligned} L_1 &= -\log P(H_1, \beta_1|V_1) \\ &= -\log P(V_1|H_1) - \log P(H_1|\beta_1) - \log P(\beta_1). \end{aligned} \quad (3)$$

Based on $\hat{V}_t \simeq H_t H_t^T$, we assume that $V_{ij,1} \sim \text{Poisson}(\sum_k H_{ik,1} H_{jk,1}^T)$, the generation of all edges are independent, and $\sum_k H_{ik,1} H_{jk,1}^T$ is the Poisson rate. The first term of (3) is the log

likelihood of the network data over all variables at snapshot $t = 1$ of the temporal network. So this term can be rewritten as

$$\begin{aligned} P(V_1|H_1) &= P(V_1|H_1 H_1^T) \\ &= \prod_{i=1}^{N_1} \prod_{k=1}^{K_{initial}} \frac{\hat{V}_{ij,1}^{V_{ij,1}} \exp^{-\hat{V}_{ij,1}}}{V_{ij,1}}. \end{aligned} \quad (4)$$

Based on β_1 , we place independent half normal-priors over the columns of H_1 with parameters $\beta_1 \in R^{K_{initial}} = \{\beta_{1,1} \cdots \beta_{K_{initial},1}\}$, we get the log priors over H_1 as

$$\begin{aligned} P(H_1|\beta_1) &= \prod_{i=1}^{N_1} \prod_{k=1}^{K_{initial}} HN(0, \beta_{k,1}^{-1}) \\ &= \prod_{i=1}^{N_1} \prod_{k=1}^{K_{initial}} \frac{\sqrt{2\beta_{k,1}}}{\sqrt{\pi}} \exp\left(-\frac{H_{ik,1}^2 \beta_{k,1}}{2}\right), \end{aligned} \quad (5)$$

260 where HN represents half-normal distribution.

Here each element of β_1 controls the importance of every column of the membership matrix H_1 based on the observed snapshot network $t = 1$. We assume all $\beta_{k,1}$, $k = 1, 2, \dots, K_{initial}$ are conditional independent and drawn from a Gamma distribution with hyper-parameters a_1 and b_1 . Then we have

$$\begin{aligned} P(\beta_1) &= \prod_{k=1}^{K_{initial}} Ga(\beta_{k,1}|a_1, b_1) \\ &= \prod_{k=1}^{K_{initial}} \frac{b_1^{a_1} \beta_{k,1}^{a_1-1} e^{-\beta_{k,1} b_1}}{\Gamma(a_1)}, \end{aligned} \quad (6)$$

265 where Ga is the standard Gamma distribution. Considering equations (4) (5) (6) and (3), the objective function L_1 can be rewritten as

$$\begin{aligned} L_1 &= -\log P(V_1|H_1) - \log P(H_1|\beta_1) - \log P(\beta_1) \\ &= -\sum_{i=1}^{N_1} \sum_{j=1}^{N_1} (V_{ij,1} \log \frac{V_{ij,1}}{\sum_{k=1}^{K_{initial}} H_{ik,1} H_{jk,1}} + \sum_{k=1}^{K_{initial}} H_{ik,1} H_{jk,1} - V_{ij,1}) \\ &\quad + \sum_{i=1}^{N_1} \sum_{k=1}^{K_{initial}} \left(\frac{1}{2} \beta_{k,1} H_{ik,1}^2\right) - \frac{N_1}{2} \log \beta_{k,1} \\ &\quad + \sum_{k=1}^{K_{initial}} (\beta_{k,1} b_1 - (a_1 - 1) \log \beta_{k,1}) + c, \end{aligned} \quad (7)$$

where c is a constant.

we add a paragraph here we consider the snapshots at time $t > 1$. Here, the number of rows and the number of columns of H_t , respectively, denotes the number of nodes and the number of communities in snapshot t . In general, the number of nodes in snapshot t and that in snapshot $t - 1$ are not equal. As explained in the aforementioned example, we delete the rows from H_{t-1} representing the nodes which disappeared in snapshot t and add the rows newly added nodes at time t , and thus get a new H_{t-1} which has the same size as H_t and is denoted as H''_{t-1} in subsequent

equations. Then we can write the joint distribution over all variables at snapshot t as

$$P(V_t, H_t, H''_{t-1}, \beta_t, \alpha) = P(V_t|H_t)P(H_t|H''_{t-1}, \alpha) \cdot P(H_t|\beta_t)P(\beta_t). \quad (8)$$

275 Similarly to the model in the snapshot network of $t = 1$, based on the previous clustering result H''_{t-1} and the balance parameter α , the posterior probability for the snapshot network at t can be written as

$$P(H_t, \beta_t|V_t, H''_{t-1}, \alpha) = P(V_t|H_t)P(H_t|H''_{t-1}, \alpha) \cdot \frac{P(H_t|\beta_t)P(\beta_t)}{P(V_t, H''_{t-1}, \alpha)}. \quad (9)$$

At the same time, we write the negative log posterior of the posterior probability (ignoring the constant term $P(V_t, H''_{t-1}, \alpha)$) at snapshot network t as

$$\begin{aligned} L_t &= -\log P(H''_{t-1}, \beta_t|V_t, H_t, \alpha) \\ &= -\log P(V_t|H_t) - \log P(H_t|H''_{t-1}, \alpha) \\ &\quad - \log P(H_t|\beta_t) - \log P(\beta_t). \end{aligned} \quad (10)$$

280 Similarly to (4), (5), and (6), we rewrite each term of (10) as

$$\begin{aligned} P(V_t|H_t) &= P(V_t|H_t H_t^T) \\ &= \prod_{i=1}^{N_t} \prod_{j=1}^{N_t} \frac{\hat{V}_{ij,t}^{V_{ij,t}} \exp^{-\hat{V}_{ij,t}}}{V_{ij,t}}, \end{aligned} \quad (11)$$

$$\begin{aligned} P(H_t|\beta_t) &= \prod_{i=1}^{N_t} \prod_{k=1}^{K_{initial}} HN(0, \beta_{k,t}^{-1}) \\ &= \prod_{i=1}^{N_t} \prod_{k=1}^{K_{initial}} \frac{\sqrt{2\beta_{k,t}}}{\sqrt{\pi}} \exp\left(-\frac{H_{ik,t}^2 \beta_{k,t}}{2}\right), \end{aligned} \quad (12)$$

$$\begin{aligned} P(\beta_t) &= \prod_{k=1}^{K_{initial}} Ga(\beta_{k,t}|a_t, b_t) \\ &= \prod_{k=1}^{K_{initial}} \frac{b_t^{a_t} \beta_{k,t}^{a_t-1} e^{-\beta_{k,t} b_t}}{\Gamma(a_t)}, \end{aligned} \quad (13)$$

for the second term of (10), we denote that nodes $V_t - V_{t-1}$ are the newly added in the snapshot t and nodes $V_{t-1} - V_t$ are the disappearing from the snapshot $t - 1$, so we rewrite H_t with adding newly nodes and deleting disappearing nodes based on indexes of nodes in the temporal network, which ensures H_t has the same size with H''_{t-1} and we have

$$\begin{aligned} P(H_t|H''_{t-1}, \alpha) &= \{P(H_t|H''_{t-1})\}^\alpha \\ &= \left\{ \prod_{i=1}^{N_t} \prod_{k=1}^{K_{initial}} \frac{H''_{ik,t-1} \exp^{-H''_{ik,t-1}}}{H_{ik,t}} \right\}^\alpha, \end{aligned} \quad (14)$$

285 thus the objective function L_t can be rewritten as

$$\begin{aligned}
L_t &= -\log P(V_t|H_t) - \log P(H_t|H''_{t-1}, \alpha) \\
&\quad - \log P(H_t|\beta_t) - \log P(\beta_t) \\
&= -\sum_{i=1}^{N_t} \sum_{j=1}^{N_t} (V_{ij,t} \log \frac{V_{ij,t}}{\sum_{k=1}^{K_{initial}} H_{ik,t} H_{jk,t}} + \sum_{k=1}^{K_{initial}} H_{ik,t} H_{jk,t} - V_{ij,t}) \\
&\quad - \alpha \sum_{i=1}^{N_t} \sum_{k=1}^{K_{initial}} (H_{ik,t} \log \frac{H_{ik,t}}{H''_{ik,t-1}} + H''_{ik,t-1} - H_{ik,t}) \\
&\quad + \sum_{i=1}^{N_t} \sum_{k=1}^{K_{initial}} (\frac{1}{2} \beta_{k,t} H_{ik,t}^2) - \frac{N_t}{2} \log \beta_{k,t} \\
&\quad + \sum_{k=1}^{K_{initial}} (\beta_{k,t} b_t - (a_t - 1) \log \beta_{k,t}) + c,
\end{aligned} \tag{15}$$

where c is a constant.

