Edge Intelligence-enabled Dynamic Overlapping Community Discovery and Evolution Prediction in Social Media Data Streams

Jun Ge[1,2,4], Lei-lei Shi[1,4], Lu Liu[3], Hongwei Shi[2], John Panneerselvam[3]
[1]School of Computer Science and Telecommunication Engineering, Jiangsu University, China
[2] School of Information Engineering, Suqian University, China
[3]School of Informatics, University of Leicester, UK
[4]Jiangsu Key Laboratory of Security Tech. for Industrial Cyberspace, Jiangsu University, China
Corresponding author: Lu Liu, Email: l.liu@leicester.ac.uk

Abstract: **Edge intelligence (EI) is recognized by academia and industry as one of the key emerging technologies for future Cyber-Physical-Social Systems (CPSS), which provides the ability to analyze data at edge rather than sending it to the cloud for analysis, and will be a key enabler to realize a world of a trillion hyper-connected smart sensing devices. As a part of future Cyber-Physical-Social Systems, online social networks are large-scale complex networks that consist of a large number of network nodes and links. The dynamic discovery of communities, especially overlapping communities, is important to understand the evolution of online social networks. However, traditional community discovery algorithms cannot effectively discover overlapping communities in social networks. In order to address this challenge, an Edge Intelligence-enabled dynamic overlapping community Discovery and Evolution Prediction model (EIDEP) is proposed in this paper. This model encompasses a Label Propagation Algorithm based Extension (LPAE) algorithm, which is able to efficiently discover the user community structures in online social networks. Based on the LPAE community discovery algorithm, a User Interest Behavior based Evolution Prediction algorithm (UIBEP) is incorporated in our EIDEP model in order to realize a fast yet accurate community evolution for online social networks, by considering the interest similarity of unlinked nodes in a given community. The performance of our proposed LPAE and UIBEP models is validated and evaluated against notable state-of-the-art community discovery algorithms, through extensive experiments conducted based on a Twitter dataset.**

Index Terms- Edge Intelligence, Dynamic Community, Community Prediction, Community Evolution, Social Network

I. INTRODUCTION

With the rapid development of the Internet and web services [54, 42], the proliferation of the Internet of things (IoT) and the burgeoning of 5G networks are generating unprecedented volumes of data [1-6]. Edge intelligence (EI) in online social networks is recognized by academia and industry [62] as one of the key emerging technologies for future Cyber-Physical-Social Systems (CPSS). The outreach of CPS across various application domains has led to the deployment of data-centric edge nodes that generate and process considerable volumes of heterogeneous data [7-9]. As a distributed intelligent computing paradigm [63] that computation is largely or completely performed on distributed device nodes, EI-enable online social networks provides the rapid development of artificial intelligence and edge computing resources to support a real-time insight and analysis for applications in CPS, which brings memory, computing power and processing ability closer to the location where it is needed [64], reduces the volumes of data that must be moved, the consequent traffic, and the distance the data must travel. Moreover, as an emerging intelligence computing paradigm, EI-enable online social media can accelerate content deliveries and improve the quality of services and applications, which is attracting more and more relevance from academia and industry because of its advantages in throughput, delay, network scalability and intelligence in CPS. It is also becoming more data-centric, where users are typically connected to one another [10-13], for instance, one follows another in Twitter or one is a friend of another in Facebook. Social network services allow users to receive and spread information via diffusion actions on the connections [59-61], e.g., share, retweet, and reply [14, 52].

Online social networks characterize real-time processing of huge volumes of data, and such data usually comes with stronger semantics due to the wide array of user connectedness and their complex relationship [53]. Such attributes of social media are inviting research from various dimensions in recent years [15-19]. As an extension of human society in a virtual network, online social networks play an increasingly important role in people's daily life [20]. Therefore, online social networks also have community attributes, which is a reflection of the modularity of online social networks [21-25].

At present, there is no unified definition of the community attributes of online social networks. For example, if a social network is modelled as a graph structure, users can be regarded as nodes, user relationships can be regarded as links between nodes, and the community can be viewed as a sub-graph structure. The node links within a given community are usually intensive than those between communities. Such a community structure reflects the real-world social relationship to a certain extent. Different communities often represent different user groups, such as relatives, friends in the same city, fans/followers of celebrities and so on [16]. Users within a given group often have the same interest or attribute characteristics. The growing expansion of online social network scale has put forward further challenges within the context of community discovery. Due to the higher time complexity, classical community discovery algorithms [17-20] are witnessed to fall behind when meeting the performance requirements. Community discovery algorithm [21] works with the local expansion principle, such that the network gradually grows based on local information rather than evolving based on the entire network. With this localized principle,

community discovery algorithms can quickly discover the community structure that can also be suitable for large-scale networks. However, the stability of the community discovery algorithm based on the idea of local extension suffers various issues. With a different choice of initial seed and extension direction, the algorithm may produce varied results.

Community structure is an important feature of online social networks. Discovering a community in a network involves identifying and mapping similar nodes into a set, whereby interaction between nodes in a given set is made stronger than the interaction of nodes located beyond. That is to say, the links between nodes in the same community are denser, and the links between different communities are sparser. In fact, real-world network communities may not be independent of each other. Nodes in the network can belong to two or more communities at the same time, that is, the communities are overlapped, which brings new challenges within the realms of community discovery.

A network with more overlapping nodes characterizes a higher degree of overlapping communities. On the one hand, overlapping nodes characterize a "multi-faceted nature", and the mining of overlapping nodes can more comprehensively discover the characteristics of nodes. Currently, online social networks are all based on users, whereby analyzing user behavior is of primary importance. Characteristics and preferences of users are of great significance to business recommendation and user management. On the other hand, overlapping nodes form the bridge between communities and play a key role in community evolution and information exchange between communities. A higher overlapping degree between communities indicates that most of the members are shared; such overlapping communities can be integrated into a large community. On the contrary, a lower overlapping degree between communities reflects the need for more comprehensive analysis of the community overlapping structure in order to understand the evolution trend of online social networks.

To this end, a simple community clustering algorithm cannot effectively resolve the detection problem of overlapping communities and nodes, especially in front of large networks. Effectively and accurately mining the overlapping communities of social networks has been the focus of current online social network analysis, which is also addressed in this paper.

With this in mind, an edge intelligence-enabled dynamic Overlapping Community Discovery and Evolution Prediction model, named EIDEP model is proposed in this paper. In our proposed model, the hub value of each node is calculated during the initial link selection stage. Secondly, the hub value of nodes at both ends of the network link is taken as the authority value of the corresponding link and further sorted in a descending order. The link with the greatest authority value is added to the network structure as the initial link. This approach avoids the instability caused by the random selection strategy of the Label Propagation Algorithm (LPA) algorithm [5, 9], and further improves the discovery performance and the quality of the generated community.

The main contributions of this paper are as follows:

1) An edge intelligence-enabled overlapping community detection algorithm based on LPA algorithm Extension (LPAE) is proposed. First of all, the hub value of nodes is calculated based on an improved HITS algorithm [1, 2]. The hub value is used as a measure of the importance of nodes, such that a node with more hub value is likely to be located in the community. Secondly, the sum of the hub value of the nodes at both ends of the link is regarded as the link centrality, and the community internal link is selected as the initial seed link to avoid the instability of the algorithm.
2) Based on the results of community discovery with LPAE, this paper proposes a User Interaction Behavior and Subgraph Matching based Evolution Prediction algorithm, named UIBEP, in order to realize fast and accurate community evolution for social networks. Firstly, the network data is filtered and the network matrix in the field of linear algebra is allied to solve the weights of all kinds of interaction behaviors in online social networks. Secondly, the LPAE algorithm is used to divide the network into communities, and the communities are used as the search scope of community evolution, and the interaction similarity of unlinked nodes in the community is calculated. Then the community evolution results of each community are summarized into a set, and the Cosine Similarity method is applied in the formed set for predicting the community evolution for the entire social network.
3) Finally, we conducted experiments to evaluate the performance of our proposed models. The experimental results based on a Twitter dataset demonstrate the efficiency and accuracy of our proposed models in both dynamic community discovery and evolution prediction.

The rest of this paper is organized as follows. In section II, we introduce previous studies of community discovery and evolution. In section III, we present our preliminaries. We describe our proposed LPAE method and introduce the UIBEP model in section IV section V respectively. We discuss our experimental analysis and the obtained results in section VI and in Section VII, we draw our conclusions and future works.

II. RELATED WORK

As an extension of human society in the virtual network, online social network plays an increasingly important role in people's daily life. Therefore, online social networks also characterize a community attribute [26], which is a reflection of the characteristics of social network module. The research on community discovery provides basic theoretical support for community evolution [27]. Further mining the community structure based on local network information is required for understanding the process of evolution in large-scale social networks.

