Research Article

A Novel Emerging Topic Identification and Evolution Discovery Method on Time-Evolving and Heterogeneous Online Social Networks

Xiaoyan Xu ¹, Wei Lv ¹, Beibei Zhang ², Shuaipeng Zhou ³, Wei Wei ²,
and Yusen Li ¹

¹School of Science, Xi'an Shiyou University, Xi'an 710056, China

²School of Computer and Engineering, Xi'an University of Technology, Xi'an 710048, China

³Aamaze Data Company, Xi'an, China

Correspondence should be addressed to Beibei Zhang; beibeizhang115@hotmail.com

Received 3 September 2020; Revised 4 July 2021; Accepted 2 August 2021; Published 27 August 2021

Academic Editor: Jia Wu

Copyright © 2021 Xiaoyan Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the fast development of web 2.0, information generation and propagation among online users become deeply interweaved. How to effectively and immediately discover the new emerging topic and further how to uncover its evolution law are still wide open and urgently needed by both research and practical fields. This paper proposed a novel early emerging topic detection and its evolution law identification framework based on dynamic community detection method on time-evolving and scalable heterogeneous social networks. The framework is composed of three major steps. Firstly, a time-evolving and scalable complex network denoted as KeyGraph is built up by deeply analyzing the text features of all kinds of data crawled from heterogeneous online social network platforms; secondly, a novel dynamic community detection method is proposed by which the new emerging topic is detected on the modeled time-evolving and scalable KeyGraph network; thirdly, a unified directional topic propagation network modeled by a great number of short texts including microblogs and news titles is set up, and the topic evolution law of the previously detected early emerging topic is identified by fully utilizing local network variations and modularity optimization of the “time-evolving” and directional topic propagation network. Our method is proved to yield preferable results on both a huge amount of computer-generated test data and a great amount of real online network data crawled from mainstream heterogeneous social networks.

1. Introduction

In recent years, with the fast development of web 2.0, social network sites such as Facebook, Sina microblog, and Twitter rise in a short time, a huge heterogeneous online social networks have gradually formed on which the functional role of online users is changing from the information consumers to both diffusers and generators [1]. Information of different online social networks propagates in a deeply mingled way. For example, information from news websites is reposted to sites like personal (micro) blogs that have specific following groups; BBS (bulletin board system) of which the information bares the broadcasting attribute is shared to the personal (micro) blogs. Activities above making the information production and

propagation among various users become a huge data-heterogeneous, time-evolving, and scalable complex network. Thus, how to efficiently and timely identify and reveal the new emerging topic and even its evolution process (law) on this scalable, time-evolving, and heterogeneous online social network has become the hot research spot in topic detection and tracking fields lately.

It is widely known that a complex network bears the community structure, which represents a clustering of network nodes with densely interweaved edges within groups but spares connections between them [2]. Community structure not only reveals the coarse structure of the network but also plays an important role in the functioning of the network [3, 4]. For example, community in the social

network represents the real social groups composed of people having the same backgrounds or interests; community in collaboration network represents the related papers with the same research subject; community in the biology or circuit network represents nodes group with the same network function. Deng et al. [5] proposed a hot topic detection algorithm based on a community of networks. Lin and Guan-Zhong. [6] also found the forum hot topic using the community detection concept applied on the BBS network and validated the efficiency and consistency by manual calibration of the identified topic and the community. Some researchers [7, 8] discovered the epidemic spreading mechanism by using the community structure analysis of complex networks. Identification of community structure in the complex network usually can realize certain application-based purposes and thus, the community identification and topology evolution discovery have become the most significant focus in the complex network structure analysis field. Although community detection/identification in networks has been studied for many years, most existed approaches are designed simply for the static network [9, 10] and unified homogeneous network [11]. However, in the real-world and in this paper, the studied social network is time-evolving and heterogeneous due to the time-varying social communications and the time-dependent interactions of different social network platforms.

Although it has been concluded that the traditional static community detection methods could be applied to the time-evolving heterogeneous network by converting the time-evolving network into a series of static snapshots of sub-networks through rearranging the nodes and links belonging to the same time stamps, while in this process, the semantic relationships and dynamic properties of communities may have violently been damaged or even lost. Another concluded method is to identify network community not from scratch but by storing and using the historical results of execution of static community detection algorithm along with the network evolving process [12, 13]. However, a great amount of time and space computing costs are needed and the efficiency of the algorithm becomes slower with time flying. Lately, community discovery in the time-evolving and heterogeneous network has emerged as an outstanding challenge and has attracted much attention of researchers. Sun et al. [14] proposed a Dirichlet Process Mixture Model based algorithm to describe the community detection in a heterogeneous star-model network.

In summary, when applying existing community detection methods for time-evolving and heterogeneous networks, three main problems are usually encountered: (1) most existed community detection methods are proposed simply for the static and homogeneous network; (2) the semantic relationships and dynamic properties of communities are violently damaged and even bluntly lost due to the man-made segmentation of network; (3) a great amount of computing time and space cost is required by storing the historical community structure information as the initial input values.

Aiming to tackle these problems, we propose an emerging topic identification and evolution topology discovery framework based on a novel dynamic community

detection method on the time-evolving and heterogeneous social network. Firstly, a unified short-text network is constructed by modeling heterogeneous short texts crawled from different online social networks into a network. It is denoted as KeyGraph based on the cooccurrence of keywords of the crawled short texts. Secondly, a novel dynamic community detection method with a well-known static community detection algorithm as the Louvain algorithm corely embedded is proposed and is applied on the KeyGraph, as a result of which, the new emerging topic in the form of newborn community is detected. Finally, the topic evolution topology is discovered by deep analysis of the community scale and nodes variation with time-evolving of the detected communities.

The rest of this paper is presented as follows. We briefly review the related research work in Section 2, our methodology is presented in Section 3, numerical results and evaluations are presented in Section 4, and finally, we conclude and discuss our future work in Section 5.

2. Related Work

Topic detection and tracking (TDT) is firstly put forward by DARPA in 1996. Its original objective is to automatically identify online public sentiments in the form of a topic from the network media stream and further to track the propagation and diffusion process of the previously identified/detected topic. Later, TDT has become the key technology in the fields of Internet public opinion/sentiment mining field. Classic TDT methods include the latent Dirichlet analysis (LDA) [15] and the probabilistic latent semantic analysis (PLSA) [16] methods. Their main concepts are that a topic is based on the probability distributions of sets of words such that the probability distribution of word cooccurrences among sets of words is maximized/optimized.