Now, the evaluation of the model parameters becomes an optimization problem, i.e., minimizing the objective function L_t at each snapshot t . To analyze the parameter α , we rewrite (15) as

$$\begin{aligned}
L_t &= -\log P(V_t|H_t) - \log P(H_t|H''_{t-1}, \alpha) \\
&\quad - \log P(H_t|\beta_{k,t}) - \log P(\beta_t) \\
&= -\alpha \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} (V_{ij,t} \log \frac{V_{ij,t}}{\sum_{k=1}^{K_{initial}} H_{ik,t} H_{jk,t}} + \sum_{k=1}^{K_{initial}} H_{ik,t} H_{jk,t} - V_{ij,t}) \\
&\quad - (1 - \alpha) \sum_{i=1}^{N_t} \sum_{k=1}^{K_{initial}} (H_{ik,t} \log \frac{H_{ik,t}}{H''_{ik,t-1}} \\
&\quad + H''_{ik,t-1} - H_{ik,t}) \\
&\quad + \sum_{i=1}^{N_t} \sum_{k=1}^{K_{initial}} (\frac{1}{2} \beta_{k,t} H_{ik,t}^2) - \frac{N_t}{2} \log \beta_{k,t} \\
&\quad + \sum_{k=1}^{K_{initial}} (\beta_{k,t} b_t - (a_t - 1) \log \beta_{k,t}) + c,
\end{aligned} \tag{16}$$

290 with the first term denoting the likelihood of the model in snapshot network t where α represents the weight, and the second term is referred to as a penalty based on the clustering result at snapshot network $t-1$. Other terms could be adjusted by hyper-parameter a_t and b_t , so the objective function in equations (16) and (15) are equal.

4. Optimization algorithm

In this section, we propose a gradient descent algorithm to optimize the objective functions (7) and (16). The update rule of the gradient descent algorithm with suitable step sizes can be turned into the multiplicative update rule [35]. In this way, we used the procedure of iteratively updating

H_t and β_t until convergence. In the following, we will give the gradient of the objective functions with respect to H_t and β_t , and then select the suitable step sizes for H_t and β_t to perform the optimization.

300 The gradient of the objective function with respect to $H_{ik,t}$ at $t = 1$ is

$$\frac{\partial L_1}{\partial H_{ik,1}} = \sum_j (H_{jk,1} - \frac{V_{ij,1}}{\hat{V}_{ij,1}} \cdot H_{jk,1}) + \beta_{k,1} H_{ik,1}, \quad (17)$$

and that for $t > 1$

$$\begin{aligned} \frac{\partial L_t}{\partial H_{ik,t}} &= \alpha [\sum_j (H_{jk,t} - H_{ik,t} \cdot \frac{V_{ij,t}}{\hat{V}_{ij,t}})] + (1 - \alpha) \cdot \\ & (I(N_t, K_{initial}) - \frac{H''_{ik,t-1}}{H_{ik,t}}) + \beta_{k,t} H_{ik,t}. \end{aligned} \quad (18)$$

Similarly, for $\beta_{k,t}$ we have

$$\frac{\partial L_t}{\partial \beta_{k,t}} = \sum_i \frac{1}{2} H_{ik,t}^2 + \sum_j \frac{1}{2} H_{jk,t}^2 - \frac{N_t}{\beta_{k,t}} + b_t - \frac{a_t - 1}{\beta_{k,t}}. \quad (19)$$

We then rewrite equation (17) and (18) in the matrix formulation as

$$\frac{\partial L_1}{\partial H_1} = (\frac{H_1 H_1^T - V_1}{H_1 H_1^T}) \cdot H_1 + H_1 \text{diag}(\beta_1), \quad (20)$$

$$\begin{aligned} \frac{\partial L_t}{\partial H_t} &= \alpha \cdot (\frac{H_t H_t^T - V_t}{H_t H_t^T}) H_t \\ & + (1 - \alpha) \frac{H_t - H''_{t-1}}{H_t} + H_t \text{diag}(\beta_t), \end{aligned} \quad (21)$$

where $\text{diag}(\beta_t)$ is the diagonal matrix with each $\beta_{k,t}$ as the elements.

305 Using gradient descent algorithm, a general update rule for the $H_{ik,t}$ is

$$H_{ik,t} \leftarrow H_{ik,t} + \lambda_t \frac{\partial L_t}{\partial H_{ik,t}}, \quad (22)$$

where λ_t is referred to as the step sizes of update $H_{ik,t}$. According to the analysis in [35], we set

$$\lambda_1 = - \frac{H_{ik,1}}{\sum_j H_{jk,1} + \sum_k \beta_{k,1} H_{ik,1}},$$

and $\lambda_t =$

$$- \frac{H_{ik,t}}{\alpha \sum_j H_{jk,t} + (1 - \alpha) I(N_t, K_{initial}) + \sum_k \beta_{k,t} H_{ik,t}},$$

for $1 < t \leq T$. For $\beta_{k,t}$, we set

$$\beta_{k,t} = \frac{2(N_t - a_t - 1)}{(\sum_i H_{ik,t}^2 + \sum_j H_{jk,t}^2) + 2b_t}.$$

Then we give the optimization algorithm for the DBNMF model as follows.

Algorithm 1 Overlapping community detection in temporal networks by DBNMF

Input: A temporal network $G = V_1, V_2, \dots, V_T$; initial $K_{initial}$, the hyper-parameters a, b , and the balance parameter α

Output: The number of communities $K_t, t = 1, 2, \dots, T$ and the clustering result H_t for each snapshot network

Define: Matrix operation $\frac{W}{M}$ represents element by element division

For $t = 1$ **initialize** a, b and H_1 ;

For $n = 1 : n_{iter}$

1. $H_1 \leftarrow \left(\frac{H_1}{I(N_1, N_1)H_1 + H_1 \text{diag}(\beta_{K,1})} \right) \cdot \left[\left(\frac{V_1}{H_1 H_1^T} \right) H \right]$
2. $\beta_{k,1} \leftarrow \frac{2(N_1 - a_1 - 1)}{(\sum_i H_{ik,1}^2 + \sum_j H_{jk,1}^2) + 2b_1}$

EndFor

EndFor

For $t = 2, \dots, T$

remove $V_{t-1} - V_t$, and add $V_t - V_{t-1}$ of H_{t-1} and get H''_{t-1}

For $n = 1 : n_{iter}$

1. $H_t \leftarrow \left(\frac{(1-\alpha)H''_{t-1} + \alpha[H_t \cdot \left(\frac{V_t}{H_t H_t^T} H_t \right)]}{(1-\alpha)I(N_t, K_{initial})H_t + \alpha H_t + H_t \text{diag}(\beta_{K,t})} \right)$
2. $\beta_{k,t} \leftarrow \frac{2(N_t - a_t - 1)}{(\sum_i H_{ik,t}^2 + \sum_j H_{jk,t}^2) + 2b_t}$

EndFor

ForAll H_1, \dots, H_T

remove $\sum_i H_{ik,t} \leq \epsilon$ $H_{ik,t} = \frac{H_{ik,t}}{\sum_k H_{ik,t}}$

EndFor

Notice that, the operation $V_{t-1} - V_t$ represent the nodes occurring at snapshot network $t - 1$ and disappearing at snapshot network t . Similarly, the operation $V_t - V_{t-1}$ represents the newly added nodes.