1. Community discovery

Community discovery is being researched from various dimensions in recent years, leading to the development of various community discovery algorithms. The GN algorithm is one of the classical community discovery algorithms, which works based on the splitting idea proposed by Newman [26]. GN algorithm continues to delete links that characterize the maximum number of sides until all links in the network are deleted. The time complexity of GN algorithm is $O(n^3)$. Therefore, the GN algorithm is not suitable for large-scale network structure, but it stands as a pioneer in the field of community discovery. In order to optimize the time complexity of the GN algorithm, Newman et al. [36] then proposed the CNM algorithm based on a clustering idea. First, each node in the network initially forms a community, and then the community is merged in Q increments until the Q value in the whole network cannot be increased further. However, the CNM algorithm suffers from a resolution problem [17], and its neutral performance is poor when large and small communities coexist in a network. Reghavan et al [18] proposed the LPA algorithm for the first time. By passing labels to neighbor nodes, the nodes with the same labels are assigned to the same community. LPA algorithm characterizes a linear time complexity and has a wider utilization. Gregory [19] proposed the copra algorithm to identify overlapping communities, which allows one node to carry multiple labels and membership degrees, and randomly select nodes to update labels. Liu et al [20] proposed an overlapping community discovery algorithm based on the tag propagation probability, where the tag propagation probability is computed based on the network structure information. Liang et al [2] proposed a user interest community detection method in social media, based on collaborative filtering, for user interest community recommendation. Shi et al [21] proposed an event community detection and Multi-source Propagation model for online social network management. Despite the contributions of the aforementioned algorithms, overlapping community discovery has not received enough emphasis.

At present, the mainstream community discovery algorithms in academia can be classified into local expansion optimization and label propagation according to different local optimization strategies [28].

(1) Global optimization community discovery algorithm

The concept of community was first proposed by Newman et al. [26,] and at the same time, the most classical community discovery algorithm, the GN algorithm, was proposed. Based on the idea of splitting, the algorithm defines the number of sides as the shortest path through the link. Due to the modularity of community structure, the number of sides within the community structure is less than the number of sides between the links across communities, so the link with a large number of sides is more likely to be the link between communities. Based on this feature, the GN algorithm continuously deletes the link with the largest side in the network structure, and recalculates the remaining links until all the links in the network are deleted, and the algorithm converges. The whole process of the GN algorithm forms a hierarchical tree, where each level represents a community discovery result, but the algorithm does not tell readers how to select the optimal community discovery result. The GN algorithm should recalculate the number of edge mediations of all the remaining links after deleting the links with the maximum number of edge mediations. The time complexity is $O(n^3)$, which is only suitable for small network structure.

(2) Local extension optimization community discovery algorithm

The idea of local extended optimization community discovery algorithm works based on similarity measurement, which extends the optimization from the initial node to the neighboring nodes. Such a kind of algorithm consists of two core steps: initial seed node and extension direction selection. This strategy only requires the local network information and can quickly discover the community structure, along with exhibiting high performance in large-scale network. However, this kind of method is very strict in the selection of seed nodes, which is easy to generate instability.

Lancichinetti et al [30] put forward the local fitness community discovery algorithm (LFM) in 2009. The algorithm randomly selects the community section as the initial seed node. The LFM algorithm is simple and fast, but the random selection strategy of the seed node directly results in the uneven quality of the seed node. Baumes et al. [31] proposed the Rank Removal community discovery algorithm, named RaRe, which deletes nodes with less influence, and sets the nodes with more influence as initial seed nodes. However, there may be isolated nodes in this seed set, which may lead to unsatisfactory community discovery results after the algorithm is executed. The selection of seed nodes is directly related to the quality of community discovery, herein many researchers have focused on selecting seed nodes accurately. Su et al. [32] put forward a RWA algorithm based on random walk, which uses node intensive sub-graph as the initial community, then selects *K* local maximum nodes as initial nodes, and the maximum nodes and their common neighbors form seed communities are assigned.

(3) Label communication community discovery algorithm

The label propagation algorithm (LPA) [29] was first proposed in 2007. Because of its simplicity and low time complexity, it is widely used. LPA algorithm initially assigns each node with a unique label. In each iteration step, the label with the largest number of neighbors is updated as the label of the node itself. If there are multiple labels with the largest number of neighbors, then one of these randomly selected labels is updated as its own label. Finally, nodes with the same label belonging to the same community are selected. However, the LPA algorithm only allows nodes to carry one label, thus it cannot discover the overlapping community structure. Gregory [33] proposed a multi-label propagation community discovery algorithm, named COPRA, which improved the performance of the LPA algorithm. The COPRA algorithm supports nodes to carry multiple labels, to help identifying the overlapping community structure. However, the COPRA algorithm randomly updates the label strategy, which results in the instability of the algorithm. Thus, in order to address this drawback, Liang et al. [9] proposed an

efficient evolutionary user interest community discovery model in dynamic social networks for the Internet of People, which works based on an improved LPA algorithm. It updates the label according to the hub value of nodes.

2. Community Evolution

The rapid emergence of online social networks has eventually led to the construction of a database that facilitates research in community evolution. It is difficult to maintain a social network platform with hundreds of millions of people using artificial measures. An efficient community evolution algorithm is urgently required to accurately predict the social relationship in the network, so as to provide a better social experience for the Internet users [43] and to ensure the optimization of the social network platform in a benign direction. In this context, community evolution has been widely researched, which has led to the development of various community evolution algorithms. Most of such existing algorithms have focused on predicting and measuring the similarities between nodes. Existing community evolution algorithms can be categorized into four categories as follows: based on the similarity of node attributes, based on the similarity of network structure, based on probability distribution and based on machine learning.

(1) Similarity method based on node attributes

Node attribute similarity is used for the discovery of community evolution based on extracting attribute information of online social network nodes such as gender, age, region, interest, etc. A higher similarity in the node attributes reflect a higher similarity between the nodes in the network, thereby modelling the attribute information of nodes helps calculating the similarity as the measurement of community evolution. However, due to the promulgation of relevant national policies and the general improvement of Internet users' privacy awareness, it is difficult to obtain the attribute information of social network nodes, and sometimes produces the wrong data. Herein, extracting effective attribute information of nodes is a challenging task in online social networks [56].

The topic model [34-36] can efficiently resolve the problem of information extraction. The information published by social platform users is often open and real, easy to obtain, and highly reliable. Topic modelling is an unsupervised learning statistical strategy that is primarily used for text semantic analysis. The input of the model is a text content, and the output is the probability distribution of the topic. LDA model is a hierarchical topic model proposed by BLEI et al. [37] in 2003. It is divided into three levels: document, topic and word. The document is a combined distribution of multiple topics, and the topic is determined by the probability distribution of related words. The LDA model is used to analyze the content published by users, and the probability distribution of interest (topic) of social network nodes is obtained. Finally, the interest similarity between nodes is used for the community evolution discovery. This method can accurately extract the interest information of social network nodes, and can further predict the effective links in the network. However, with the expansion of the network scale and the diversification of information (text, pictures, audio and video, etc.), the subject model or some existing collaborative-based classification model is not quite adaptive to large-scale social network platform due to the problem of computational efficiency [57].

(2) Similarity method based on network structure

The strategy of using network structure similarity for community evolution discovery is simple and intuitive, which works completely based on the network structure information. The accuracy of this method depends on the selection of the structural similarity index. An appropriate structure similarity index can well represent the structural characteristics of a community network, and further can improve the accuracy of community evolution detection. Researchers have put forward many structural similarity indicators, the simplest of which is the common neighborhood Neighbor method, where nodes with more common friends may characterize more links in the present or in the future. However, the common neighbor's methods do not consider the issue of node influence, and its subsequent structural similarity indicators are based on the improvement of common neighbors. In 2003, Jaccard [38] proposed the Jaccard index, which not only considers the number of common neighbors, but also the proportion of common neighbors in all neighbors, that is, the influence factors of different nodes. Salton et al [39] proposed the Salton index, as an improvement of the CN index, and introduced the information of node degree as an influential factor. Adamic et al [40] believed that nodes with different degrees have different contributions, and the degree is inversely proportional to the contribution value. Therefore, each node is given a contribution degree, which considers the sum of the contribution degrees of the common neighbors for the discovery of community evolution.

In addition to considering the contribution of common neighbors, the local path (LP) method considers the contribution of the third-order neighbors. Katz [41] proposed the Katz index. In comparison with the common neighbor and the third-order neighbor methods, the Katz index considers all the possible paths. The shorter a given path is, the greater the weight will be.

(3) Machine-learning based method

With the rapid growth of computing, the field of machine-learning has its influence in the field of community evolution in online social networks in the recent years. Mostly of the machine-learning algorithms in such context has been predominantly supervised and semi-supervised.

Messaoudi [49] proposed a multi-objective Bat Algorithm to obtain high performance, which generate the initial population using the mean-shift algorithm. However, due to the computational complexity, machine-learning based community evolution methods are not suitable for large-scale network structures. On the premise of maintaining the prediction accuracy, reducing the computation time of machine-learning method is of far-reaching significance for the further application of machine-learning algorithms [53, 58].

III. CONCEPT AND DEFINITION

1. Social networks

In this section, we first present the definition of network topology. As the extension of real society in the network world, a social network can be abstracted as a graph structure $G(V.E)$, where $V = \{V_1, V_2, \cdots, V_n\}$ is the set of nodes in the network, $n = |V|$ is the number of nodes in the network, $e = \{e_{ij} | < V_i, V_j >, V_i, V_j \in V, i \neq j\}$ is the set of link relations, where $< V_i, V_j >$ represent the set of link relations between nodes V_i and V_j.

2. Community structure

Community is a sub-graph structure in the network topology, as shown in Figure 1, the density of node links within the community structure is higher than that of between communities, which means that the internal relationship of communities is closer, and also in line with the cognition of real-world social communities.