Community is another important property of a complex network except for the small-world and scale-free properties. It not only provides a coarse view of the network structure but also actually plays and represents certain functions of the network. Community describes the closeness among nodes which means that nodes are closely interrelated within groups, while nodes between different communities are loosely connected.

Sayyadi and Raschid [17] proposed a graph analytical method to detect and identify the topic; they proposed the KeyGraph algorithm to transform original texts data into a term graph based on properties of cooccurrence relations of texts data with each other. Furthermore, they utilized a community detection method to part the constructed KeyGraph network into community topology and they deemed each identified community as a detected topic. Nowadays, community detection is a fundamental technique of network structure analysis, many creative methods for discovering communities in a static and homogeneous network have been deployed in the past decades. It can be commonly divided into two classes: graph theory-based algorithm and sociology-based algorithm. Sociology-based algorithms can be generally divided into division and aggregation methods. The classical GN algorithm [2] belongs

to the division method, and its fundamental principle is to obtain network community by finding the edge with the highest score of betweenness and by removing it from the network. Newman proposed a fast aggregation algorithm [9], which has similar accuracy with GN, but the performance has been significantly improved. Blonde et al. [18] proposed the Louvain algorithm based on the modularity optimization method, which is a simple, efficient, and easy-to-implement method for finding community structure in a large-scale network. The method is actually a greedy optimization method that attempts to optimize the tag-indexed “modularity” of every possible partition of the target network.

However, the resolution limit problem is commonly encountered using the modularity-based community detection method in a static and homogeneous network. Here, the “static” mainly refers to both the time and the whole network structure staying static without variation with time flying by. The resolution limit problem means when the scale of the network is large enough, a small community in a large network cannot be properly and efficiently detected, which results in the overlapping community phenomena. This phenomenon is called the resolution limit problem in a large-scale network when using the modularity method for network community detection purposes [19, 20]. These community detection methods, however, lack the capability of dealing with time-evolving and cannot be directly used for heterogeneous networks. As our problem of identifying and revealing the emerging topic and its evolution topology on the large-scale, time-varying, and heterogeneous online social network, the previous and classic community detection methods face great challenges.

2.1. Dynamic Community Detection of Homogeneous Social Network. Tracking the evolution topology of detected community need to take the dynamic property of the time-evolving network into consideration. A commonly utilized and concluded framework [21–24] is to apply the static community detection algorithms for each static snapshots subnetwork composed of nodes and edges with the same timestamp of the time-evolving network and then to generate the evolution of community by computing the community closeness between two adjacent static snapshots subnetworks. Toyoda and Kitsuregawa [25] firstly selected these web pages with high focus numbers (thumb-up number) as seed web pages and finally obtained the community including the seed web pages by utilizing the page closeness calculation algorithm with hyperlink-induced topic search as core calculation framework. Palla et al. [26] obtained the community topology of one snapshot using the clique percolation clustering method and evaluated its usefulness by applying it to the scientists’ cooperation network and the telecommunication users network. Chakrabarti et al. [27] proposed an evolution clustering model with k-means and hierarchy clustering method to identify the community evolution law in the process of dynamic community detection.

Another policy to track community evolution in the time-evolving network is by integrating the optimization of both modularity and the structure of local network variations into a multiobjective optimization problem. Its main concept is that treating the community topology of the previous timestamp $t - \Delta t$ as baseline network, at the present time epoch t , the variations of the network during the time range Δt , i.e., $[t - \Delta t, t]$ are the main focus other than the whole network of the present time being t . By detecting the community topology of the variation part of the network during the time range Δt while the other part of the network remains unchanged to improve the whole efficiency of the community detection algorithm [28–33]. Yang and Liu [28] proposed the physical incremental model by modeling the relationship between nodes of a network controlled by attraction force and repulsive force proposed in Newtonian mechanics. Other incremental dynamic community detection methods usually utilize the key feature of the network; these algorithms firstly obtain the community topology of network at the initial snap shot commonly using the static community detection method, then with time flying by, the variations of network during the man-set time range is recalculated and its community topology is identified [29–33]. However, these methods are usually designed for homogeneous networks.

2.2. Dynamic Community Detection of Heterogeneous Social Network. Lately, community detection in a heterogeneous network has become a hot research spot. Zhao et al. [34] proposed a uniform framework for detecting and tracking community evolution. They firstly model the entities and their relationships with the same timestamp into a heterogeneous network. They secondly extracted the snapshot-based feature and delta-based feature by utilizing the autoregression method to finally obtain the community topology and its evolution law. Sun et al. [35] introduced the community evolution in multimode networks and proposed a framework that partitioned the multimode network into a set of bipartite networks. Sun et al. used net clusters [14] to describe the community and proposed the Evo-NetCluster to detect the community automatically. Wu et al. [36] proposed a tensor decomposition framework to detect community in the general time-evolving heterogeneous network. Nevertheless, these methods either need to know the topology schema like star or bipartite or need to satisfy the requirements of tensor factorization, which are intractable/hard to use in real applications. Tang et al. [37] proposed a principal modularity maximization method, in which they first analyzed the modularity of different relational dimensions, then according to its eigenvalue and eigenvector of each relational dimension, the principal structural feature was extracted; thirdly, they correlated every network principal structural features to acquire the shared community topology of the whole network which would make the whole network modularity optimized.

With the rapid development of web 2.0 and mobile networks, event detection on heterogeneous data has drawn more attention in recent years. Yang et al. [38] proposed a

unified model to dynamically learn how to represent the data with different features of a heterogeneous social network. Liu et al. [39] treated the breaking news as a heterogeneous social data stream and developed how to extract events from the dynamic data stream. Liu et al. [40] extended the heterogeneous data stream into a multilingual scenario; Cao et al. [41] developed a knowledge-preserving and incremental social event detection framework using the GNNs and they applied it on the heterogeneous social network.

In summation, the TDT methods given above face three challenges as follows. Firstly, though most previous topic detection methods have good results in static online social networks, they rarely relate to the research of new emerging topic detection under the time-evolving and dynamic social network situations. Secondly, topic detection researchers mainly focus on finding new methods to detect prominent or distinct topics. They pay little attention to reveal the topic evolution process with time flying in the meantime of the topic detection process. Thirdly, the resolution limit problem has still not been well solved in the existing modularity community detection methods.

In this paper, we propose an original emerging topic detection and topology evolution identification framework to firstly detect the newly emerging online topic and secondly to uncover its evolution topology on the global heterogeneous online social networks.