310 In the following, we give the complexity analysis of the proposed algorithm. The most time-consuming part is the updating of H_t . At snapshot $t = 1$, the time cost is $O(2p_1(N_1^2 K_{initial} + N_1 K_{initial}))$, where p_1 denotes the number of iterations. For the snapshot networks $t > 1$, the time cost is $O(\sum_{t=2}^T \sum_t p_t (6N_t K_{initial}))$. Then the time complexity of the whole algorithm is $O((Tp)(N^2 K_{initial} + NK_{initial}))$, where p and N represent the average number of the number of
 315 iterations at snapshot networks t and the average number of nodes of all the snapshot networks, respectively. In fact, the real temporal networks are very sparse. When we denote the edges of the temporal networks as $e_t, t = 1, 2, \dots, T$, so the time complexity will degrade to $O((Tp)(eK_{initial}))$, where e denotes the average edges for all snapshots of the temporal network.

5. Experiments

To evaluate the performance of the proposed model and method, we conduct extensive experiments both on the dynamic Girven-Newman synthetic network [25] and on some widely-used real-world networks. In this section, we first introduce the evaluation metrics we used. Second, we give the performance on the dynamic Girven-Newman synthetic network and some real-world temporal networks, which shows the superior performance of our method compared with the state-of-the-art methods. By using experiment as analysis, we demonstrate that the performance of our algorithm is almost not sensitive to the hyper-parameters a and b , and also $\alpha = 0.9$ is a good choice. So we empirically set $a = 8$, $b = 5$ and $\alpha = 0.9$ in all of our tests. Later, we will give the detailed analysis of the parameters in subsection 5.4.

5.1. Evaluation Metrics

In this subsection, we introduce the evaluation metrics used in the paper, including the Normalized Mutual Information (NMI) [36], the error rate (CA) [37], the modularity [38], as well as the fuzzy modularity [39]. The NMI and error rate are used when the ground truth of the community structure of the temporal networks are available; otherwise, the modularity is used. We also evaluate the performance of our method for the detection of overlapping community structure of temporal networks based on the fuzzy modularity. Notice that, these metrics are all widely-used for the static networks. Here we employ them for each time snapshot of the temporal networks. So the descriptions are just based on each snapshot of the temporal networks for convenience.

The Normalized Mutual Information is defined as

$$NMI = \frac{\sum_{i=1}^{K^r} \sum_{j=1}^{K^s} n_{ij} \log\left(\frac{n \cdot n_{i,j}}{n_i^r \cdot n_j^s}\right)}{\sqrt{(\sum_{i=1}^{K^r} n_i^r \log \frac{n_i^r}{n})(\sum_{j=1}^{K^s} n_j^s \log \frac{n_j^s}{n})}}, \quad (23)$$

where n is the number of nodes in the network; K^r and K^s the number of communities of the ground-truth structure and that of the the structure obtained by our method, respectively; n_i^r , n_j^s and n_{ij} the number of nodes in the i -th community of the ground truth, the number of nodes in the j -th community obtained by our method, and the number of common nodes in i -th and j -th communities, respectively. The NMI values vary from 0 to 1. If the value is closer to 1, the obtained result is closer to the ground truth.

The error rate is defined as

$$CA = \|ZZ^T - RR^T\|_F^2, \quad (24)$$

where $Z \in R^{N \times K}$ is the community membership matrix of nodes corresponding to the ground truth. We have $Z_{ik} = 1$ if the node i belongs to the k -th community, and $Z_{ik} = 0$ otherwise. R presents the community membership matrix obtained by our method which has the same definition as Z .

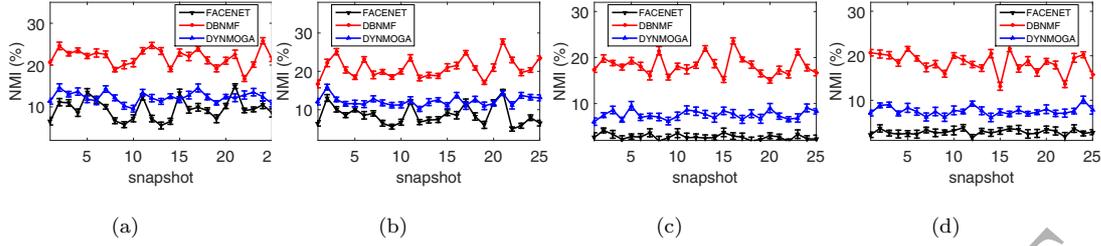


Figure 3: The NMI accuracy on the Girven-Newman Synthetic Networks: the network size is 128, the average degree is 16. (a): $z = 5$, $nC = 3$; (b): $z = 5$, $nC = 9$; (c): $z = 6$, $nC = 3$; (d): $z = 6$, $nC = 9$. The black lines represent the results of the FaceNet method, the blue lines represent the results of the DYNMOGA method, and the red lines represent that of the DBNMF proposed in this paper. Error bars show the standard deviations estimated on 10 networks with the same setting of the parameters.

The value of the error rate is usually increasing with the number of nodes in the network. For a given network, a large value of CA means a relatively poor result.

The modularity is defined as

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{d_i d_j}{2m}) \delta(C_i, C_j), \quad (25)$$

where m is total number of edges in the network, A_{ij} the element of adjacent matrix of the network, d_i the degree of node i , and C_i the community which node i belongs to. $\delta(C_i, C_j) = 1$ if nodes i and j belong to a same community and $\delta(C_i, C_j) = 0$ otherwise. In general, a larger Q value corresponds to a better result.

The fuzzy modularity is defined as

$$FQ = \frac{1}{2m} \sum_c \sum_{i,j \in c} (A_{ij} - \frac{k_i k_j}{2m}) a_{ic} a_{jc}, \quad (26)$$

where a_{ic} and a_{jc} are the probabilities of nodes i and j belonging to community c . A_{ij} , k_i and m have the same definitions as that defined for modularity in (25). The FQ has the ability of evaluating the goodness of the result of fuzzy (overlapping) community structure obtained by a method, which has the similar properties with modularity Q .

Besides, the existing metrics for dynamic community structure in temporal networks, such as the adjusted Rand index (ARI)[40], are not suitable for our model. This is because they often have an inherent assumption, i.e., the number of nodes and the number of communities are both constants in temporal network. Here we give a relative reconstruction error $R - error$, and use it to analyze the temporal evolution of dynamic networks and communities, which is defined as

$$R - error = |1 - \frac{|V_t - R_{t+1} * R_{t+1}^T|_F^2}{|V_t - Z_{t+1} * Z_{t+1}^T|_F^2}|, \quad (27)$$

where V_t is the adjacency matrix of the temporal network at t . The Z_t and R_t denotes, respectively, the community membership matrix of nodes corresponding to the ground truth and the community

membership matrix obtained by the method at the snapshot t . The R -error reflects the temporal evolution when we reconstruct the snapshot t of the network by the community membership matrix at $t + 1$. A similar value between the $|V_t - R_{t+1} * R_{t+1}^T|_F^2$ and $|V_t - Z_{t+1} * Z_{t+1}^T|_F^2$ means that the evolution pattern of the community structure obtained by the model is also similar to the ground truth of the temporal network. So, a smaller R -error value, a better community structure of the temporal network. Here the designed metric R -error still do not support the variety of the number of the nodes, but it is suitable for the situation that the number of communities changes. Thus we can use it to evaluate the performance of difference methods on the generated datasets, although this metric is still not suitable for the real temporal networks. But in the future we try to improve the R -error and make it more powerful.

5.2. Dynamic Girven-Newman Synthetic Networks

The dynamic Girven-Newman synthetic benchmark is proposed by [25], which contains 128 nodes and 4 communities with each community including 32 nodes at each snapshot network. The benchmark is generated as follows. At the first snapshot network, each link is independently generated among all the node pairs with a link probability p_{in} if the pair of nodes are in the same community, and p_{out} otherwise, which is controlled by a mixing parameter or say noise level z . At the snapshot network $t > 1$, there are nc nodes in every community leaving their original communities and joining the other communities, which means that there are $4nc$ nodes changing their original communities. In fact, the value of nc represent the dynamic level in the temporal network. A detailed description of this temporal network can be found in [37]. Note that, there are no changes for the number of the nodes and the number of the communities in this temporal network. Here we set the number of the snapshot networks $T = 25$, the number of nodes in the network is $bS = 128$, the average degree of every node is $aD = 16$, the mixing parameter is $z = 5$ or $z = 6$, and the dynamic level is $nc = 3$ or $nc = 9$. Therefore, there are four situations that generated this temporal network in our experiments. For each generated network, we compare the performance of our algorithm with the results of two competing methods FaceNet [25] and DYNMOGA [19]. Error bars show the standard deviations estimated on 10 networks with the same setting of the parameters.