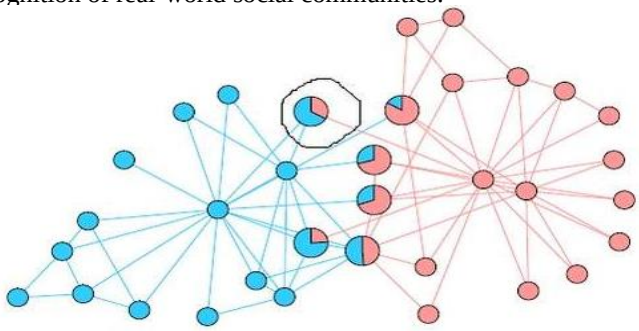

Figure 1. Community Structure

3. Evaluation index

(1) Modular Q function

The modular Q function [26] was first proposed by Newman and Girvan. It is defined as the difference between the proportion of community's internal links in the network structure and the expectation of the proportion of community's internal links in the random network. The quality of community discovery is usually measured by the modular Q value. The larger the Q value is, the more accurate the community structure will be. The modular Q function can be represented as in equation (1).

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \partial \left(C_i, C_j \right) \tag{1}$$

where, m is the total number of links in the network, and A_{ij} is an element of the adjacency matrix corresponding to the network. When there is a link relationship existing between V_i and V_j, A_{ij} is 1, otherwise 0; k_i is the degree of the user i; $\partial(C_i, C_j)$ denotes whether the community C_i and C_j are the same, and $C_i = C_j$ is 1, otherwise 0.

(2) Surprise function

The surprise function [17] is a modular evaluation index different from the traditional Q function. It describes the distance between a partition of the network and the expected distribution of community nodes and links under a given random model. The surprise function can be represented as in equation (2).

$$S = -log \sum_{j=p}^{min(M,n)} \frac{\binom{M}{j}\binom{F-M}{n-j}}{\binom{F}{n}} \tag{2}$$

where, F represents the maximum number of links in the network structure; n represents the actual number of links in the network; m represents the maximum number of links between communities in the network; p represents the actual number of links within the community. Experiments show that there is no resolution problem in the sure function. A higher price value reflects a more accurate community discovery.

(3) AUC

AUC (area under curve) is an important index to measure the quality of the community evolution algorithm. It compares the possibility of existing link and non-existent link in the test set. AUC index measures the community evolution results as a

whole. We randomly select an existing and non-existing link from the test set, and used the community evolution to predict the possibility of the above two community links, which are respectively recorded as pre_{linked} and $pre_{unlinked}$. If $pre_{linked} > pre_{unlinked}$, the evaluation index accumulates 1 point; if $pre_{linked} = pre_{unlinked}$, the evaluation index accumulates 0.5 points; the calculation of AUC is shown in equation (3).

$$AUC = \frac{n' + 0.5n''}{n} \tag{3}$$

where, n is the number of measurements, n' is the value of $pre_{linked} > pre_{unlinked}$, and n'' is the value of $pre_{linked} = pre_{unlinked}$. The higher the AUC value is, the better the quality of link prediction algorithm will be.

(4) F-measure

In order to compare the UIBEP model with other methods, three validation measures are introduced: community precision (*Precision*), community recall (*Recall*), *F-measure*:

$$Precision = \frac{|x \cap y|}{|x|} \tag{4}$$

$$Recall = \frac{|x \cap y|}{|y|} \tag{5}$$

$$F - measure = 2\frac{Precision*Recall}{Precision + Recall} \tag{6}$$

4. LDA model and Topical HITS algorithm

This section introduces the LDA model [1, 2] in detail, which forms the foundation for the principle and reasoning process of the Topical HITS algorithm [5, 9].

The community discovery model plays an extremely important part in the problem of community evolution. The best way to judge whether a node is interested in commodities or events is to analyze their topic distribution. As a key indicator of the node's communication ability, topic distribution can well describe the process of communication. In addition, topic information, community influence, node authority and centrality also play a key role in the community discovery process, hence are integrated into our new model. This section briefly introduces the classical community discovery model, and based on which, we propose a new community discovery model by integrating topic, community influence and other elements.

1) LDA topic model

LDA topic model is a Bayesian probability model with three layers of variable parameters proposed by David M. Blei et al. [37] in 2003. It is called the Potential Dirichlet Distribution Model. The three layers of variable parameters are words, topics and documents. LDA involves many theories such as Bayes theory, Dirichlet distribution and so on. It belongs to unsupervised machine learning technology and is used to infer potential topics contained in a document set or corpus.

LDA treats each document as a word vector, which is used to perform complex mathematical calculations, thus transforming text information into digital information that can be easily modelled. A document should contain several topics, and words are obtained by calculating the probability distribution of topics. The polynomial distribution of words is used to represent a topic distribution. Similarly, the polynomial distribution of topics is used to represent a document.

2) Topical HITS Algorithm

It has been proved that besides text, documents contain some properties that can represent the characteristics of nodes. Jon Kleinberg [5, 9] believes that documents have two potential attributes: Authority and Hub.

Jon Kleinberg believes that if a page has a high degree of authority, then the page will be linked by many centrality nodes; at the same time, if a page has a high degree of hub, then the page will also be linked by many authoritative pages. Accordingly, Jon Kleinberg proposed the Hyperlink Induced Topic.

Since topic factors are not considered seriously as an important factor in the initial version of the HITS algorithm, HITS performed well in most of the text-based search engines. However, with the increasing influence of the topic factor, the HITS algorithm is no longer applicable when it plays a decisive role in determining the effect of the algorithm. In view of this situation, Shi et al. [51] integrated topic factors into HITS algorithm and proposed the Topical HITS algorithm.

As we can see from Figure 2, in the Topical HITS algorithm, Authority vector and Hub vector of authority degree are considered instead of a single authority degree and centrality degree. Each dimension of the Authority vector and the Hub vector maps a topic, and the dimension is the number of topics contained in the current document. A topical HITS algorithm uses a multi-surfer semantic model as a random-access model. According to the behavior of surfer *A*, authority *A* can be obtained, and centrality *H* can be obtained according to the behavior of surfer *H*. We extend the HITS algorithm to exploit the inseparable connection between users and their corresponding posts for the purpose of extracting only high-quality posts and influential users. Thus, the Topical HITS algorithm method can effectively filter out random low-quality posts and ordinary users, and thereby avoid a phenomenon known as a bump, which generally reduces the efficiency and accuracy of event detection as well as the identification of influential spreaders.

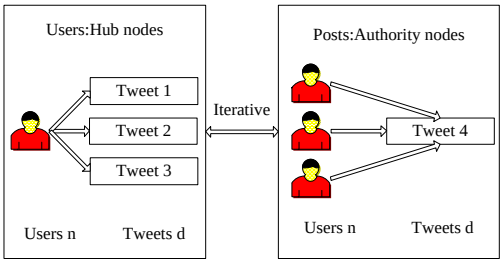

Figure 2. Iterative Model for Topical HITS algorithm[14]

IV. LPAE ALGORITHM

1. Seed Link Selection

The LPAE algorithm is a local optimal community discovery algorithm, which can efficiently resolve the instability issue of the classical LPA algorithm. As we can see from Figure 3, in the initial seed link selection stage, the random strategy is no longer used. Based on the modularity characteristics of social networks, the hub value of the nodes is calculated. The sum of hub value of nodes at both ends of the link is considered as the link centrality of the corresponding link, and further the link with the largest link centrality is confirmed and considered as the initial seed link to avoid the instability caused by the random selection of the initial seed link.

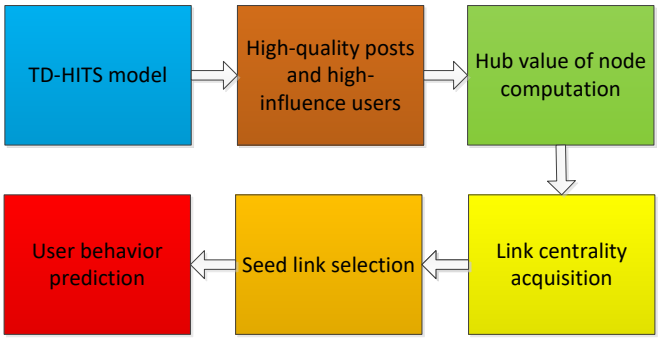

Figure 3. The Procedure of the LPAE algorithm

(1) Problem description

The undirected graph is used to represent the given target network, by including the set of all nodes in the network, and the set of links between nodes. Now, we mine the sub-graph structure set of the network from the given target network G, that is, the community set, so that the internal node links of the sub-graph structure are more intensive, and the links between the sub-graph structures are more sparse.

(2) Calculate node centrality

In order to solve the instability issue of the LPA algorithm caused by the random selection of initial seed links, the LPAE algorithm first determines the internal nodes of the community. According to the concept of community structure, the nodes within the community are closely linked, and the nodes with higher influence are likely to be within the community. Based on the introduction of centrality assessment, it can be seen that the concept of centrality assessment index is simple and its calculation performance is high, which can be used to assess the influence of nodes in network G. Therefore, the first step of the LPAE algorithm is to calculate the centrality of all nodes V in network G, where the node with the largest degree of centrality is determined as the internal node of the community. This internal node is used in subsequent steps of the process of determining the internal link of the community.

(3) Calculate link centrality

In order to obtain a high quality community structure, the LPAE algorithm should ensure that the initial seed link is located in the community structure, because when the seed link is located between communities, the algorithm accumulates relevant nodes belonging to different communities into a single community. This action directly leads to a low degree of community structure module. At present, no strategies exist to measure the importance of network links. Starting from the concept of the hub value of nodes, the nodes with higher hub values are likely to be located in the community, and when the hub value of two associated nodes via a given link is high, the link is also likely to be located in the community. Therefore, in this paper, the sum of the hub value of the nodes associated via a given link is considered as the measure of the link located in the community, which is called link centrality, and its definition is shown in equation (7). Finally, the link with the greatest link centrality is identified as the community's internal link, and it is regarded as the initial seed link.

$$C_L(e) = C_H(u) + C_H(v) \tag{7}$$

where, $C_H(u)$ and $C_H(v)$ represent the hub value of nodes u and v at both ends of link e.