3. Problem Formulation and Method

3.1. KeyGraph Network Modeling. Before introducing our proposed dynamic community detection method, we firstly build up the KeyGraph network for short texts crawled from heterogeneous social networks platforms in two steps. Firstly, every short text is modeled as a node/vertex of the key graph network. Connections between any two short texts are modeled as the edge state between them. Secondly, we acquire the keywords set corresponding to each short text by using the word segmentation technology. Thus the original short-text network can be abbreviated as a complex network based on the closeness of keywords.

In this paper, we denote and name it as KeyGraph $G = \{V, E\}$ in the following way, where i, j represent the i th and j th short texts crawled from heterogeneous social networks and marked with a number, C_i is the keyword set of the i th short text using word segmentation technology, N_{ij} is the count number of common keywords belonging to keyword sets of both C_i and C_j ; V_i is the i th node of the network, E_{ij} represents the edge between the i th and j th short texts which is closely related to the common keywords number N_{ij} . The relationship of E_{ij} with N_{ij} is shown in the following formula:

$$\begin{aligned} E_{ij} &= 1, & \text{if } N_{ij} > 0, \\ E_{ij} &= 0, & \text{if } N_{ij} = 0, w_{ij} = N_{ij}. \end{aligned} \quad (1)$$

For illustration purposes, 406 short texts containing both news titles and microblogs are crawled in which the number of people participating overcomes 1000 in October 1st, the year 2019. Its KeyGraph in the stochastic and Fruchterman Reingold distributions is shown in Figure 1. Figure 1(a) is the

random distribution of the KeyGraph, and Figure 1(b) is the Fruchterman Reingold distribution of the KeyGraph. It shows a clear community structure.

3.2. Dynamic Community Detection and Topic Detection on the Time-Evolving KeyGraph Network. Different from the static network, it should be noticed that the network formed by short texts crawled from heterogeneous online social network actually evolves with time flying by in this paper, and so do the relationships of the network. Thus the Key-graph network modeled in Section 3.1 is actually the time-evolving and scalable network. In this paper, we denote it as the time-evolving and scalable network $G_t = \{V_t, E_t\}$, the scale of which increases in sizes of either node V_t or edges E_t or even both of them with time flying by.

In this paper, we propose a dynamic community detection method which not only can effectively alleviate the resolution limit problem, but also can discover the community structure of the time-evolving and scalable network G_t . Its main idea is that at given time epoch t , the community structure of network at time epoch $t - \Delta t$, i.e., $G_{t-\Delta t}$ is assumed as clearly detected using static Louvain algorithm and already known, the part of network changing/variation during the time interval $(t, t + \Delta t]$ rather than the entire network, i.e., $G_{t+\Delta t}$ at time $t + \Delta t$ is our main focus. By calculating the closeness of local changed subnetwork during the time interval $(t, t + \Delta t]$ with historical communities of network G_t , a local bipartite graph is obtained. The local bipartite graph is composed of two groups of nodes. One is the group of nodes having loose closeness with communities of G_t and is denoted as $Bi_G_{\text{loose}, t, t+\Delta t}$, and the other is the group of nodes having close connections with communities of G_t denoted as $Bi_G_{\text{close}, t, t+\Delta t}$. By applying the static community detection, i.e., Louvain algorithm on subnetwork is composed of both the $Bi_G_{\text{close}, t, t+\Delta t}$ and historical communities of G_t ; also by applying the Louvain algorithm on $Bi_G_{\text{loose}, t, t+\Delta t}$, the whole new emerging community loosely connected with historical communities of G_t will be detected. The community structure of time-evolving network at time epoch $t + \Delta t$, i.e., $G_{t+\Delta t} = \{V_{t+\Delta t}, E_{t+\Delta t}\}$ is discovered by combining community detection results on these bipartite graphs during the time range $(t, t + \Delta t]$. By simply utilizing the local changed network property, the complexity and running time only depend on the local changed part of network rather than the whole network at $t + \Delta t$, i.e., $G_{t+\Delta t} = \{V_{t+\Delta t}, E_{t+\Delta t}\}$, which enables its applications in the large-scale network. The flow chart of our proposed dynamic community detection method is presented in Figure 2.

Before explicitly unfolding the specific dynamic community detection method, we present some related and important definitions closely related to our algorithm in advance as follows.

3.2.1. Related Definitions

Definition 1. Closeness degree of node i with the network at time epoch t , i.e., G_t is defined as $k_{i,t}$ and is calculated using the following formula:

FIGURE 1: Illustration of KeyGraph. (a) Random Distribution. (b) The Fruchterman Reingold distribution.

$$k_{i,t} = \sum_{j_t \in G_t} A_{ij_t}, \quad (2)$$

where A_{ij_t} is the adjacent matrix of the network $G_{t+\Delta t}$, j_t means node j belonging to network G_t , and node i belonging to network $G_{t+\Delta t}$.

Definition 2. Closeness of node i belonging to the local changed network of time range $(t, t + \Delta t]$ with networks G_t and $G_{t+\Delta t}$ is defined as r and is calculated using the following formula:

$$r = k_{i,t} - k_{i,t+\Delta t}, \quad (3)$$

if $r > 0$, it is believed that the node i during time range $(t, t + \Delta t]$ have a close relationship with the historical communities of network G_t compared with network $G_{t+\Delta t}$, if $r \leq 0$, then it is believed that node i during time range $(t, t + \Delta t]$ have a loose relationship with the historical communities of G_t compared with network $G_{t+\Delta t}$.

Definition 3. The modularity model proposed by Newman and Girvan is presented in the following formula:

$$Q = \frac{1}{2m} \sum_{i,j \in \Omega} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta_{ij}, \quad (4)$$

where A_{ij} is the adjacent matrix, δ_{ij} is the Kronecker function, $\delta_{ij} = 1$ when both nodes i and j are in the same community; otherwise, $\delta_{ij} = 0$. Ω is the set of total network nodes, k_i, k_j is the degree of nodes i and j within the whole network, m is the total weights of all edges of the whole network.

By rewriting the Kronecker function δ_{ij} , the modularity function Q can be rewritten in the following formula:

$$\begin{aligned} Q &= \frac{1}{2m} \left(\sum_{i,j \in \Omega} A_{ij} - \frac{\sum_{i \in \Omega} k_i \sum_{j \in \Omega} k_j}{2m} \right) \\ &= \frac{1}{2m} \sum C \left(\sum \text{in} - \frac{(\sum \text{tot})^2}{2m} \right), \end{aligned} \quad (5)$$

where C represents any community of network, $\sum \text{in}$ represents the total edge weights within community C , $\sum \text{tot}$

represents the summation of total edge weights connected with community C .