As we can see from Figures 3, 4 and 5, in terms of the error rate (CA) which expresses the fitness between the model and the observed temporal network, the performance of our model DBNMF is much better than that of FaceNet and DYNMOGA. In terms of the NMI index which measures the accuracy of the community detection result, our DBNMF also outperforms the other two methods. In terms of the R-error which measures the temporal evolution of the community structure, the DBNMF has a smaller value, which means that our method still has a better result. Besides, we have compared these methods on the networks with different values of the mixing parameter z in a larger range from 1 to 6. We do not give the corresponding performance of the community detection results here, because all three methods have good performance when the mixing parameter $z \leq 4$ of

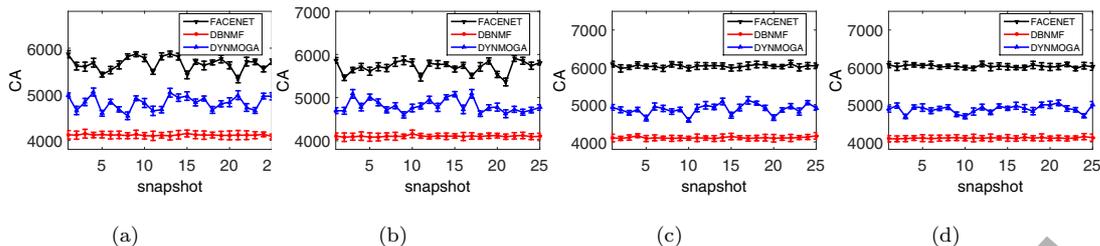


Figure 4: The error rate accuracy on the Girven-Newman Synthetic Networks: the network size is 128, the average degree is 16. (a): $z = 5$, $nC = 3$; (b): $z = 5$, $nC = 9$; (c): $z = 6$, $nC = 3$; (d): $z = 6$, $nC = 9$. The black lines represent the results of the FaceNet method, the blue lines represent the results of the DYNMOGA method, the red lines represent that of DBNMF proposed in this paper. Error bars show the standard deviations estimated on 10 networks with the same setting of the parameters.

the temporal network. But in this situation, our method still has a better performance based both
 405 on the NMI and error rate.

Considering the model selection problem, there are four actual communities in each snapshot
 of the dynamic Girven-Newman synthetic networks, and our method can always get the correct
 number of communities in all of the experiments. Only when the community structure is highly
 unclear under the mixing parameter $z = 6$, it gives five or six communities for several snapshot
 410 networks. This further validates the effectiveness of our new method.

5.3. Real World Data

In this subsection, we show the experiments on some real world temporal networks, including
 four temporal networks with known ground truth of the community structure, which are evaluated
 based on the NMI and the error rate. Here we have only compared the performance of our method
 415 with that of the FaceNet on these temporal networks. This is because both methods have the ability
 to deal with adding new nodes and removing existing nodes, while the DYNMOGA method does
 not have this ability. Please note that for DBLP citation network which does not have the ground
 truth of community structure, the other methods are not suitable, so we have only used our method
 on this network and evaluate its results based on modularity Q and fuzzy modularity FQ and also
 420 give a case study analysis.

5.3.1. KIT-email data

This email data provides a temporal network which is constructed by the email senders, recipients
 and their interactions over time. Here, every email sender or recipient represents a node, an email
 from a sender to a recipient represents an edge. We assume that this network is undirected and
 425 unweighed, the ground truth of community structure in this temporal network is constructed by the
 same way that the students (the senders or the recipients) who are guided by the same supervisor

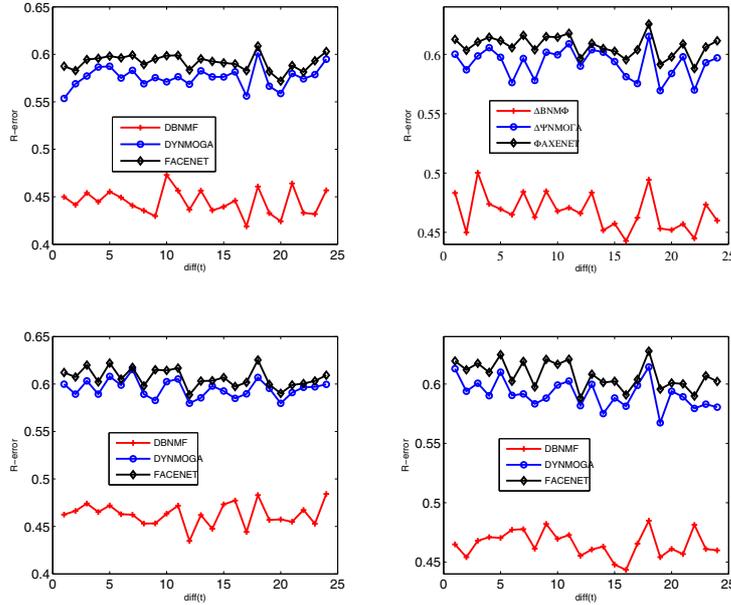


Figure 5: The reconstruct error on the Girven-Newman Synthetic Networks: the number of nodes is 128, and the average degree of the nodes is 16. left-top: $z = 5$, $nC = 3$; right-top: $z = 5$, $nC = 9$; left-down: $z = 6$, $nC = 3$; right-down: $z = 6$, $nC = 9$. The black lines represent the results of the FaceNet method, and the blue lines represent the results of the DYNMOGA method.

belong to the same community. More detailed description can be found in ¹. Here we select the data with the time stamp ranging from July 2007 to December 2009 with 1,218 different students. Thus we have 1,218 nodes and 38 clusters in the data. The time duration of each snapshot network is divided into three different situations for community detection. We set the number of snapshots in the temporal network as $T = 10$, $T = 15$, and $T = 30$ corresponding to the time duration of three months, two months, and one month, respectively. Thus, at each snapshot network, the number of nodes and the number of clusters are smaller than that in the whole dataset. Let t denote the snapshot and N , E , NA , ND and K represent the number of nodes, the number of edges, the number of added nodes based on last snapshot network, the number of deleted nodes based on the last snapshot network and the number of communities, respectively. At each snapshot network t , $NJ = \frac{N_{t-1} \cap N_t}{N_{t-1} \cup N_t}$ represents the evolution of the temporal network (see Table 1 for details). Note that, the data in Table 1 is based on the temporal network with the number of snapshots $T = 15$.

We evaluate the performance of different methods in terms of the NMI and the error rate CA . Here, we compare the result of our proposed method with that of the FaceNet method. The results

¹<http://i11www.itl.uni-karlsruhe.de/en/projects/spp1307/emaildata>

Table 1: Statistical properties of the KIT-email-network with $T = 15$

t	N	E	NA	ND	NJ	K
1	467	2511	no	no	no	36
2	484	2679	66	49	0.78	35
3	545	2641	111	50	0.73	35
4	508	2740	68	105	0.72	35
5	521	2727	77	64	0.76	34
6	538	2810	84	67	0.75	35
7	522	2741	67	83	0.75	35
8	560	3083	99	61	0.74	34
9	529	2942	67	98	0.74	35
10	535	2905	78	72	0.75	34
11	578	3074	97	54	0.76	34
12	563	2948	75	90	0.75	36
13	598	3057	100	65	0.75	35
14	607	3260	97	88	0.73	35
15	573	2631	68	102	0.75	35

Table 2: Statistical properties of the NEC blog network

t	N	E	NA	ND	NJ	K
1	288	688	no	no	no	2
2	289	725	49	48	0.71	2
3	272	564	38	55	0.72	2
4	256	539	45	61	0.67	2
5	241	499	42	57	0.67	2
6	215	415	36	62	0.65	2
7	205	348	49	59	0.59	2
8	187	280	34	52	0.64	2
9	139	187	29	77	0.51	2
10	131	162	36	44	0.54	2
11	109	119	28	50	0.51	2
12	108	126	40	41	0.46	2
13	123	174	38	23	0.58	2
14	107	143	18	34	0.63	2
15	74	64	17	50	0.46	2

are shown in Figure 6. As we can see, in terms of both NMI and error rate, our method DBNMF has a better performance than that of FaceNet under the three cases, which represents DBNMF's stronger ability to both find the true community structure and fit the observed data in the temporal network. Besides, the mean NMI of the result of our method DBNMF for the temporal network with snapshots $T = 10$ is minimum, and that for $T = 30$ is maximum; opposite to the error rate. This further validates that, our method could reflect the temporal characteristics of the temporal networks.