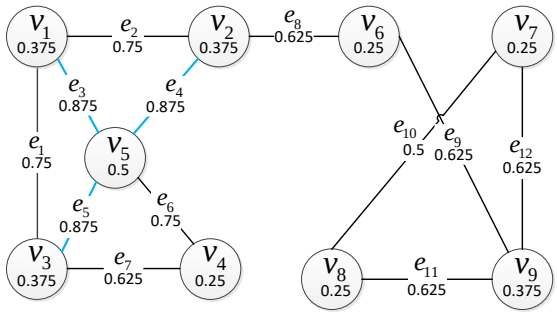

Figure 4. Seed link selection strategy of LPAE algorithm

The seed link selection strategy of the LPAE algorithm is shown in Figure 4. First, the hub value $C_H(v)$ of all nodes $v(v \in V)$ in the social network G is calculated by the hub value (for example, the hub value of node v_1 is 0.375), and then the link centrality $C_l(e)$ (for example, the link centrality of link e_8 is 0.625) of all links $e(e \in E)$ in the social network *G* is calculated using the link centrality formula (7). Thus, we can confirm that the most central links are e_3, e_4, and e_5. It can be seen from Figure 4 that the link with the greatest link centrality is obviously located in the community structure. Finally, a link set $\{e_3, e_4, e_5\}$ with the greatest link centrality is considered as the initial seed link.

2. User Interest Behavior Prediction

The diversity of online social network interaction methods brings users a new experience. For a given content of interest, users can participate in the interaction through likes, comments, forwarding, collection, sharing and reminders. In addition, there is a wide range of citation relationships in the literature citation network.

(1) Like: the most basic and common interaction behavior in online social networks. Users can express their support and appreciation for their favorite content through like. Praise is often used as the heat evaluation index of contents, which can be used for network information recommendation. This interactive way of liking is of great significance to the development of social networks.

(2) Comment: users express their opinions by commenting on the content, and can also communicate with other commentators.

(3) Forward: for the content of special interest, users can publish the content in their own space by forwarding them, and can simultaneously post their comments at the contents.

(4) Collection: for the content that users are interested in and intend to read for many times, users can save the content as their favorites through collection, so that users can read it again later. In social platforms, collection information is often only visible to users themselves, which is not easy to access publicly.

(5) Share: if users want to send content to other users outside the platform, users can make the content visible to other users by sharing. This interaction behavior is helpful for the promotion of social network platform. Because sharing operation occurs between internal users and other users outside the platform, furthermore, sharing data is not easy to obtain.

(6) Reminders (at): reminders often occur at the same time as comments, and are used to invite users to participate in interactive behaviors. When users are reminded, they will receive a message prompt, which can be easily used to locate the reminded location and participate in the communication and discussion on time.

Online social network interaction often characterizes forwarding and comments appearing at the same time. When users are reminded in a blog post, commenting behavior inevitably occurs, and praising behavior is the most common in the network. Therefore, the impact of different interaction behaviors is different, and their corresponding weights are also different. For example, forwarding is often more meaningful than likes, frequent commenting behavior means that users are more likely to know each other; liking and forwarding behavior represents users' recognition of social content, where users and content publishers are likely to have the same interest. In comparison with the lagging structural information, the interaction behavior can be better used to reflect the social trend and dynamics of users.

However, there are differences with the real weights, which ultimately have led to the lack of accuracy in the community evolution part. Therefore, in order to further improve the accuracy of community evolution, this paper incorporates such complex interaction behaviors into the community evolution problem, at the same time, evaluates the weight of various

interaction behaviors in the network, and proposes the complex interaction behavior based community evolution model (UIBEP).

V. UIBEP MODEL

Community evolution is an important focus in the community structure research. Evolution is the basic characteristic of real networks. The communities in social networks are evolving continuously over time, which results from the interaction between network's own structure and the frequent interactions occurring on it. Addressing the community evolution, we mainly build a community evolution model based on the historical characteristics of the community in the network, and further predict the possible changes in the future. The study of community evolution also facilitates researchers to analyze changes in user interest and predict user behavior and hotspot trends anticipated in the future. The information presented in social networks is updated rapidly over time. Moreover, various social events in relation to the changes in users' social relations, behaviors and interests etc., also lead to changes in the community.

In this paper, as an enhancement to the sub-graph increment method proposed by Liu et al. [45], a novel complex interaction behavior based community evolution algorithm (UIBEP) using Cosine Similarity [9, 51] is proposed to accurately detect targeted communities during community evolution. Our proposed model characterizes high query efficiency and good scalability. The sequential process involved in our proposed model is illustrated in Figure 5.

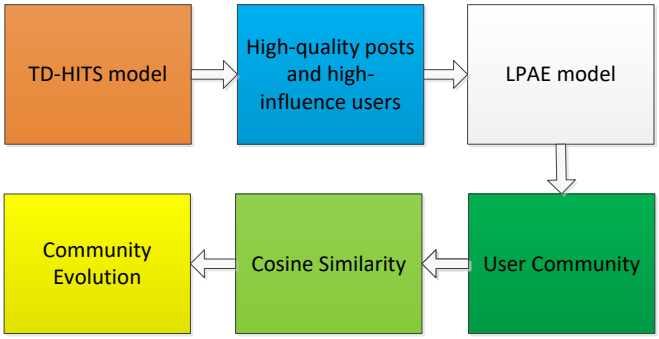

Figure 5. The Procedure of the UIBEP method

A. Community Interaction Behaviors

A complex relationship usually exists between the interaction behaviors in online social networks. The type of interaction and their effects usually vary across different networks. For example, cooperation and reference behaviors are more common in citation networks. Therefore, a generalized model to capture the influence of all kinds of interaction behaviors in online social network is quite complex. It is necessary to quantify the influence of all kinds of interaction behaviors in the network based on network information, that is, interaction weight w. However, online social networks are large in scale and sparse in data, and nodes often only interact closely with the surrounding nodes. Calculating the interaction weight for the entire network usually involve higher time overheads and other performance issues. Therefore, in order to avoid data sparsity and reduce time overheads, it is necessary to filter the social network data, sample relevant nodes and use their interaction data as the training set, and further regard the analysis results of the training data as the approximate value of the interaction weight of the entire network.

However, social network data is highly heterogeneous. If the random sampling strategy is adopted, especially when the frequency of interaction between the network nodes is low, the representativeness of training data becomes poor. This further leads to an inaccurate calculation of the interaction weight and affects the subsequent measurement of node interaction similarity, and ultimately affects the accuracy of community evolution. In order to ensure a timely performance and community evolution accuracy, this paper selects the part of network data with the highest interaction frequency as the training data.

(1) Interactive behavior data statistics

Interaction behavior in online social networks is an abstract concept. The process of modelling the interaction behavior should be concrete and symbolic. First of all, we study the interaction behavior in the online social network, analyze all kinds of interaction behavior types, and define the set of interaction behavior types, as $k = |b|$, where each element represents an interaction type. Secondly, the original dataset of the social network interaction behavior is obtained by technical means, which indicates that there is an interaction behavior b_i between nodes u and v, while nodes u and v may generate the same interaction behavior under different blogs, for example, node u likes under 10 blogs of node v. Therefore, the frequency of a given interaction between corresponding node pairs is computed and denoted as r, and the interaction behavior data is represented by a three-dimensional interaction behavior matrix. Here, as defined above, the aggregation of

all nodes in the social network G is represented, for a set of interaction behavior types, $n = |v|$, $k = |b|$, section x represents the interaction between node v_x and all other nodes. Each matrix element $R[x, y, z]$ corresponds to the frequency of interaction behavior b_z between node v_x and node v_y.

(2) Determine the set of candidate nodes

Online social networks tend to be large-scale and sparse, so it is necessary to select the most representative part of the data in the entire network, and the nodes with the most number of neighbors are usually located in the most central part of the network, which is more representative. Therefore, first, the centrality $C(v)$ of all nodes is calculated according to the number of neighbor nodes. Then, the node with the largest hub value v_{cmax} is selected, and its neighbor node set can be defined.

B. Cosine Similarity

In online social networks, people's community undergoes changes every day. When the users in the community change, it is important to quickly identify the influential users [44].

Thus, this paper applies cosine similarity [50] to our field. First of all, it is difficult to obtain useful information from online social networks, because most of the microblogs posted or forwarded by users are short texts, disguised as various topics, and will not be published or forwarded in strict chronological order. Therefore, in order to solve this problem, this paper proposes the concept of long text document. Long text documents are composed of key microblogs in each user's topic community to identify some keywords in each topic community that can represent the hot events. In each long text document, because the microblog topic is similar, the keyword combination is more likely to belong to the same topic. This solves the problem of sparse features of short text microblog, and also enhances the ability to learn from the topic, thus improving the quality of topic analysis. In addition, in the process of learning topics, users' interest topic keywords can be obtained through the topic content keywords of long text documents. Then, the microblog is reprocessed by eliminating stop words and extracting keywords. Finally, the evolution chain of hot events is identified according to cosine similarity.