In our paper, a modularity gain index is defined as the modularity difference between the modularity value before and after reassigning node i into the community where node j belonging to the modularity gain is calculated using formulae (4) and (6).

$$\begin{aligned} \Delta Q_{i,j} &= \left[\frac{\sum \text{in} + k_{i,\text{in}}}{2m} - \left(\frac{\sum \text{total} + k_i}{2m} \right)^2 \right] \\ &\quad - \left[\frac{\sum \text{in}}{2m} - \left(\frac{\sum \text{total}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] = \frac{k_{i,\text{in}}}{2m} - \frac{k_i \sum \text{total}}{2m^2}. \end{aligned} \quad (6)$$

3.2.2. Dynamic Community Detection Method. After presenting the essential and necessary definitions and formulations, we give the specific framework of our dynamic community detection method for the time-evolving network as follows.

Firstly, for time-evolving network G_t formed by short texts and its relationships crawled from heterogeneous social network platforms before time epoch t , the static community detection algorithm here referred to as Louvain algorithm is utilized to obtain the community structure of G_t . Secondly, by bisecting the local varied network during the time range $[t, t + \Delta t]$ into two groups as a bipartite graph; one is the subnetwork denoted as $Bi_G_{\text{close},t,t+\Delta t}$ composed of the new emerging nodes that have close relationship/connections with the historical communities of G_t ; the other is the subnetwork denoted as $Bi_G_{\text{loose},t,t+\Delta t}$ composed of the new emerging nodes that have loose relationship with the historical communities of network G_t , and we propose formula (3) to quantify the closeness of the local varied nodes within time range $[t, t + \Delta t]$ with historical communities of network G_t . Finally, by applying the static community detection method into the local varied networks $Bi_G_{\text{loose},t,t+\Delta t}$ and $Bi_G_{\text{close},t,t+\Delta t}$, we can identify the community subordinate attributes of these new emerging nodes within time range $[t, t + \Delta t]$, i.e., which nodes should belong to the historical communities of network G_t as the new emerging nodes within historical communities during time range $[t, t + \Delta t]$,

FIGURE 2: Flow chart of the community detection method based on the Louvain algorithm.

and which nodes should be assumed as the whole new emerging communities during time range $[t, t + \Delta t]$, respectively.

As mentioned above, there is a static community detection algorithm embedded and applied when we identify the community structure of bipartite networks $Bi_G_{close,t,t+\Delta t}$

and $Bi_G_{loose,t,t+\Delta t}$; here the Louvain algorithm is chosen as the embedded static community detection method. It is well known that a high value of modularity indicates a good community partition of the target network. Maximizing this criterion by utilizing all kinds of optimization algorithms has always been a popular research focus during the past

decades. However, it is intractable to find the exact global optimal modularity. Thus, many approximation optimization algorithms have been proposed. Among these algorithms, the greedy concept introduced by Blondel et al. [18] and called as Louvain algorithm has been proved to be among the most efficient algorithm with excellent performance, especially in large-scale networks. Louvain algorithm is actually the hierarchical cluster community detection method and mainly consists of two steps. In the first step, modularity is optimized locally in the neighborhood of each node; in the second step, it aggregates the nodes in the same community into supernodes and thus forms a new coarse-grained aggregated network. These two procedures are iteratively performed until the global value of network modularity stops to increase by any movement of nodes in the network, specific Louvain algorithm is presented as follows.

Step 1: treating each node of the targeted network as every single community

Step 2: for node i and its neighbor node j , we calculate the modularity gain $\Delta Q_{i,j}$ and its maximum value $\max \Delta Q_{i,j}$, if $\max \Delta Q_{i,j} > 0$, then we deem that node i and node j should belong to the same community

Step 3: repeating step 2 for all the node i and its neighbor node j until there is no community change for all the nodes of the network

Step 4: compressing the network with the community as an aggregated node, the degree of the aggregated node is the original degree of the corresponding community

Step 5: repeating step 1 to step 4 for the compressed network until the modularity gain $\Delta Q_{i,j}$ of the whole network does not increase, the algorithm stops

The modularity gain $\Delta Q_{i,j}$ in step 2 after node i joining into the communities of its neighbor node j is calculated using formula (6).

$$\Delta Q_{i,j} = \left[\frac{\sum \text{in} + k_{i,\text{in}}}{2m} - \left(\frac{\sum \text{total} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum \text{in}}{2m} - \left(\frac{\sum \text{total}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] = \frac{k_{i,\text{in}}}{2m} - \frac{k_i \sum \text{total}}{2m^2}. \quad (7)$$

Thus, we have fully presented the dynamic community detection algorithm with the Louvain method embedded for the time-evolving network $G_t = \{V_t, E_t\}$.

Its main algorithm is presented as follows:

Step 1: Utilizing Louvain algorithm to obtain the community structure of network G_t

Step 2: bisecting the local varied network during time range $[t, t + \Delta t]$ into two groups as a bipartite graph using formula (3), denoted as $Bi_G_{\text{close},t,t+\Delta t}$ and $Bi_G_{\text{loose},t,t+\Delta t}$;

Step 3: applying the Louvain method for networks $Bi_G_{\text{close},t,t+\Delta t} + G_t$ and $Bi_G_{\text{loose},t,t+\Delta t}$, and combining

the community detection results as the final community structure of network $G_{t+\Delta t}$.

3.2.3. Topic Detection Method. A more detailed process of our new emerging topic detection method is presented as follows. We first construct the named KeyGraph network using keywords of short texts crawled from heterogeneous social network platforms according to modeling rule shown in Section 3.1; secondly, by employing the time-evolving property of the constructed KeyGraph, we utilize the proposed dynamic community detection method to identify the community structure of time-evolving KeyGraph network; thirdly, for each detected community of KeyGraph, we calculate the total number of people participating in (i.e., reviewing, thumbing-up, retweeting) each detected community, which actually reflects the keywords of the original short texts belonging to the detected community; finally, we rank and select the top- N detected communities according to the sequence of the statistic value. According to the selected top- N communities, the highest frequently mentioned keywords of the detected community are chosen as the keywords of the newly detected emerging topic.