5.3.2. NEC blog network

The NEC blog network is constituted by the blogs and their relations. A detailed introduction can be found in [37]. There are 14,8681 edges among 407 blogs during 15 months in this data. We assume that the temporal network has 15 snapshot networks and the data in each month is taken as a snapshot network. We assume that the network is unweighted for analysis. The characteristics of the blog network are presented in Table 2 and the relevant symbols defined here are consistent with those in Table 1. From this table, we find that the value of NJ is low in some snapshot networks, which means that the network may change a lot from one snapshot network to the next. The performances of DBNMF and FaceNet are both shown in Figure 7.

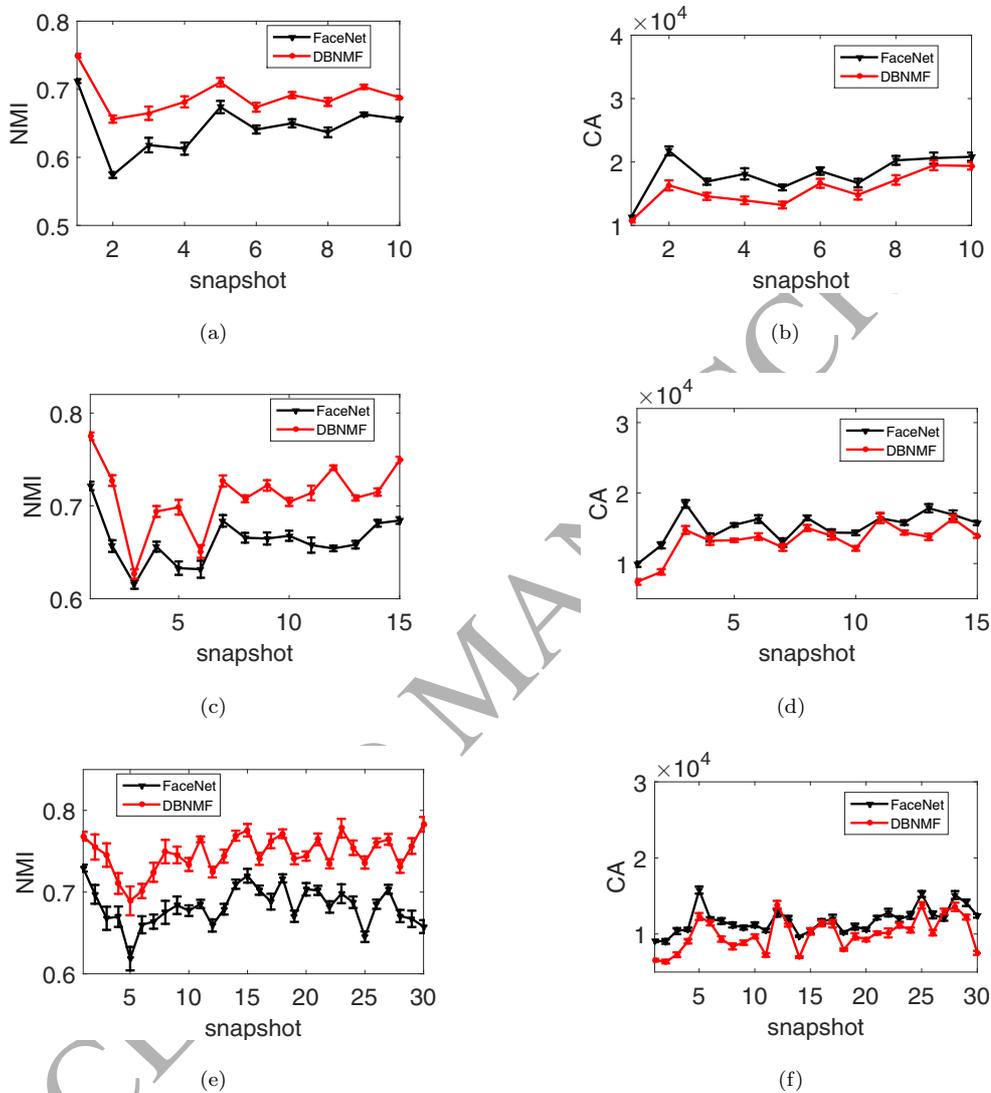


Figure 6: The result on the KIT-email network: (a): The NMI on the network with $T = 10$; (b): The error rate on the network with the number of snapshots $T = 10$; (c): The NMI on the network with the number of snapshots $T = 15$; (d): The error rate on the network with the number of snapshots $T = 15$; (e): The NMI on the network with the number of snapshots $T = 30$; (f): The error rate on the network with the number of snapshots $T = 30$. The red lines and the black lines represent the results of our method and the FaceNet method, respectively. Error bars show the standard deviations estimated on the network with 10 random initialization.

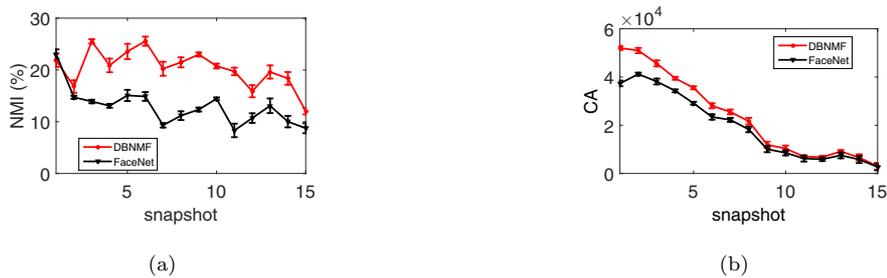


Figure 7: The result on the NEC blog network: (a) The NMI accuracy; (b) The error rate. The black lines represent the results of FaceNetod, the red lines represent the results of our DBNMF. Error bars show the standard deviations estimated on the network with 10 random initialization.

As we can see, although the NMI and the error rate of these two methods vary fast from one snapshot to the next, our DBNMF still outperforms FaceNet for this temporal network in terms of both these two metrics, especially when we use the NMI index.

5.3.3. DBLP cooperation network

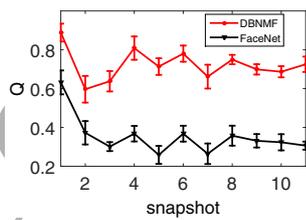
Here we analyze a rapidly changing temporal network named the DBLP cooperation network, in which the authors are denoted as the nodes and a cooperation of two authors in one paper represents a link in the network. This network contains three main areas: data mining, databases and artificial intelligence from 1996 to 2006. We assume that the temporal network has 11 snapshot networks and the data in each year is taken as a snapshot network. The characteristics defined can be seen in Table 3, which shows that there is a very low edge density and a fast rate of new nodes adding or odd nodes removing in this network, so it makes the community detection of this temporal network more challenging. The experimental results are shown in Figure 8, which was evaluated based on the error rate and the modularity Q . Because the performances of our DBNMF and FaceNet have a nearly equal value based on the NMI, we do not give their comparison in terms of the NMI metric. Here, the results show that our DBNMF method has a better performance comparing with the FaceNet based both on the error rate and the modularity Q .

5.3.4. AI cooperation network

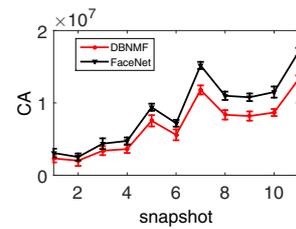
Furthermore, we analyze a temporal cooperation network in the area of artificial intelligence which is a subset of the DBLP cooperation network. The definition of this network is the same as that defined in the DBLP cooperation network. This is a highly dynamic network in which nodes and edges change very quickly, and especially the aforementioned index NJ is much smaller now. The time stamp of this temporal network ranges from year 2008 to 2012, We assume that the temporal network has 5 snapshot networks and the data in each year is taken as a snapshot network. The experimental results of these two methods in term of the NMI and the error rate are given in Figure 9. As we can see, the performance of our method outperforms that of FaceNet based on both these

Table 3: Statistical properties of the DBLP cooperation network

t	N	E	NA	ND	NJ	K
1	2635	3917	no	no	no	3
2	2336	3931	1595	1894	0.18	3
3	3088	5120	2227	1475	0.19	3
4	3240	5645	2166	2014	0.20	3
5	4550	8572	3291	1981	0.19	3
6	4005	6867	2565	3110	0.20	3
7	5816	10217	4148	2337	0.20	3
8	4966	9382	3012	3862	0.22	3
9	4919	9781	3034	3081	0.24	3
10	5085	10067	3201	3035	0.23	3
11	6138	11936	4012	2959	0.23	3



(a)



(b)

Figure 8: The result on the DBLP cooperation network: (a) The modularity Q ; (b) The error rate. The black lines represent the results of FaceNet, the red lines represent results of our DBNMF. Error bars show the standard deviations estimated on the network with 10 random initialization.