Specifically, each microblog d_i is a set of vectors composed of keywords. The topic of each event can also be represented by the keywords in the key microblog d_k. Finally, the cosine theorem is used as a criterion to judge the correlation d_k between microblogs and key microblogs. The cosine distance between microblog d_i and key microblogs [50] is calculated as follows:

$$\cos\left(d_i, d_k\right) = \frac{\vec{d}_i \cdot \vec{d}_k}{\left\|d_i\right\|\left\|d_k\right\|} \tag{8}$$

If the value of $cos(d_i, d_k)$ is higher, it means that there is a higher community similarity. The threshold is defined as λp. When $cos(di, dk)$ of d_i is greater than λp, d_i and d_k are considered to be the same community. In addition, η is calculated as follows.

$$\eta = \frac{n(d_k)}{r} \tag{9}$$

where, $n(d_k)$ represents the number of users similar to the key user d_k, and the threshold λp is confirmed by training η.

VI. EXPERIMENTS

In this section, we detail the experiments conducted on real-world short-text dataset to demonstrate the effectiveness of our proposed LPAE method and UIBEP model.

In the rest of this section, we describe our collection of the dataset, experimental setup and analysis, the baseline approaches, and model evaluation.

A. Dataset

The dataset used in this experiment is shown as follows.

(1) Zachary Karate Club dataset

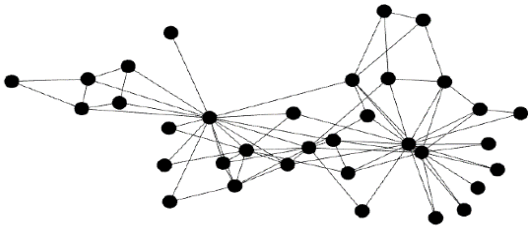

Figure 6. Network structure of Zachary Karate Club

Zachary karate club network [46] is a relationship network obtained by sociologist Zachary over two years of observing social relations among 34 members of a karate club in the United States. It contains 34 nodes and 78 links. Its network structure is shown in Figure 6. This network is used as a common dataset in the research of community discovery.

Table I Zachary Karate Club Dataset

Network	Node	Edge	Community
Zachary	34	78	2

(2) LFR artificial network

LFR network [47] is a manually generated dataset very similar to the real-world social dataset. It can simulate various scenarios through configuration parameters, including number of nodes n , hybrid parameter μ , node average k, node maximum $k_{\max}$ minimum community size $c_{\min}$ and maximum community size $c_{\max}$. In order to verify the performance of our LPAE algorithm in a large network structure, this paper generates several artificial network structures as shown in Table II.

Table II LFR Data Structure

Network	Node	Degree	Minimum	Maximum
LFR-1000	1000	10	10	50
LFR-5000	5000	15	10	50
LFR-10000	10000	20	20	100

(3) Twitter social network

We generated our dataset from Twitter (http://twitter.com/) via Twitter API. This dataset consists of 40,000 posts from October 25–28, 2015. We included only those users who published or commented upon posts in our dataset. After filtering unwanted users and posts, the dataset comprises 2139 high-quality posts and 1887 users.

B. Experimental Settings

We conducted the experiments on a computer with an Intel I7 3.4 GHz CPU and 16G memory.

We tuned the parameters via a grid search. For LDA [1, 2, 51], α = 0.5 and β = 0.1. In all of our experiments, we used Gibbs sampling for 1,000 iterations. The results reported in this paper are obtained as a five-run average. In the process of filtering high-quality dataset, we set all of the initial authority scores *d.a* and hub scores *u.h* to 1.

C. Baseline Approaches

LPAE algorithm solves the instability problem of the LPA algorithm, and in theory, the LPAE algorithm has lower time complexity. In order to verify the performance of the LPAE algorithm on real networks, this section conducts experiments on Zachary karate club network and multiple LFR artificial generated networks, and evaluates our LPAE model against the classical LPA algorithm under the same experimental conditions. Because of the instability of the initial seed link in the LPA algorithm, we run the LPA algorithm for 10 times respectively, and take the optimal module degree as the final result. In this summary, the module degree *Q* value, the Surprise value and the running time of the algorithm are used as the evaluation and analysis indexes.

We validated the improved efficiency and effectiveness of the proposed LPAE and UIBEP by evaluating our model against LPA [48], which is a classical latent semantic analysis algorithm. Meanwhile, in this paper, the social network data is used as the input of UIBEP algorithm, and the weights of three kinds of interaction behaviors, i.e. likes, comments and forwards, are calculated, which are 0.13, 0.34 and 0.55 respectively. Then, they are substituted into the UIBEP algorithm to record the community evolution results. Secondly, the community evolution algorithm based on LDA model, CN and Jaccard index is used for our comparative evaluation, under the same experimental environment and duration.

D. Evaluation Methods

1) The experimental results

The results of the LPAE algorithm and the LPA algorithm on Zachary Karate Club dataset are shown in Table III. Because the initial link selection of the LPA algorithm is random, we run the experiment for 10 times to obtain 10 sets of modular *Q* value and Surprise value, where "LPA-n" represents the number of experiment.

Table III Zachary Karate Club

Method	*Q*	*Surprise*
LPAE	0.3688	260.0215
LPA-1	0.2276	195.6582
LPA-2	0.2959	208.6581
LPA-3	0.2558	180.3647

LPA-4	0.3705	241.2368
LPA-5	0.3688	226.3648
LPA-6	0.3336	235.3614
LPA-7	0.2962	191.5621
LPA-8	0.2254	146.0305
LPA-9	0.3505	250.3923
LPA-10	0.2742	185.0827

It can be seen from Table III that the LPA algorithm runs on the Zachary karate club network to get the fluctuation state of the modular *Q* value and the *Surprise* value. The results of the fourth experiment (LPA-4) of the LPA algorithm are the best, which are also consistent with the results of the LPAE algorithm. That is to say, when the initial link of the LPA algorithm is selected as the maximum link, the optimal module degree *Q* and the *Surprise* value can be obtained. When the initial link is selected as the community internal link, the LPA algorithm exhibits better results; and when the initial link is selected as the community inter link, the LPAE algorithm exhibits better results, the modular *Q* and *Surprise* values of the LPAE algorithm are low. Figure 7 and Figure 10 illustrate the curves of modular *Q* value and *Surprise* value of LPAE algorithm and LPA algorithm respectively.

According to the modularity characteristics of social networks, the links between communities only account for a small part of the whole network link set. Therefore, in the case of smaller probability, the LPA algorithm selects the links between communities as the initial seed links, resulting in the low modularity of the community structure divided by the LPA algorithm. In most cases, the LPA algorithm achieves reliable results. However, the LPAE algorithm indirectly determines the link located in the community through the hub value of nodes. In the case of smaller probability, the effect of LPAE algorithm is the same as that of LPA algorithm; in other cases, the community quality obtained by the LPAE algorithm is better than that of the LPA algorithm, which shows that our LPAE algorithm can discover the community structure with better stability and reliability.

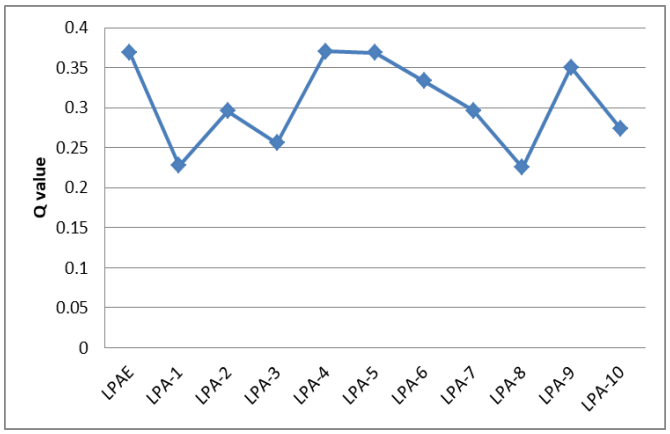

Figure 7. Comparison of *Q* value

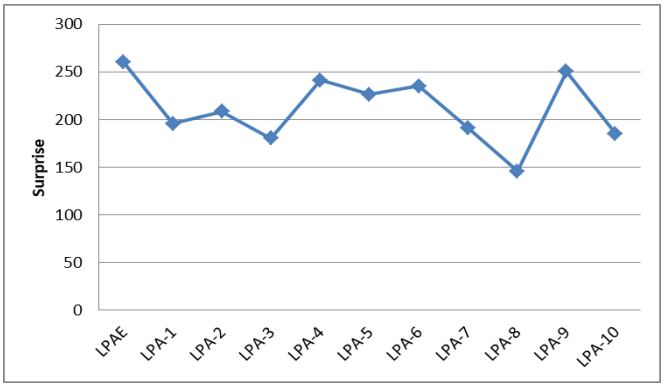

Figure 8. Comparison of *Surprise* value with LPAE and LPA 1-10

Figure 8 compares the community discovery module degrees between LPAE algorithm and LPA algorithm on LFR networks of different scales. In the initial seed link selection stage, when compared with the random selection strategy of the

LPA algorithm, our proposed LPAE algorithm initially selects the community internal link by calculating the link centrality. Hence, the community structure module degree obtained by our LPAE algorithm is better than that of the LPA algorithm. Figure 9 shows the runtime duration of the LPAE algorithm and the LPA algorithm on LFR artificial network respectively. It can be observed clearly that the runtime of the LPAE algorithm is lower than that of the LPA algorithm, especially when the scale of the network further evolves. It is obvious from the analysis that our LPAE algorithm incurs more node centrality calculation and link centrality calculation during the process of seed link selection, but further optimization is carried out during the process of local expansion. Herein, unnecessary repeated calculations are omitted, thus the runtime of our LPAE algorithm is lower than that of the LPA algorithm. Thus, we conclude that the performance of our LPAE algorithm is better than that of the LPA algorithm.