Until now, our topic detection method based on the dynamic community detection method has been fully uncovered, and its specific algorithm is presented as follows:

Step 1: the KeyGraph model is used to map the original short-text network into the KeyGraph network, thus the data set of the short texts before time epoch t is changing into the KeyGraph network G_t

Step 2: identifying the community structure of network G_t using Louvain algorithm

Step 3: adding the new emerging short-text data during time range $[t, t + \Delta t]$ into the network G_t and forming the new network at time epoch $t + \Delta t$, i.e., $G_{t+\Delta t}$

Step 4: identifying the community structure of network $G_{t+\Delta t}$ using the dynamic community detection method

Step 5: repeating step 3 and step 4 until the community structure of the short-text formed network $G_{t+\Delta t}$ has been fully identified

Step 6: calculate the total number of people participating in each detected community belonging to network $G_{t+\Delta t}$

Step 7: selecting the top- N communities which have the most statistical total number of people participating

Step 8: for the selected top- N communities, calculating the frequency of keywords subordinating to each community, chose the top- n keywords with the highest frequency as the keywords of the corresponding topic

3.3. Method Alleviating Resolution Limit Problem. Another advantage needs to be pointed out is that our proposed dynamic community detection method can effectively alleviate the resolution limit problem commonly encountered in modularity-based community detection methods of complex network. For illustration purposes,

Figure 3 is presented to show how our dynamic community detection method can alleviate the resolution limit problem by adaptively presetting the discrete time step $\Delta t_{0,i}, i = 1, 2, \dots$ in the proposed dynamic community detection method.

As presented in Figure 3, at the initial time epoch t_0 , we set the first discrete time step $\Delta t_{0,1}$, during the time range $[t_0, t_0 + \Delta t_{0,1}]$, the local changed networks are composed of nodes marked in orange circle and are denoted as community 1, light blue circles are denoted as community 2, and cyan circles are denoted as community 3; at time epoch $t_0 + \Delta t_{0,1}$, we adaptively set the second discrete time step as $\Delta t_{0,2}$, during time range $[t_0 + \Delta t_{0,1}, t_0 + \sum_{i=1}^2 \Delta t_{0,i}]$, the corresponding local changed networks are composed of nodes marked as cyan circle, pink circle, and light blue circle, which are denoted as community 3, community 4, and community 2, respectively, where there are new emerging vertices in community 2 and community 3 compared with those communities at the previous time epoch $t_0 + \Delta t_{0,1}$, and community 4 is the whole new emerging community during time range $[t_0 + \Delta t_{0,1}, t_0 + \sum_{i=1}^2 \Delta t_{0,i}]$; also at time epoch $t_0 + \sum_{i=1}^2 \Delta t_{0,i}$, we set the time step $\Delta t_{0,3}$, during time range $[t_0 + \sum_{i=1}^2 \Delta t_{0,i}, t_0 + \sum_{i=1}^3 \Delta t_{0,i}]$, the local changed/varied networks are composed of nodes marked as pink circle, dark blue circle, where the scale of community 4 enlarges with new emerging vertices and in the mean time the whole new emerging community 5 is detected. Following this schema, this process continues in the nested and recursive way with time flying by. Thus, for the time-evolving and scalable network, by selecting the proper time step $\Delta t_{0,i}$, the resolution limit problem can be effectively alleviated.

3.4. Topic Evolution Law Identification. In this section, we focus on how to discover the community topology evolution of the detected topic in Section 3.2.2. It should be noticed that the constructed KeyGraph network is actually a directed graph under the topic propagation situation instead of simply topic detection scenario because the topic propagation commonly reflects information propagation direction during the topic diffuses process, while our topic detection method proposed in Section 3.2.2 is illustrated using the time-evolving, undirected KeyGraph network. Thus, in this section, we should first expand the modularity formula (4) for undirected networks into the directed network as shown in the formula.

$$Q = \frac{1}{n} \sum \left(a_{ij} - \frac{k_i^{\text{in}} k_j^{\text{out}}}{n} \right) \delta_{ij}, \quad (8)$$

where n is the total number of edges, a_{ij} is the i, j elements values of adjacent matrix of the directed network, k_i^{in} is the in-degree of node i , k_j^{out} is the out-degree of node j . δ_{ij} is the Kronecker function as defined before.

Also, we expand the modularity gain ΔQ for undirected graph into the directed graph and present it in the following formula:

$$\Delta Q = \frac{k_{i,\text{in}}}{n} - \frac{k_i \sum \text{tot}}{n^2}, \quad (9)$$

where k_i is the degree of node i , and $k_i = k_i^{\text{in}} + k_i^{\text{out}}$, $k_{i,\text{in}}$ represents a new number of edges connecting the local changed nodes with historical communities, $\sum \text{tot}$ represents the total number of edges with community C .

Hence, by substituting formulae (8) and (9) into our dynamic community detection method in Section 3.2.2, we can propose the algorithm for topic evolution topology identification purposes.

Step 1: modeling the users at time epoch t participating in the topic identified Section 3.2 as the directed topic propagation network G_t ;

Step 2: identifying the community structure of the directed network G_t by utilizing our proposed dynamic community detection method;

Step 3: adding the local changed users participating into the directed topic propagation network G_t during time range $[t, t + \Delta t]$ and forming the scalable network $G_{t+\Delta t}$;

Step 4: separating the local changed users during time range $[t, t + \Delta t]$ into a bipartite graph, the one having a loose relationship with network G_t , denoted as $Bi-G_{\text{loose},t,t+\Delta t}$, and the other having close relationship with network G_t , denoted as $Bi-G_{\text{close},t,t+\Delta t}$;

Step 5: for subnetwork composed of G_t and $Bi-G_{\text{close},t,t+\Delta t}$, employing the proposed dynamic community detection algorithm, the incremental information of the historical communities of G_t during time range $[t, t + \Delta t]$ is found.

Step 6: for subnetwork composed of $Bi-G_{\text{loose},t,t+\Delta t}$, the whole new emerging community is identified by using the proposed dynamic community detection algorithm;

Step 7: merging the community detection results of steps 5 and 6, the community detection results of the topic propagation network at time epoch $t + \Delta t$ is identified;

4. Experiments and Results

To validate the effectiveness of our proposed topic detection and evolution law identification method, we use both the artificial complex networks composed of computer-generated data and the real network constructed by original data crawled from heterogeneous and popular social media platforms. By comparison of the community detection results under the static Louvain algorithm and the proposed dynamic community detection method with the Louvain algorithm embedded, our proposed dynamic community detection method yields better results which validate its effectiveness and feasibility.

4.1. Experiments on Artificial Computer-Generated Network.

In the artificial complex network composed of computer-generated data, we generate the artificial complex networks by choosing the nodes' connection probability p within the same community, while the nodes connection probability between the communities is set as $1-p$, and values p and $1-p$ satisfying $p > 1 - p$, which means that the closeness of nodes

FIGURE 3: Illustration of dynamic community detection method to alleviate resolution limit problem.

within the community is larger than nodes between communities. Here, the artificial computer-generated network, which contains 68 nodes in total, with p set as 0.78. Community detection results using the static Louvain algorithm are presented in Figure 4(a).