Figure 9: The result on the AI cooperation network: (a) The NMI accuracy; (b) The error rate. The black lines represent the results of FaceNet, the red lines represent the results of our DBNMF. Error bars show the standard deviations estimated on the network with 10 random initialization.

metrics, especially when we use the NMI index.

5.3.5. DBLP citation network

We analyze a larger and sparser temporal network namely DBLP citation network. We analyze the data provided by KDD Cup 2003, which is a knowledge discovery and data mining competition held in conjunction with the Ninth Annual ACM SIGKDD Conference.

This temporal network has already been used for the analysis of community detection and network evolution. Similarly to [41], we construct this temporal network as follows. The nodes represent the papers, and the edges represent the citation relations between the papers. We select the data whose time stamp ranges from years 1994 to 2002, and there are 26,399 nodes in the network. Notice that, as a paper published in 1994 may also be cited by other papers in 2002, we have a big reason to analyze the network by cumulative snapshot networks. That is, the size of this temporal network would become larger and larger, and there are no nodes and edges removal. From this viewpoint, we construct the temporal network in this way. We set the time duration of every snapshot network as the following four situations. They denote $\delta = 3$ months, $\delta = 4$ months, $\delta = 6$ months and $\delta = 12$ months, which corresponds to the snapshots $T = 36$, $T = 27$, $T = 18$ and $T = 9$, respectively. For example, in the situation with the number of snapshot networks $T = 9$, each snapshot network is formed with the data that comes from year 1994, 1994 – 1995, \dots , 1994 – 2002, respectively. As the number of nodes in this network is changing and there is no ground truth of the community structure for each snapshot of the temporal network, we cannot give the result of the other benchmarked methods because the FaceNet method needs the number of communities in each snapshot and the DYNMOGA cannot be used to deal with the change of the number of nodes in the network. Thus, we only show the results of our method here. But we will further explain the reasonability of our obtained overlapping community structure using the case study analysis.

For this temporal network, the number of the nodes and the number of the edges in each snapshot network are shown in Figure 10. As we can see, the growth of the number of the nodes is near linear, and the growth of the number of the links is exponential.

We give the community detection results of our DBNMF method evaluated based on the mod-

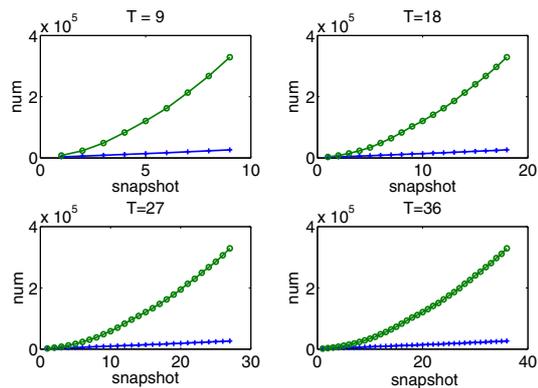


Figure 10: The statistical properties of the DBLP citation network: the number of the nodes and the number of the edges in the DBLP citation network with different time duration of each snapshot network. The cyan lines represent the number of the edges in each snapshot network, and the blue lines represent the number of the nodes in each snapshot network.

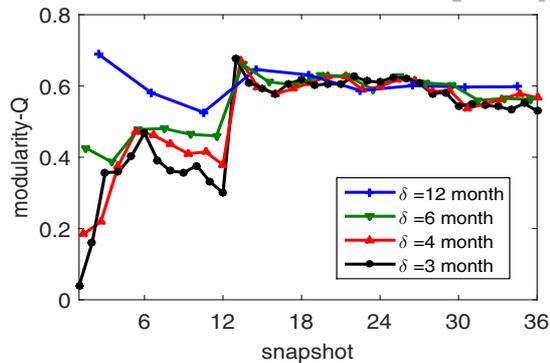


Figure 11: The modularity Q on the DBLP citation network.

ularity Q which is shown in Figure 11. We find that, the modularity Q does not increase since
 510 year 1997, which is independent with the size of each snapshot network. This may mean that the
 number of citation relations of these papers increase among different research fields. Using the fuzzy
 modularity as the quality metric, we analyze the detected overlapping community structure of our
 method DBNMF, shown as Figure 12. And we find similar tendency with that which is based on
 the modularity Q .

515 Furthermore, we analyze some popular nodes (papers) in this temporal network. For instance,
 the paper '*the large N limit of superconformal field theories and supergravity*' has been cited 2,369
 times in our data but 12,740 times as recorded by Google scholar. The papers which cite it refer
 to almost all of the physical areas, including High Energy Physics, Mathematical Physics, Thermal
 Gauge Theories, Nuclear Physics, and so on. In Figure 13, we demonstrate some most popular nodes
 520 in the temporal network, and show the comparison between the papers that cite these popular nodes

in our data and these in Google scholar. As a result, there is an obvious trend that a paper with a large citation number in our data will have a large citation number in Google scholar, and also it will have a large probability to be cited by papers in other fields.

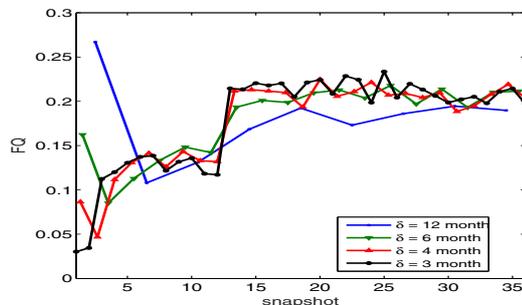


Figure 12: The fuzzy-modularity value of our algorithm on the DBLP citation network.

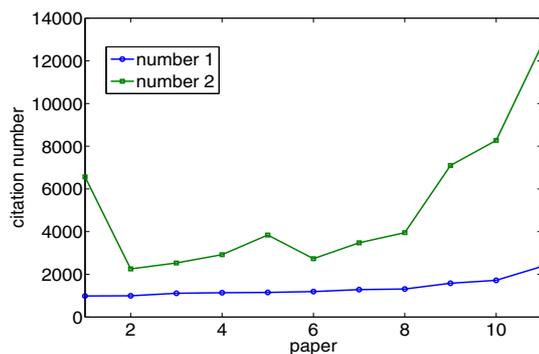


Figure 13: Top 11 cited papers in DBLP citation network: the blue line represents the citation number in the DBLP citation data; the cyan lines represent the citation number count by the Google scholar recently. The title of these paper are in turns as: 'an alternative to compactification'; 'Monopoles, Duality and Chiral Symmetry Breaking in $N=2$ Supersymmetric'; 'String Theory Dynamics In Various Dimensions'; 'Dirichlet-Branes and Ramond-Ramond Charges'; 'String Theory and Noncommutative Geometry'; 'M Theory As A Matrix Model: A Conjecture'; 'Electric-magnetic duality, monopole condensation, and confinement in $N= 2$ supersymmetric Yang-Mills theory'; 'Large N Field Theories, String Theory and Gravity'; 'Gauge Theory Correlators from Non-Critical String Theory'; 'Anti De Sitter Space And Holography'; 'The large N limit of superconformal field theories and supergravity'.