In conclusion, our proposed LPAE algorithm exhibits good time performance, solves the instability issue of the LPA algorithm, can quickly explore the community structure in online social network with stability, and improves the quality of community discovery.

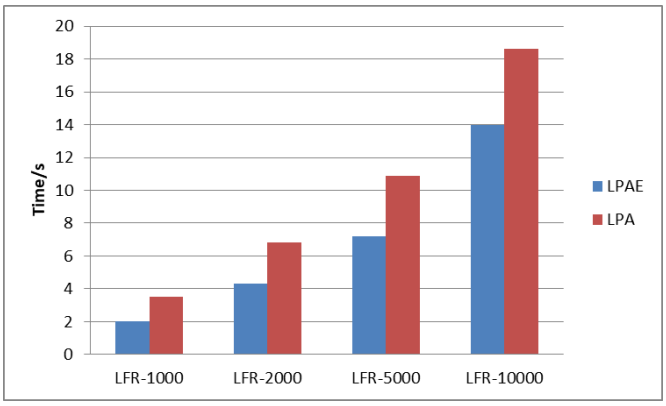

Figure 9. Comparison of Time

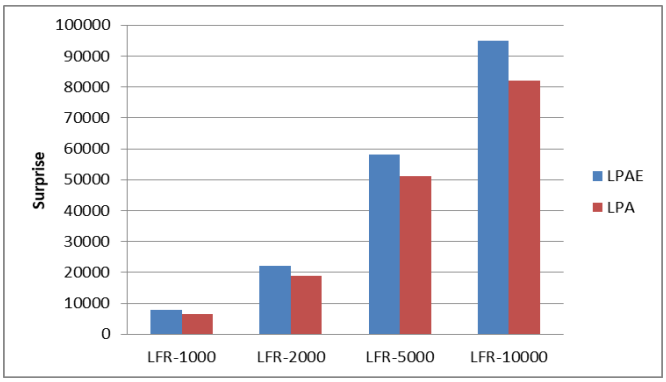

Figure 10. Comparison of Surprise value with LPAE and LPA

2) Comparative analysis of operation time

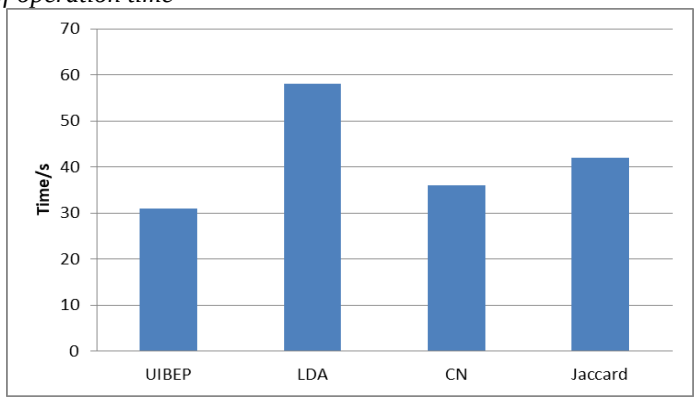

Figure 11. Comparison of operation time

Figure 11 shows the operation of each of the studied algorithm based on the undirected Twitter network. It is obvious that our proposed UIBEP algorithm exhibits shorter runtime duration and higher performance than other studied algorithms. Further analysis of this situation shows that the community evolution algorithm based on the LDA model incurs more computation time when analyzing the probability distribution of interest topics of microblog users' blogs. The LDA model algorithm calculates the similarity between nodes and the topics of all other unlinked nodes in online social network, which results in the runtime of the LDA algorithm being much longer than other algorithms, especially when the network scale is larger. With large-scale networks, this phenomenon is more obvious. CN index and Jaccard index are simple in principle and can quickly measure the structural similarity of node pairs. However, they are still incurring full graph range calculations, and so the final runtime duration is higher than the UIBEP algorithm. Thus, our proposed UIBEP algorithm characterizes significant advantages in terms of time complexity and efficiency.

3) Comparative analysis of AUC and Precision

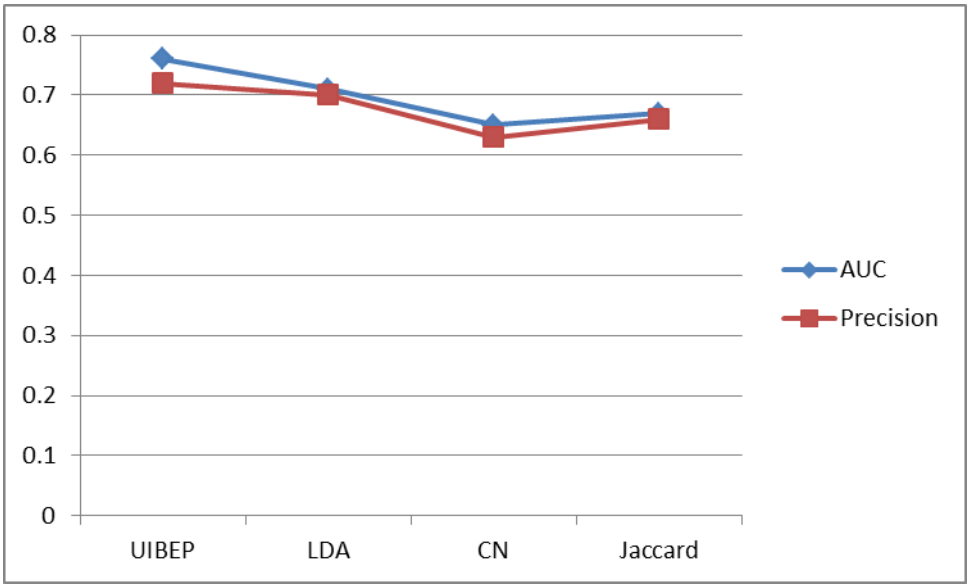

Figure 12. Comparisons of AUC and Precision

The AUC and Precision evaluation index curves of our UIBEP algorithm and other studied algorithms, based on the undirected Twitter network, are presented in Figure 12. It can be observed from Figure 12 that the community evolution effect of our propsoed UIBEP algorithm is slightly better than that of the LDA model. Based on the CN index and Jaccard index, our UIBEP algorithm exhibits greater improvement in accuracy. First, the community evolution algorithm based on the LDA model obtains the probability distribution of interest topics of nodes through the model before comparing the similarity of nodes. Users with similar interest topics are more likely to have attention behaviors, that is, the corresponding nodes are more likely to have links presently or in the future. This phenomenon is also consistent with people's consistent cognition. However, there are some errors in the calculation of probability distribution, especially when the user posts are less or not published, the LDA model has less input data, and the distribution of user interest obtained from the analysis is quite different from the real situation. This caused the community evolution algorithm of the LDA model to characterize a slightly lower evaluation index than the UIBEP algorithm. The community evolution algorithm based on CN index and Jaccard index is completely based on the static structure of the network. When the nodes solely calculate the similarity with the surrounding nodes, they can get higher similarity. However, for the nodes that are far away, the structure similarity is low, and the community evolution results are not ideal. The UIBEP algorithm is more intuitive and simple than the LDA model, directly relies on the interaction behavior data to measure the similarity between users. The results of AUC and precision show that our proposed UIBEP algorithm can deliver more accurate community evolution results.