While considering the time-evolving property in our practical situations, properties of nodes and edges of the artificial computer-generated network remains unchanged. While we randomly chose part of the nodes with time epoch t marked, those randomly selected nodes and edges at time epoch t compose the network G_t , and the remaining nodes are treated as the new emerging nodes of the network during time range $[t, t + \Delta t]$, the community detection results using our dynamic community detection methods are shown in Figure 4(b).

Figure 4 shows community detection results of the artificial computer-generated network with 68 nodes in total and edges randomly connected with parameter $p = 0.78$ described (a) using the static Louvain community detection algorithm and (b) using the proposed dynamic detection method with Louvain community detection algorithm, respectively. Nodes with the same color represent that they belong to the same community using corresponding community detection algorithms. The blue lines paralleling to the x -axes are the separating lines of different communities. The blue lines paralleling to the y -axes are the separating lines of time; as can be seen from Figure 4(b), the three blue lines paralleling to the x -axes separate the three different communities.

In order to verify the efficiency of the proposed dynamic community detection method with the Louvain algorithm embedded, we select five groups of artificial computer-generated data and use both the static Louvain algorithm and the dynamic community discovery method with the Louvain algorithm embedded to detect the communities structure of the computer-generated artificial network. Under the same operating environment, the time efficiencies

of these two algorithms are compared and are shown in Figure 5.

It can be found from Figure 5 that the running time of the static Louvain algorithm is basically the same as those of the dynamic community discovery method with Louvain algorithm embedded when the network scale is relatively small (the number of vertices and edges of the network is relatively small), but with the fast growth of the network scale, the running time of our dynamic community detection method with Louvain algorithm embedded is far less than that of simple static Louvain community detection algorithm.

4.2. Experiments of Topic Detection on Real Short-Text Data Crawled From Real Social Networks Platforms. In this section, we randomly choose 86 short texts from October 1, 2019, to October 3, 2019, from heterogeneous online social networks for manual annotation. Later, the results of manual annotation are used to verify results validation of our proposed dynamic community detection method. Out of the 86 short texts, 46 short texts are crawled from Sina Weibo social platform, 17 short texts are crawled from Sina News site social platform, 15 short texts are crawled from Sohu News site social platform, and 8 short texts are crawled from the Fenghuang News site social platform. In the results of manual annotations, the total 86 short texts are divided into 11 communities, out of which the largest community contains 28 pieces of short texts, and the smallest community contains 2 pieces of short texts.

According to the modeling rule of the KeyGraph network, the above-mentioned manual annotation short-text data are transformed into an undirected KeyGraph network, which has 86 nodes and 530 edges, among which 244 edges with a weight of 1 account for 46.04% of the total number of edges of the whole network. The weight range of the KeyGraph network is 0, 1, 2.

FIGURE 4: Community detection results of computer-generated complex network.

FIGURE 5: Performance comparison of the static Louvain algorithm and dynamic community detection method.

As we know that the weight of edge has a great effect on the final community detection results, and by applying the proposed dynamic community detection method with the Louvain algorithm embedded under different edge weights, the results are compared in Table 1. The detection ratio is computed using the following formula:

$$\text{Ratio} = \frac{n - s}{n}, \quad (10)$$

where n is the total number of manual annotation short texts, and s is the total number of misdetected communities.

From Table 1, it can be found that the community detection ratio of manual annotated short texts is higher, which means the community detection result is more accurate when the weight threshold value is set as 1 other than 0 and 2. Thus, in the topic identification experiments, we choose the edge weight threshold as 1.

Except for the pilot experiment to choose the optimal edge weight threshold, next, we will use the real short-text data crawled from the Sina Weibo, Sina News site, Sohu News site, and Fenghuang News site, the mainstream news publishing sites and microblog platforms which are popular. We media means citizen Journalism, the mainstream news publishing sites and microblog platforms which are popular citizen Journalism nowadays in China. A total of 262246 pieces of short-text data from October 1, 2019, to October 3, 2019, were extracted from the above listed heterogeneous online social networks as the real experimental data set.

According to the time stamp, the total number of crawled short texts on October 1 is 85980 pieces, the total number of crawled short texts on October 2 is 86768 pieces, and the total number of crawled short texts on October 3 is 89498 pieces. Chinese word segmentation and keywords extraction are performed on these crawled original experimental short-text

TABLE 1: Community detection ratio under different edge weights.

Weight	Community detected	Misdetected community	Detection ratio
0	10	3	0.965
1	12	1	0.988
2	24	22	0.744

data, and five keywords are selected to represent the original short texts for contents of news titles and microblogs.

Here, the time interval is set as 1 day (24 hours) and the edge weight threshold is set as 1 when we use the dynamic community detection method with the Louvain algorithm embedded. Firstly, we construct a network G_1 using the short-text data of October 1, and the network G_1 has 85980 nodes and 420800 edges based on the model definition of the KeyGraph network. Applying the proposed dynamic community detection method with Louvain algorithm embedded on KeyGraph network G_1 . Secondly, adding the original crawled short-text data of October 2, then the network is denoted as KeyGraph network G_2 , then the KeyGraph network G_2 has 172748 nodes and 1506583 edges, applying the proposed dynamic community detection method with Louvain algorithm embedded on KeyGraph network G_2 ; thirdly, adding the crawled original short-text data of October 3 into the KeyGraph network G_2 , then the newly varied KeyGraph network is denoted as G_3 , KeyGraph network G_3 has 262246 nodes and 3250235 edges. We apply the proposed dynamic community detection method with Louvain algorithm embedded on KeyGraph network G_3 . The number of nodes corresponding to each community is shown in Figure 6, in which the abscissa represents the total detected community number and the ordinate represents the number of nodes belonging to each corresponding community.

The modularity of the KeyGraph network G_3 is 0.886, and 222133 communities are discovered by using the dynamic community detection method with the Louvain algorithm embedded; nodes contained in each corresponding community are shown in Figure 6. The abscissa represents the total detected number of communities, and the ordinate represents the number of nodes in each community.

From Figure 6, it can be found that almost all node sizes of the community are less than 250 and the proportion of community sizes less than 100 is larger than 99%. From the community discovery results, we can find the sparsity of the information distribution of the online social networks of different sources (222133 communities are found out of 262246 short texts), and the scale of the communities is generally small.