5.4. Parameter Analysis

525

In this section, we analyze the sensitivity of the parameters of our method DBNMF in detail. We show that the performance of DBNMF is almost not sensitive with the changes of the hyper-parameters a and b . Besides, $\alpha = 0.9$ is a good choice irrespective of it being evaluated by the NMI or by the error rate (CA). We offered a detailed analysis on the KIT-email network with the number of snapshot networks $T = 15$ and the results are shown in Figures 14 and 15. The experiments on



Figure 14: The parameter analysis based on KIT email network with the number of snapshot networks T : (a) The NMI on the KIT email network; (b) The error rate on the KIT email network. Every node in the figure represents the mean and standard deviation of T snapshot networks with different parameters α

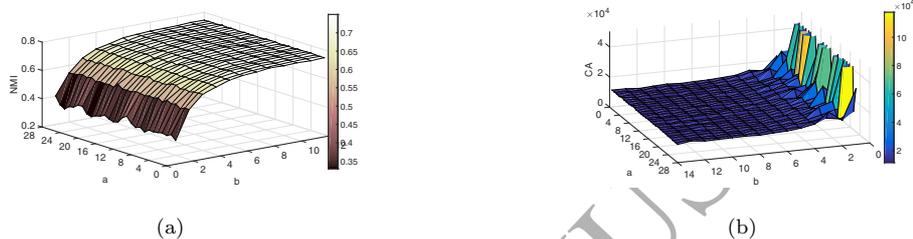


Figure 15: The hyper-parameter analysis based on KIT email network with the number of snapshot networks $T = 15$: (a) The NMI on the KIT email network; (b) The error rate on the KIT email network. Every node in the figure represents the expectation values with different hyper-parameter a and b based on the temporal network

530 the dynamic Girven-Newman synthetic networks and that on other dynamic real networks offered similar conclusion, although we do not show their results here.

Figure 14 represents the analysis on the KIT-email network with the number of snapshot networks $T = 10$, $T = 15$ and $T = 30$. Each point shows the average values of NMI and error rate (CA) on the full temporal network. The results show that our DBNMF method often has the best result at $\alpha = 0.9$. Figure 15 shows that the performance of the method is almost not sensitive to the hyper-parameter a and $b > 4$. In fact, b is the scale parameter in gamma distribution and this is consistent with our intuition and the property of gamma distribution.

As we can see from the experiments on the artificial and real-world temporal networks, the DBNMF model has a better performance than all the methods compared. The advantages of our proposed model can be summarized in the following. First, the model is interpretable and intuitive for both overlapping and non-overlapping community detection based on the probabilistic community memberships we derived. Second, with the assumption that each link in the temporal network is drawn from a Poisson distribution, the generated model shows a good fitness to the observed network, and meanwhile, the regularization term in our model are also interpretable which helps to ensure the smooth evolution of the temporal network. Third and most importantly, using a Bayesian method the model can determine the number of communities automatically at each snapshot of the temporal

network, which makes the model more appropriate for the real and unexplored temporal networks. Besides, a good initialization we introduced for the algorithm and a suitable parameter selection based on the half normal and Gamma prior can also help the proposed DBNMF model and make it often have well and stable results in the experiments.

6. Conclusions and Discussion

In this paper, we propose a probabilistic model named Dynamic Bayesian Dynamic Non-negative Matrix Factorization (DBNMF), which belongs to the framework of the evolutionary clustering, for the detection of communities in temporal networks. The DBNMF model can automatically determine the number of communities and detect the overlapping community structure in temporal networks. The determination of the number of communities is based on the automatic relevance determination. The overlapping community structure is obtained based on the probabilistic group membership of nodes in every snapshot network which has a good theoretical interpretation based on the nonnegative matrix factorization (NMF). Finally, we proposed a gradient descent algorithm for the optimization of our DBNMF model based on the multiplication update rule.

We give experimental results on both dynamic Girven-Newman synthetic networks and some real world datasets. First, the experiments on the dynamic Girven-Newman synthetic networks show that, compared with the state-of-the-art methods, our method is better suited to this synthetic networks and to a higher accuracy of the community detection. Second, a detailed analysis of the result on the KIT email networks shows that our method can get a better performance for the temporal network with different time duration of each snapshot when compared with that of the FaceNet, and the result on the two cooperation networks gives a similar conclusion. Furthermore, we analyze a larger and sparser temporal citation network, the experimental results also show that our method has a strong ability to detect the overlapping community structure in terms of fuzzy modularity. To sum up, all results on the synthetic networks and real world data with known ground truth of community structure indicate that our method can automatically determine the number of communities. Finally, our parameters analysis demonstrates that the performance of the method is almost not sensitive to the hyper-parameters, and the balance parameter $\alpha = 0.9$ is always a good selection for our method.

As we can see from the experiments on different temporal networks, the proposed model DBNMF is more suitable to be used on sparse and temporal networks, especially for the networks when the number of communities are unknown, and the number of nodes and communities are varying with time. There are also some problems for our method to be further studied. Ideally, the selection of the parameter α should be modified automatically on the temporal network, and not just confirmed by the experiments. Adaptive Evolutionary Clustering [42] may work in this situation, but it is not suitable for large real temporal networks. So a general and simple method with automatic

determination of the parameter α will be our future work. Another important and formidable problem is the evaluating metric for the dynamic community structure. Although the designed $R - error$ can be applied in some cases, a more general and standard metric is imperative to the dynamic community structure. This new metric should be suitable for the changes of both the number of nodes and the number of communities in temporal networks, and it may be our next work in future. In addition, although we have analyzed multiple situations with different snapshots for the temporal networks, what is the best selection of the number of snapshots for a given temporal network is still an issue.

Acknowledgments

The work was supported by the major research plan of the National Natural Science of China (91224009, 51438009), the National Natural Science Foundation of China (61303110), and the PhD Programs Foundation of Ministry of Education of China (20130032120043).

References

- [1] M. Girvan, M. E. J. Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences 99 (12) (2002) 7821–7826. `arXiv:http://www.pnas.org/content/99/12/7821.full.pdf`, `doi:10.1073/pnas.122653799`.
URL `http://www.pnas.org/content/99/12/7821.abstract`
- [2] V. Spirin, L. A. Mirny, Protein complexes and functional modules in molecular networks, Proceedings of the National Academy of Sciences 100 (21) (2003) 12123–12128. `arXiv:http://www.pnas.org/content/100/21/12123.full.pdf`, `doi:10.1073/pnas.2032324100`.
URL `http://www.pnas.org/content/100/21/12123.abstract`
- [3] D. Lusseau, The emergent properties of a dolphin social network, Proceedings of the Royal Society of London B: Biological Sciences 270 (Suppl 2) (2003) S186–S188. `doi:10.1098/rsbl.2003.0057`.
- [4] G. Palla, I. Derenyi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (7043) (2005) 814–818.
URL `http://dx.doi.org/10.1038/nature03607`
- [5] D. Jin, B. Gabrys, J. Dang, Combined node and link partitions method for finding overlapping communities in complex networks, Sci. Rep. 5 (2015) –.
URL `http://dx.doi.org/10.1038/srep08600`

- [6] P. Schuetz, A. Caffisch, Efficient modularity optimization by multistep greedy algorithm and vertex mover refinement, *Phys. Rev. E* 77 (2008) 046112. doi:10.1103/PhysRevE.77.046112. URL <http://link.aps.org/doi/10.1103/PhysRevE.77.046112>
- 615 [7] M. J. Newman, A measure of betweenness centrality based on random walks, *Social Networks* 27 (1) (2005) 39 – 54. doi:<http://dx.doi.org/10.1016/j.socnet.2004.11.009>. URL <http://www.sciencedirect.com/science/article/pii/S0378873304000681>
- [8] I. Derényi, G. Palla, T. Vicsek, Clique percolation in random networks, *Phys. Rev. Lett.* 94 (2005) 160202. doi:10.1103/PhysRevLett.94.160202.
- 620 URL <http://link.aps.org/doi/10.1103/PhysRevLett.94.160202>
- [9] J. Reichardt, D. R. White, Role models for complex networks, *The European Physical Journal B* 60 (2) (2007) 217–224. doi:10.1140/epjb/e2007-00340-y. URL <http://dx.doi.org/10.1140/epjb/e2007-00340-y>
- [10] G. Palla, A.-L. Barabasi, T. Vicsek, Quantifying social group evolution, *Nature* 446 (7136) (2007) 664–667.
- 625 URL <http://dx.doi.org/10.1038/nature05670>
- [11] P. Holme, J. Saramki, Temporal networks, *Physics Reports* 519 (3) (2012) 97 – 125, temporal Networks. doi:<http://dx.doi.org/10.1016/j.physrep.2012.03.001>. URL <http://www.sciencedirect.com/science/article/pii/S0370157312000841>
- 630 [12] R. Kumar, J. Novak, A. Tomkins, Structure and evolution of online social networks, in: P. S. Yu, J. Han, C. Faloutsos (Eds.), *Link Mining: Models, Algorithms, and Applications*, Springer New York, 2010, pp. 337–357. doi:10.1007/978-1-4419-6515-8_13. URL http://dx.doi.org/10.1007/978-1-4419-6515-8_13
- [13] J. Leskovec, L. Backstrom, R. Kumar, A. Tomkins, Microscopic evolution of social networks, in: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08*, ACM, New York, NY, USA, 2008, pp. 462–470. doi:10.1145/1401890.1401948.
- 635 URL <http://doi.acm.org/10.1145/1401890.1401948>
- [14] T. Yang, Y. Chi, S. Zhu, Y. Gong, R. Jin, Detecting communities and their evolutions in dynamic social networks bayesian approach, *Machine Learning* 82 (2) (2011) 157–189. doi:10.1007/s10994-010-5214-7.
- 640 URL <http://dx.doi.org/10.1007/s10994-010-5214-7>
- [15] D. Chakrabarti, R. Kumar, A. Tomkins, Evolutionary clustering, in: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '06*,