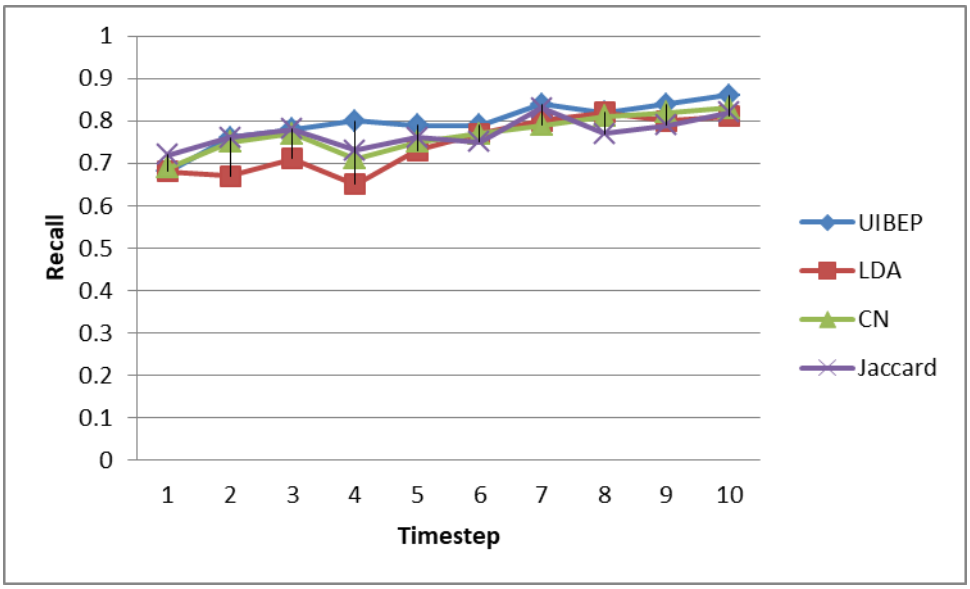

Figure 13. Comparison of Recall rate

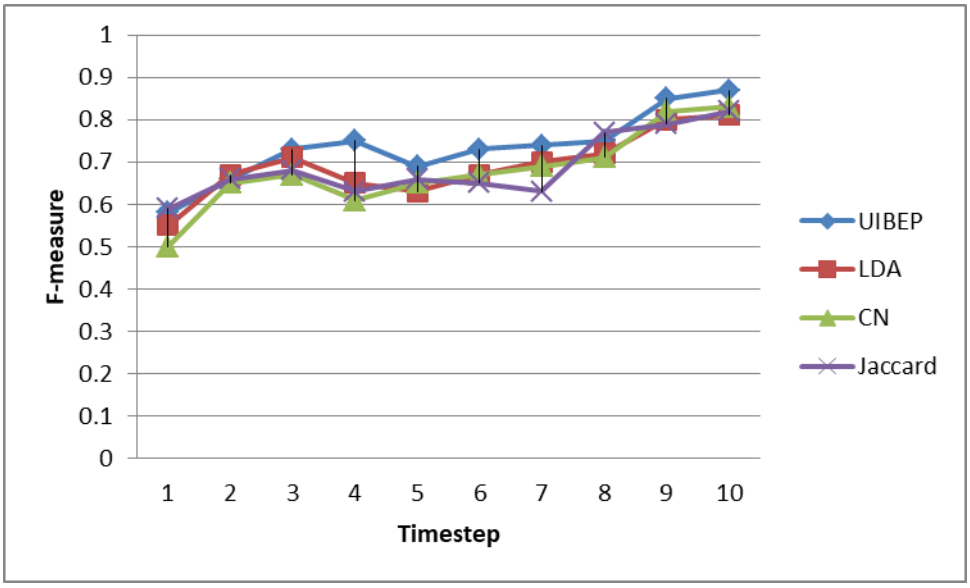

Figure 14. Comparison of F-measure value

4) Core user replacement based on Cosine Similarity

In online social networks, people's interests may change at any time. It is unavoidable to replace core users based on Cosine Similarity when tracking dynamic user interest community evolution. We compare the UIBEP model with other existing algorithms, and used Recall, F-measure as the evaluation criteria.

As we can see from Figure 13 and Figure 14, the results of the studied method do not exhibit significant difference in performance. This is because these methods use the respective pre-processing methods to process the data, and the core sub-graphs can be obtained ideally without considering the efficiency. However, there are many options that exist for core user replacement, but managing the core sub-graph after replacement is important and significantly affects the quality of the method.

5) Summary of experiment

Based on the analysis of the runtime duration and the evaluation index, our UIBEP algorithm makes reasonable use of the modularity characteristics of online social network. Based on the idea of divide and rule, the prediction of the entire network is reduced to community structure, which can effectively reduce the time complexity. By introducing interactive behavior data into the community evolution problem, the accuracy of community evolution algorithm can be effectively improved by

quantifying the interaction similarity between nodes, as the index of community evolution. Therefore, our proposed UIBEP algorithm exhibits good performance in terms of time overheads and prediction accuracy, and has comprehensive advantages.

VII. CONCLUSIONS AND FUTURE WORK

This paper studied and proposed an edge intelligence-enabled community discovery algorithm (LPAE algorithm) based on node centrality and link centrality expansion. Firstly, the hub value of all nodes in online social networks is calculated. Secondly, the sum of the hub value of the nodes at both ends of the link is taken as the link centrality. The link corresponding to the maximum link centrality is selected as the initial seed link of the network to ensure that the seed link is located in the community. The instability problem caused by the random selection of the seed link in the LPA algorithm is resolved. Secondly, based on a greedy optimization, the local extension is achieved quickly. When the link is added into the network, the added link is compared with all other links. Our LPAE algorithm only includes those links around the chosen link into the local candidate set, so as to improve the efficiency of the algorithm. Finally, the LPAE algorithm can efficiently resolve the instability problem and can identify the community structure quickly and accurately.

The UIBEP algorithm based on complex interaction behavior is studied and designed. First of all, all kinds of interaction behaviors are analyzed in online social networks. Secondly, the network data is filtered using an improved HITS algorithm, and the sample matrix of interaction behaviors are obtained to construct the augmented matrix for calculating the weight vector of all kinds of interaction behaviors. Finally, our LPAE algorithm is used to detect the network community for obtaining the high-quality community structure, which can be used as the unit of community evolution.

As future work, we plan to focus on the users' interest communities and its evolution for achieving user behavior prediction. And we will study the links prediction between influential users.

ACKNOWLEDGEMENT

The work reported in this paper has been supported by National Natural Science of Foundation of China Program (61502209 and 61502207), and Suqian Municipal Science and Technology Plan Project in 2020 (S202015).

REFERENCES

1 J. Ge, L. -L. Shi, Y. Wu and J. Liu, "Human-Driven Dynamic Community Influence Maximization in Social Media Data Streams," in IEEE Access, vol. 8, pp. 162238-162251, 2020.

2 L. Jiang, L. Shi, L. Liu, J. Yao, and M. A. Yousuf, "User interest community detection on social media using collaborative filtering," *Wireless Networks*, vol. 8, pp. 1-24, 2019.

3 Y.H. Guo, L. Liu, Y. Wu, J. Hardy, "Interest-Aware Content Discovery in Peer-to-Peer Social Networks", ACM Transactions on Internet Technology, vol.18, no. 3, pp. 152-172, 2017.

4 Z. Han, W. K. S. Tang, and Q. Jia, "Event-Triggered Synchronization for Nonlinear Multi-Agent Systems With Sampled Data," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 67, pp. 3553-3561, 2020.

5 M. Y. Wu, L. Zhang, and M. Imran, "Research on Evolution Model of Employees' Relationship and Task Conflict in Enterprise Clusters: A Complex Network Perspective," Wireless Personal Communications, vol. 102, pp. 3489–3501, 2018.

6 G. Chen, L. Xu, X.-B. Liu, Wang, and Zhao, "Picking Robot Visual Servo Control Based on Modified Fuzzy Neural Network Sliding Mode Algorithms," Electronics, vol. 8, pp. 605-622, 2019.

7 A. Monney, Y. Zhan, J. Zhen, and B. B. Benuwa, "A Multi-Kernel Method of Measuring Adaptive Similarity for Spectral Clustering," Expert Systems with Applications, vol. 159, pp. 157-168, 2020.

8 L. Liu, N. Antonopoulos, M. H. Zheng, Y. Z. Zhan, Z. J. Ding, "A Socio-ecological Model for Advanced Service Discovery in Machine-to-Machine Communication Networks", ACM Transactions on Embedded Computing Systems, Vol 15(2), 2016, pp 38:1–38:26.

9 L. Jiang, L. L. Shi, L. Liu, B. Yuan, Y. J. Zheng, "An Efficient Evolutionary User Interest Community Discovery Model in Dynamic Social Networks for Internet of People", IEEE Internet of Things Journal, vol. 6, no. 6, pp. 9226 - 9236 , 2019.

10 M. Ahmed, L. Liu, J. Hardy, B. Yuan, and N. Antonopoulos, "An efficient algorithm for partially matched services in internet of services," Personal & Ubiquitous Computing, vol. 20, pp. 283-293, 2016.

11 M. Wang, L-L. Shi, L. Liu, M. Ahmed, J. Panneerselvam, Hybrid Recommendation based QoS Prediction for Sensor Services, International Journal of Distributed Sensor Networks, 14 (5), 2018, pp 1-10.

12 L. Liu, N. Antonopoulos, M. Zheng, Y. Zhan, Z. Ding, A Socio-ecological Model for Advanced Service Discovery in Machine-to-Machine Communication Networks, ACM Transactions on Embedded Computing Systems, Vol 15(2), 2016, pp 38:1–38:26.

13 S. Tang, S. Yuan, and Y. Zhu, "Convolutional Neural Network in Intelligent Fault Diagnosis Toward Rotatory Machinery," IEEE Access, vol. 8, pp. 86510-86519, 2020.

14 S. Zeng, B. Zhang, Y. Lan, and J. Gou, "Robust collaborative representation-based classification via regularization of truncated total least squares," Neural Computing & Applications, vol. 31, pp. 5689–5697, 2019.

15 B. Yuan, J. Panneerselvam, L. Liu, N. Antonopoulos and Y. Lu, "An Inductive Content-Augmented Network Embedding Model for Edge Artificial Intelligence," in IEEE Transactions on Industrial Informatics, vol. 15, no. 7, pp. 4295-4305, 2019.

16 M. Nilashi, P. F. Rupani, M. M. Rupani, H. Kamyab, W. Shao, H. Ahmadi, et al., "Measuring sustainability through ecological sustainability and human sustainability: A machine learning approach," Journal of Cleaner Production, vol. 240, p. 118-162, 2019.

17 Wilson C, Sala A, Puttaswamy K P N, et al. "Beyond social graphs: User interactions in online social networks and their implications". ACM Transactions on the Web (TWEB), Vol 6(4),pp. 1-31, 2012

18 Clauset A, Moore C, Newman M E J. "Hierarchical structure and the prediction of missing links in networks". Nature, Vol 453(7191), pp. 98-101, 2008.

19 Almansoori W, Gao S, Jarada T M, et al. Link prediction and classification in social networks and its application in healthcare//2011 IEEE International Conference on Information Reuse & Integration. IEEE, 2011: 422-428.