According to our proposed dynamic community detection method with Louvain algorithm embedded, we then calculate the total number of people participating in (replying, retweeting, thumbs-up) each short text (corresponding to each community); we rank it and choose the community in which the total number of people participants are larger than 100000; then we calculate the frequency of the keywords of each community,

rank them, and obtain the top 5 keywords as the representative keywords of our detected topic, as shown in Table 2.

As shown in Table 2, the top 5 keywords of the 5 short-text communities are presented in the first column, the selected keywords are assumed as the keywords of topic detected as shown in the second column, and the total number of people participating in the short texts in ways either discussing, retweeting, or thumbs-up is shown in the third column.

It can be seen from Table 2, the top 5 keywords baring the highest appearing frequency in the short-text community are “Men’s basketball, Asian championships, Iran, Chinese team, Asia;” they form the detected topic as “China Men’s basketball Asian Championship;” the total number of people participating are the highest, about 391238 online users from heterogeneous social network platforms.

4.3. Validation of Topic Evolution Law. In this section, the detected topic “China men’s basketball Asian Champion,” which has the highest number of people participating, is chosen to validate our topic topology evolution identification purpose. 368 pieces of related news and microblogs are obtained, in which the participating people are 141318 in summation, out of which there are 1324 people on October 1, 8045 people on October 2, and 3872 people on October 3 in the year 2019.

After modeling the related data of the detected topic “China men’s basketball Asian Champion” into the directed topic propagation network as the definition is shown in Section 3.4, there are 14318 nodes and 14962 edges in total. The community detection result of our dynamic community detection method on the directed topic propagation network is shown in Figure 7.

336 communities are detected and the modularity value is 0.514 by setting the time step as one day (24 hours). The dynamic properties of the topic propagation network is shown in Table 3. It can be found that until October 3, about 92.48% of users have already participated in the topic while about 56.19% users are added just on October 2 and about 258 on October 2 are new emerging communities.

From Figure 8, we can see that the scale of the node belonging to the community in each day corresponding to the detected topic “China Men’s Basketball Asian Champion” reaches its peak at October 2 and gradually recedes to zero until October 6. The detected topic topology evolution is changing from October 1 as zero, reaches its peak value on October 2, and gradually recedes to zero from October 3 to October 6.

FIGURE 6: Size distribution of communities detected in the illustrated example.

TABLE 2: Topic detection results.

Top-5 keywords	Topic detected	People participating
Men's basketball, Asian championships, Iran, Chinese team, Asia,	China men's basketball Asian championship	391238
Liucheng county, Booming, Guangxi, package, express	The express package booms in Liucheng county, Guangxi	240673
Jewelry store, servant, Hongkong, customer, the worst	Hongkong jewelry store's servants provide the worst customer	195265
Commit adultery, condom, office, Hunan, Leiyang	Officer of Leiyang, Hunan commit adultery	167413
Innocent, Chinese, police, shoot, postgraduate	Japanese police shoot Chinese and is proved to be innocent	130957
National flag, broken, hung, county	A county in Jiangxi hang the broken national flag	123441

FIGURE 7: Community detection results of topic "China men's basketball Asian Champion."

TABLE 3: Properties of topic propagation network.

Day	Nodes	Edges	Value of modularity	Community detected
2019-10-01	1324	1451	0.558	12
2019-10-02	9369	9882	0.388	78
2019-10-03	13241	13881	0.457	336
2019-10-04	14318	14962	0.514	336

FIGURE 8: Variation of nodes in the detected topic “China men’s Basketball Asian Champion.”

5. Conclusions and Prospects

In this paper, we propose a topic detection and topology evolution identification framework based on the dynamic community detection method. Firstly, a unified time-evolving KeyGraph network is constructed based on the cooccurrence of keywords of short texts crawled from heterogeneous online social network platforms. Secondly, a dynamic community detection method for time-evolving network is proposed and the topic is detected by its utilization on the constructed KeyGraph network. Thirdly, for the detected topic in the previous step, a directional topic propagation network is built based on the short texts related to the detected topic, and the topic evolution topology is mainly reflected as the nodes scale of community is discovered.

Data Availability

The data used to support the finds of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] China Internet Network Information Center, *The 45th Development Statistic Report of Internet of China*, China Internet Network Information Center, Beijing, China, 2020, <http://www.cnnic.cn/>.
- [2] M. Girvan and M. E. Newman, “Community structure in social and biological networks,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 1, pp. 7821–7826, 2002.
- [3] E. M. Jin, M. Girvan, and M. E. J. Newman, “Structure of growing social networks,” *Physical Review E-Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 64, Article ID 046132, 2001.
- [4] A. Lancichinetti, M. Kivela, J. Saramaki, and S. Fortunato, “Characterizing the community structure of complex networks,” *PLoS One*, vol. 5, no. 8, Article ID e11976, 2010.
- [5] J. Deng, K. Deng, Y. Li, and Y. Li, “Hot topic detection based on complex networks,” in *Proceedings of the 2013 International Conference on Fuzzy Systems & Knowledge Discovery*, IEEE, Shenyang, China, July 2013.
- [6] W. Lin and D. Guan-Zhong, “Forum hot topic detection based on community structure of complex networks,” *Computer Engineering*, vol. 34, no. 11, pp. 214–217, 2008.
- [7] Z. Liu and B. Hu, “Epidemic spreading in community networks,” *Europhysics Letters*, vol. 72, no. 2, pp. 315–321, 2005.
- [8] G. Ren and X. Wang, “Epidemic spreading in time-varying community networks,” *Chaos An Interdisciplinary Journal of Nonlinear Science*, vol. 24, no. 2, p. 068701, 2014.
- [9] X. Wang, L. Tang, H. Liu, and L. Wang, “Learning with multi-resolution overlapping communities,” *Knowledge and Information Systems*, vol. 36, no. 2, pp. 517–535, 2013.
- [10] L. Tang, X. Wang, and H. Liu, “Community detection via heterogeneous interaction analysis,” *Data Mining and Knowledge Discovery*, vol. 25, no. 1, pp. 1–33, 2012.
- [11] M. Reville, C. Domeniconi, M. Sweeney et al., “Finding community topics and membership in graphs,” in *Joint European Conference on Machine Learning & Knowledge Discovery in Databases*, Springer International Publishing, New York, NY, USA, 2015.
- [12] Y. Hu, B. Yang, and C. Lv, “A local dynamic method for tracking communities and their evolution in dynamic networks,” *Knowledge-Based Systems*, vol. 110, no. 1, pp. 176–190, 2016.
- [13] R. Aktunc, I. H. Toroslu, M. Ozer, and H. Davulco, “A dynamic modularity based community detection algorithm for large-scale networks: DSLM,” in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Paris, France, August 2015.
- [14] Y. Sun, Y. Yu, and J. Han, “Ranking-based clustering of heterogeneous information networks with star network schema,” in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, Paris, France, June 2009.
- [15] M. D. Hoffman, D. M. Blei, and F. R. Bach, “Online learning for latent dirichlet allocation,” in *Proceedings of the 23rd International Conference on Neural Information Processing Systems*, Curran Associates Inc., Vancouver, British Columbia, Canada, December 2010.
- [16] Y. Akita, Y. Nemoto, and T. Kawahara, “PLSA-based topic detection in meetings for adaptation of lexicon and language