- 645 ACM, New York, NY, USA, 2006, pp. 554–560. doi:10.1145/1150402.1150467.
URL <http://doi.acm.org/10.1145/1150402.1150467>
- [16] N. P. Nguyen, T. N. Dinh, Y. Shen, M. T. Thai, Dynamic social community detection and its applications, PLoS ONE 9 (4) (2014) e91431. doi:10.1371/journal.pone.0091431.
URL <http://dx.doi.org/10.1371/journal.pone.0091431>
- 650 [17] J. Sun, C. Faloutsos, S. Papadimitriou, P. S. Yu, Graphscope parameter-free mining of large time-evolving graphs, in: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '07, ACM, New York, NY, USA, 2007, pp. 687–696. doi:10.1145/1281192.1281266.
URL <http://doi.acm.org/10.1145/1281192.1281266>
- 655 [18] D. P. Wipf, S. S. Nagarajan, A new view of automatic relevance determination, in: J. Platt, D. Koller, Y. Singer, S. Roweis (Eds.), Advances in Neural Information Processing Systems 20, Curran Associates, Inc., 2008, pp. 1625–1632.
URL <http://papers.nips.cc/paper/3372-a-new-view-of-automatic-relevance-determination.pdf>
- 660 [19] F. Folino, C. Pizzuti, An evolutionary multiobjective approach for community discovery in dynamic networks, IEEE Transactions on Knowledge and Data Engineering 26 (8) (2014) 1838–1852. doi:10.1109/TKDE.2013.131.
- [20] J. Reichardt, S. Bornholdt, Statistical mechanics of community detection, Phys. Rev. E 74 (2006) 016110. doi:10.1103/PhysRevE.74.016110.
665 URL <http://link.aps.org/doi/10.1103/PhysRevE.74.016110>
- [21] Y. Chi, X. Song, D. Zhou, K. Hino, B. L. Tseng, On evolutionary spectral clustering, ACM Trans. Knowl. Discov. Data 3 (4) (2009) 17:1–17:30. doi:10.1145/1631162.1631165.
URL <http://doi.acm.org/10.1145/1631162.1631165>
- [22] M.-S. Kim, J. Han, A particle-and-density based evolutionary clustering method for dynamic networks, Proc. VLDB Endow. 2 (1) (2009) 622–633. doi:10.14778/1687627.1687698.
670 URL <http://dx.doi.org/10.14778/1687627.1687698>
- [23] L. Tang, L. Huan, J. Zhang, Identifying evolving groups in dynamic multimode networks, IEEE Transactions on Knowledge and Data Engineering 24 (1) (2012) 72–85. doi:10.1109/TKDE.2011.159.
- 675 [24] T. Xu, Z. Zhang, P. S. Yu, B. Long, Evolutionary clustering by hierarchical dirichlet process with hidden markov state, in: 2008. ICDM '08. Eighth IEEE International Conference on Data Mining, 2008, pp. 658–667. doi:10.1109/ICDM.2008.24.

- [25] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, B. L. Tseng, Facetnet a framework for analyzing communities and their evolutions in dynamic networks, in: Proceedings of the 17th International Conference on World Wide Web, WWW '08, ACM, New York, NY, USA, 2008, pp. 685–694. doi:10.1145/1367497.1367590.
URL <http://doi.acm.org/10.1145/1367497.1367590>
- [26] I. Psorakis, S. Roberts, M. Ebdon, B. Sheldon, Overlapping community detection using bayesian non-negative matrix factorization, Phys. Rev. E 83 (2011) 066114. doi:10.1103/PhysRevE.83.066114.
URL <http://link.aps.org/doi/10.1103/PhysRevE.83.066114>
- [27] N. M. Ahmed, L. Chen, An efficient algorithm for link prediction in temporal uncertain social networks, Information Sciences 331 (C) (2016) 120–136.
- [28] Y. Yuan, G. Wang, L. Chen, B. Ning, Efficient pattern matching on big uncertain graphs, Information Sciences 339 (C) (2016) 369–394.
- [29] A. Rezvanian, M. R. Meybodi, Finding minimum vertex covering in stochastic graphs: a learning automata approach, Cybernetics and Systems 46 (8) (2015) 698–727.
- [30] L. Du, C. Li, H. Chen, L. Tan, Y. Zhang, Probabilistic SimRank computation over uncertain graphs, Information Sciences 295 (C) (2015) 521–535.
- [31] V. Y. Tan, C. Fevotte, Automatic relevance determination in nonnegative matrix factorization with the β -divergence, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (7) (2013) 1592–1605. doi:10.1109/TPAMI.2012.240.
- [32] D. Jin, H. Wang, J. Dang, D. He, W. Zhang, Detect Overlapping Communities via Ranking Node Popularities., AAAI (2016) 172–178.
- [33] B. Ball, B. Karrer, M. E. J. Newman, Efficient and principled method for detecting communities in networks, Physical Review E 84 (3) (2011) 036103–13.
- [34] B. Karrer, M. E. J. Newman, Stochastic blockmodels and community structure in networks, Physical Review E 83 (1) (2011) 016107–10.
- [35] D. D. Lee, H. S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.
URL <http://dx.doi.org/10.1038/44565>
- [36] L. Danon, A. Daz-Guilera, J. Duch, A. Arenas, Comparing community structure identification, Journal of Statistical Mechanics: Theory and Experiment 2005 (09) (2005) P09008.
URL <http://stacks.iop.org/1742-5468/2005/i=09/a=P09008>

- 710 [37] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, B. L. Tseng, Analyzing communities and their evolu-
tions in dynamic social networks, *ACM Transactions on Knowledge Discovery from Data* 3 (2)
(2009) 8:1–8:31. doi:10.1145/1514888.1514891.
URL <http://doi.acm.org/10.1145/1514888.1514891>
- [38] M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, *Phys.*
715 *Rev. E* 69 (2004) 026113. doi:10.1103/PhysRevE.69.026113.
URL <http://link.aps.org/doi/10.1103/PhysRevE.69.026113>
- [39] T. Nepusz, A. Petróczy, L. Négyessy, F. Bacsó, Fuzzy communities and the concept of bridgeness
in complex networks, *Phys. Rev. E* 77 (2008) 016107. doi:10.1103/PhysRevE.77.016107.
URL <http://link.aps.org/doi/10.1103/PhysRevE.77.016107>
- 720 [40] C. Matias, V. Miele, Statistical clustering of temporal networks through a dynamic stochastic
block model, *arXiv.org* arXiv:1506.07464v1.
- [41] J. Gehrke, P. Ginsparg, J. Kleinberg, Overview of the 2003 kdd cup, *ACM SIGKDD Explo-*
rations Newsletter 5 (2) (2003) 149–151. doi:10.1145/980972.980992.
URL <http://doi.acm.org/10.1145/980972.980992>
- 725 [42] K. Xu, M. Klinger, A. Hero III, Adaptive evolutionary clustering, *Data Mining and Knowledge*
Discovery 28 (2) (2014) 304–336. doi:10.1007/s10618-012-0302-x.
URL <http://dx.doi.org/10.1007/s10618-012-0302-x>