20 Cannistraci C V, Alanis-Lobato G, Ravasi T. From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in social network. Scientific reports, 2013, 3(1): 1-14.

21 L. L. Shi, L. Liu, Y. Wu, L. Jiang, and A. Ayorinde, "Event Detection and Multi-source Propagation for Online Social Network Management," Journal of Network and Systems Management, vol. 28, pp. 1-20, 2019.

22 Y. Wu, C. G. Yan, L. Liu, Z. J. Ding and C. J. Jiang, An Adaptive Multilevel Indexing Method for Disaster Service Discovery, IEEE Transactions on Computers, Vol 64(9), September 2015, pp 2447 - 2459.

23 H. H. Altalib, H. J. Lanham, K. K. McMillan, M. Habeeb, B. Fenton, K.-H. Cheung, and M. J. Pugh, "Measuring coordination of epilepsy care: A mixed methods evaluation of social network analysis versus relational coordination," Epilepsy & Behavior, vol. 97, pp. 197-205, 2019.

24 C.-z. Gao, Q. Cheng, X. Li, and S.-b. Xia, "Cloud-assisted privacy-preserving profile-matching scheme under multiple keys in mobile social network," Cluster Computing, vol. 22, pp. 1655-1663, 2019.

25 H. Wang, Y. Wu, G. Min, J. Xu, and P. Tang, "Data-driven Dynamic Resource Scheduling for Network Slicing: A Deep Reinforcement Learning Approach," Information Sciences, vol. 498, pp. 106-116, 2019.

26 Girvan M, Newman M E J. Community structure in social and biological networks. Proceedings of the national academy of sciences, vol. 99, no. 12,pp. 7821-7826, 2002.

27 Soundarajan S, Hopcroft J. Using community information to improve the precision of link prediction methods//Proceedings of the 21st International Conference on World Wide Web. 2012: 607-608.

28 D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," Journal of machine Learning research, vol. 3, May. 2003, pp. 993-1022.

29 Raghavan U N, Albert R, Kumara S. "Near linear time algorithm to detect community structures in large-scale networks". Physical review E, Vol 76(3), pp. 36-48, 2007.

30 Lancichinetti A, Fortunato S, Kertész J. "Detecting the overlapping and hierarchical community structure in social network". New journal of physics,Vol 11(3), pp. 33-45, 2009.

31 Baumes J, Goldberg M K, Krishnamoorthy M S, et al. Finding communities by clustering a graph into overlapping subgraphs. IADIS AC, 2005, 5: 97-104.

32 Su Y, Wang B, Zhang X. A seed-expanding method based on random walks for community detection in networks with ambiguous community structures. Scientific reports, 2017, Vol 7: 41830.

33 Gregory S. Finding overlapping communities in networks by label propagation. New journal of Physics, Vol 12(10), pp.103-108, 2010.

34 Y. Papanikolaou and G. Tsoumakas, "Subset Labeled LDA: A Topic Model for Extreme Multi-label Classification," in International Conference on Big Data Analytics and Knowledge Discovery, pp. 152-162, 2018.

35 D. Backenroth, Z. He, K. Kiryluk, V. Boeva, L. Pethukova, E. Khurana, et al., "FUN-LDA: A latent Dirichlet allocation model for predicting tissue-specific functional effects of noncoding variation: methods and applications," The American Journal of Human Genetics, vol. 102, pp. 920-942, 2018.

36 C. Huang, Q. Wang, D. Yang, and F. Xu, "Topic mining of tourist attractions based on a seasonal context aware LDA model," Intelligent Data Analysis, vol. 22, pp. 383-405, 2018.

37 .Blei D M , Ng A Y , Jordan M I , et al. Latent Dirichlet Allocation. Journal of Machine Learning Research, Vol 3:993-1022, ,2003.

38 Jaccard P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bull Soc Vaudoise Sci Nat, Vol 37(142), pp.547-579, 1901.

39 Salton G, Mcgill M J. Introduction to modern information retrieval. Auckland: McGraw-Hill, 1983.

40 Adamic L A, Adar E. Friends and neighbors on the web. Social networks, Vol 25(3) pp. 211-230, 2003.

41 Katz L. A new status index derived from sociometric analysis. Psychometrika, Vol 18(1), pp. 39-43, 1953.

42 H. Q. Wu, L. M. Wang, S. R. Jiang. "Secure Top- k Preference Query for Location-based Services in Crowd-outsourcing Environments" ,The Computer Journal, vol. 61, no. 4, pp. 496-511, 2018.

43 Xu, Z, Y. Wu, L. Liu, "A novel multilevel index model for distributed service repositories", Tsinghua Science and Technology, vol. 22, no. 3, pp. 273-281, 2017.

44 G. Li, L. Yan, and Z. Ma, "An approach for approximate Cosine Similarity in fuzzy RDF graph," Fuzzy Sets and Systems, vol. 376, pp. 106-126, 2019.

45 Y. Liu, H. Gao, X. Kang, Q. Liu, R. Wang, and Z. Qin, "Fast Community Discovery and Its Evolution Tracking in Time-Evolving Social Networks," in 2015 IEEE International Conference on Data Mining Workshop (ICDMW), pp. 13-20, 2015.

46 Zachary W W. An information flow model for conflict and fission in small groups. Journal of anthropological research, Vol 33(4), pp. 452-473, 1977.

47 Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms. Physical review E, Vol 78(4): 046110, 2008.

48 Zhu, X , and Z. Ghahramani . "Learning from labels and unlabeled data with label propagation." Tech Report 3175, Vol (2002):237-244, 2004.

49 I. Messaoudi and N. Kamel, "A multi-objective bat algorithm for community detection on dynamic social networks," Applied Intelligence, vol. 49, no. 6, pp. 2119-2136, 2019.

50 G. Qian, S. Sural, Y. Gu, et al. "Similarity between euclidean and cosine angle distance for nearest neighbor queries"//Proceedings of the 2004 ACM Symposium on Applied Computing. ACM, Nicosia, Cyprus, March 14-17, pp. 1232-1237, 2004.

51 L. L. Shi, L. Liu, Y. Wu, L. Jiang, and J. Hardy, "Event Detection and User Interest Discovering in Social Media Data Streams," IEEE Access, vol. 5, pp. 20953-20964, 2017.

52 J. Ge, L. L. Shi, L. Liu, and X. Sun, "User Topic Preferences based Influence Maximization in Overlapped Networks," IEEE Access, vol. 8, pp. 141328-141345, 2019.

53 S. Tang, S. Yuan, and Y. Zhu, "Deep Learning-Based Intelligent Fault Diagnosis Methods Toward Rotating Machinery," IEEE Access, vol. 8, pp. 9335-9346, 2020.

54 C. Y. Mao, J. F. Chen, D. Towey, J. F. Chen, X. Y. Xie, "Search-based QoS ranking prediction for web services in cloud environments", Future Generation Computer Systems, Vol. 50, pp. 111-126, 2015.

55 H. Lu, S. Liu, H. Wei, and J. Tu, "Multi-kernel fuzzy clustering based on auto-encoder for fMRI functional network," Expert Systems with Applications, vol. 159, pp. 113-123, 2020.

56 J. Gou, B. Hou, Y. Yuan, W. Ou, and S. Zeng, "A new discriminative collaborative representation-based classification method via l2 regularizations," Neural Computing and Applications, vol. 32, pp. 9479–9493, 2020.

57 J. Gou, L. Wang, B. Hou, J. Lv, Y. Yuan, and Q. Mao, "Two-Phase Probabilistic Collaborative Representation-Based Classification," Expert Systems with Applications, vol. 133, pp. 9-20, 2019.

58 T. Guan, F. Han, and H. Han, "A Modified Multi-Objective Particle Swarm Optimization Based on Levy Flight and Double-Archive Mechanism," IEEE Access, vol. 7, pp. 183444-183467, 2019.

59 H. Peng, J. Li, Q. Gong, S. Wang, L. He, B. Li, et al., "Hierarchical Taxonomy-Aware and Attentional Graph Capsule RCNNs for Large-Scale Multi-Label Text Classification," IEEE Transactions on Knowledge and Data Engineering, in press, 2019.

60 H. Peng, R. Yang, Z. Wang, J. Li, and R. Ranjan, "LIME: Low-Cost and Incremental Learning for Dynamic Heterogeneous Information Networks," IEEE Transactions on Computers, in press, 2021.

61 Peng, Hao, et al. "Streaming Social Event Detection and Evolution Discovery in Heterogeneous Information Networks." ACM Transactions on Knowledge Discovery from Data, in press ,2021.

62 Xiaozhu Liu, Rongbo Zhu, Ashiq Anjum, Jun Wang, Hao Zhang, Maode Ma, Intelligent Data Fusion Algorithm Based on Hybrid Delay-aware Adaptive Clustering in Wireless Sensor Networks, Future Generation Computer Systems, vol. 104, pp. 1-14, 2020.

63 Xiaozhu Liu, Zhiyang Xiao, Rongbo Zhu, Jun Wang, Lu Liu, Maode Ma, Edge Sensing Data-imaging Conversion Scheme of Load Forecasting in Smart Grid, Sustainable Cities and Society, vol. 62, pp. 1-10, Nov. 2020.

64 Xiaozhu Liu, Rongbo Zhu, Brian Jalaian, Yongli Sun, Dynamic Spectrum Access Algorithm Based on Game Theory in Cognitive Radio Networks, Mobile Networks and Applications, vol. 20, no. 6, pp. 817-827, Dec, 2015.