- model,” in *Proceedings of the Interspeech, Conference of the International Speech Communication Association*, Antwerp, Belgium, August 2007.
- [17] H. Sayyadi and L. Raschid, “A graph analytical approach for topic detection,” *ACM Transactions on Internet Technology*, vol. 13, no. 2, pp. 1–23, 2013.
 - [18] V. D. Blonde, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 10, no. 1, p. 10008, 2008.
 - [19] R. Jin, C. Kou, and R. Liu, “Improving community detection in time-evolving networks through clustering fusion,” *Cybernetics and Information Technologies*, vol. 15, no. 2, pp. 63–74, 2015.
 - [20] T. Murata and S. Moriyasu, “Blog community discovery and evolution based on mutual awareness expansion,” in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, Fremont, CA, USA, November 2007.
 - [21] A. Cuzzocrea, F. Folino, and C. Pizzuti, “Dynamicnet: an effective and efficient algorithm for supporting community evolution detection in time-evolving information networks,” in *Proceedings of the 17th International Database Engineering and Applications Symposium (IDEAS '13)*, Barcelona, Spain, October 2013.
 - [22] Y. Lin, C. Yun, S. Zhu, H. Sundaram, and B. L. Tseng, “Analyzing communities and their evolutions in dynamic social networks,” *ACM Transactions on Knowledge Discovery from Data*, vol. 3, no. 2, 2008.
 - [23] F. Folino and C. Pizzuti, “An evolutionary multiobjective approach for community discovery in dynamic networks,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1838–1852, 2014.
 - [24] L. Tang, H. Liu, J. Zhang et al., “Community evolution in dynamic multimode networks,” in *Proceedings of the 14th Acm Sigkdd International Conference on Knowledge Discovery & Data Mining*, Las Vegas, NV, USA, August 2008.
 - [25] M. Toyoda and M. Kitsuregawa, “Extracting evolution of web communities from a series of web archives,” in *Proceedings of the 14th ACM Conference on Hypertext and Hypermedia*, pp. 28–37, Nottingham, UK, August 2003.
 - [26] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, “Uncovering the overlapping community structure of complex networks in nature and society,” *Nature*, vol. 435, no. 3, pp. 814–818, 2005.
 - [27] D. Chakrabarti, R. Kumar, and A. Tomkins, “Evolutionary clustering,” in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 554–560, Philadelphia, PA, USA, August 2006.
 - [28] B. Yang and D.-Y. Liu, “Force-based incremental algorithm for mining community structure in dynamic network,” *Journal of Computer Science and Technology*, vol. 21, no. 3, pp. 393–400, 2006.
 - [29] H. Ning, W. Xu, Y. Chi, Y. Gong, and T. S. Huang, “Incremental spectral clustering by efficiently updating the eigensystem,” *Pattern Recognition*, vol. 43, no. 1, pp. 113–127, 2010.
 - [30] C. Dhanjal, R. Gaudel, and S. Cléménçon, “Incremental spectral clustering with the normalized laplacian,” in *Proceedings of the 3rd NIPS Workshop on Discrete Optimization in Machine Learning*, pp. 1–6, Granada, Spain, December 2011.
 - [31] D. Duan, Y. Li, R. Li, and Z. Lu, “Incremental k-clique clustering in dynamic social networks,” *Artificial Intelligence Review*, vol. 38, no. 2, pp. 129–147, 2011.
 - [32] T. N. Dinh, N. P. Nguyen, and M. T. Thai, “An adaptive approximation algorithm for community detection in dynamic scale-free networks,” in *Proceedings of the IEEE International Conference on Computer Communications*, pp. 55–59, Turin, Italy, April 2013.
 - [33] T. Falkowski, A. Barth, and M. Spiliopoulou, “Studying community dynamics with an incremental graph mining algorithm,” in *Proceedings of the 14th Americas Conference on Information Systems*, p. 29, Toronto, Canada, August 2008.
 - [34] Q. Zhao, S. Bhowmick, X. Zheng, and Y. Kai, “Characterizing and predicting community members from evolutionary and heterogeneous networks,” in *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pp. 309–318, Napa Valley, CA, USA, October 2008.
 - [35] Y. Sun, J. Tang, J. Han, M. Gupta, and B. Zhao, “Community evolution detection in dynamic heterogeneous information networks,” in *Proceedings of the 8th Workshop on Mining and Learning with Graphs (MLG '10)*, Washington, D.C., USA, July 2010.
 - [36] J. Wu, Y. Lianfei, Z. Qun et al., “Multityped community discovery in time-evolving heterogeneous information networks based on tensor decomposition,” *Complexity*, vol. 2018, Article ID 9653404, 16 pages, 2018.
 - [37] L. Tang, X. Wang, and H. Liu, “Uncovering groups via heterogeneous interaction analysis,” in *Proceedings of the 2009 International Conference on Data Mining*, pp. 503–512, Miami, FL, USA, December 2009.
 - [38] C. Yang, Y. Xiao, Y. Zhang, Y. Sun, and J. Han, “Heterogeneous network representation learning: survey, benchmark, evaluation, and beyond,” 2020, <http://arxiv.org/abs/2004.00216>.
 - [39] B. Liu, F. X. Han, D. Niu, L. Kong, K. Lai, and Y. Xu, “Story forest,” *ACM Transactions on Knowledge Discovery from Data*, vol. 14, no. 3, pp. 1–28, 2020.
 - [40] Y. Liu, H. Peng, J. Li et al., “Event detection and evolution in multi-lingual social streams,” *Revue Finance, Contrôle, Stratégie*, vol. 14, no. 5, pp. 1–15, 2020.
 - [41] Y. Cao, H. Peng, J. Wu et al., “Knowledge-preserving incremental social event detection via heterogeneous GNNs,” in *Proceedings of the Web Conference 2021*, Ljubljana, Slovenia, April 2021.