Knowledge-Based Systems 229 (2021) 107359

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A key elements influence discovery scheme based on ternary association graph and representation learning

Qian Li [a,b], Meiling Li [a], Xu Shi [a], Bin Wu [b], Yunpeng Xiao [a,*]

[a] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China
[b] School of Computer Science and Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China

ARTICLE INFO

Article history:
Received 4 May 2021
Received in revised form 13 July 2021
Accepted 2 August 2021
Available online 6 August 2021

Keywords:
Social networks
Hotspot topic
Key elements influence
Representation learning
Ternary association graph

ABSTRACT

Key elements refer to the elements that play a crucial role in disseminating information in social networks. The influence discovery of key elements can guide a series of works, such as public opinion control, user recommendation, and marketing promotion. Recently, there have been many studies on the influence of elements, but at present, many methods focus on either the influence discovery of key elements of different types or the dynamic influence discovery of a certain type of element alone, and rarely consider the combination of the two. Therefore, this study proposes a key element discovery algorithm based on a ternary association graph and representation learning, which can detect the influence of paths, users, and user groups. Additionally, the changes of different types of key elements can be analyzed according to the influence of elements in each stage of topic communication.

1. Introduction

With the rise of social platforms, such as microblog, Twitter, and blogs, people's social life has been greatly enriched [1]. There are many topics that users may pay attention to through social networks. In these topics, users can communicate anytime and anywhere; however, false and illegal information is also disseminate. The wanton dissemination of this information can destroy the security of public discourse space and may bring a negative impact on society [2]. Therefore, it is important to determine the influence of the elements on certain topics, find some key elements that play a role in promoting topic communication. It is also essential to analyze the changes in key elements in different stages of topic communication for assessing development trend of the topic. Additionally, elements influence discovery has certain application value in user recommendation, marketing promotion and other aspects.

We found an interesting phenomenon, in the process of social network topic dissemination, the influence of elements often has explicit and implicit relevance. As shown in Fig. 1, user A is just an ordinary user, but he participates in multiple user groups and multiple paths for topic propagation. Through the element relevance influence discovery algorithm to enhance its influence, user A becomes an important user more close to the real situation.

At this time, it is more accurate to increase the user's influence by taking into account the relevance of elements. In summary,

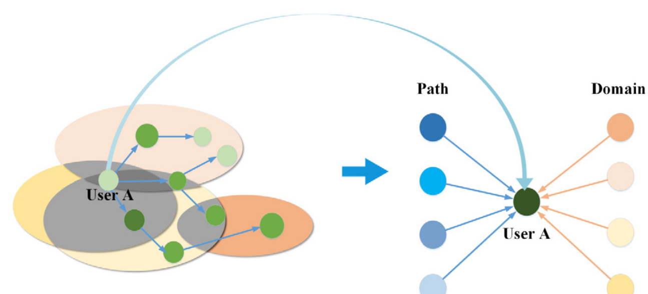


Fig. 1. A simple topic network example.

although the current research has achieved quite good results, there are still several challenges as follows:

- The relevance of elements. The topic propagation process involves many elements. How to describe the interaction between different elements is very effective for network public opinion supervision. However, many traditional studies lack unity and convenience because they are relatively independent in mining these elements, and these elements are not closely linked.
- The diversity of elements and the phenomenon of topic communication among users across multi-user groups. In social networks, how to the relevance of mine multiple elements influences discovery. Simultaneously, in the process of hot topic dissemination, there is usually the same user

* Corresponding author.
E-mail address: xiaoyp@cqupt.edu.cn (Y. Xiao).

https://doi.org/10.1016/j.knosys.2021.107359
0950-7051/

carrying out topic dissemination across multi-user groups. However, the current research ignores the phenomenon of users carrying out topic dissemination across multi-user groups and does not maximize the influence of such users on topic dissemination, the accuracy of identifying key users is reduced.
- The dynamics of influence evolution. Currently, the related research only simply excavates the topic source or key elements but does not describe how the topic links from one important user to another important user at different stages, and how it links from one important user group to another important user group.

Aiming at these challenges, this paper proposes an algorithm named PUGRank to discover overlapping user groups using representation learning to represent the user relationship network. Then, the ternary association graph model is used to associate different type of elements to discover the influence of elements. Additionally, we analyze the changes in key elements at different stages of topic communication.

Our contributions can be summarized as follows:

- Use the ternary association graph to determine the relevance of element influence. We propose a ternary association graph model based on the "path–user–user group" to represent the association of key elements simply. Aiming at the relevance problem of different types of elements, the study uses the ternary association graph to associate different elements simply and effectively.
- Aiming at the characteristics of users' topic propagation across multi-user groups, combined with the user relationship network, we propose an overlapping user group discovery algorithm for mining user groups. The accuracy of influence discovery is improved by associating the resulting user groups with other types of elements.
- Aiming at the dynamic problem of influence evolution, this paper describes the dynamic evolution law of the influence of key elements in the entire life cycle of topic communication in time discretization to realize the dynamic mining of the influence of key elements.

The organization of this paper is constructed as follows: Section 1 introduces the research background and status. Section 2 discusses related work. Section 3 we present basic definitions and issues. Section 4 presents the proposed method and related learning algorithms. In Section 5, we use real datasets for experiments and present experimental results. Section 6 presents the conclusion.

2. Related work

In the study of social topic, researchers have made many targeted improvements on existing methods and achieved remarkable results [3–5]. This paper focuses on the above challenges in the first chapter and analyzes the existing results.

Aiming at the relevance problem of elements, Wu et al. [6] used a heuristic search method to generate the user's topic behavior influence tree. They identified influential users by maximizing the number of affected users and minimizing the propagation path. Zareie et al. [7] considered the common hierarchical structure of nodes and the set of neighbor nodes to determine the key nodes. Zhao et al. [8] proposed a new Bayesian decomposition model to quantify users' preferences, authors' influence, and potential topic elements influenced by content. Li et al. [9] proposed the CoRank method considering the users and Twitter of social networks, and the complex relationship between them. Giannoulakis et al. [10] used the HITS algorithm for mining the scores of key tags corresponding to images. Most of the methods achieves performances; however, many methods consider only one type of element or based on artificially quantifying the weights of various relationships, which may lead to the inaccuracy of the results.

In view of the diversity of elements and the phenomenon of topic communication among users across multi-user groups, Cheng et al. [11] proposed an algorithm of overlapping community discovery based on node vitality considering the evolution, activity and multi-scale of nodes. Lu et al. [12] proposed an overlapping community algorithm for label propagation used on large complex networks. Gema et al. [13] proposed a new multi-objective genetic algorithm that uses measures related to network connectivity to detect overlapping communities. Li et al. [14] found communities by local diffusion of known seeds. Due to the great success in the graph neural network field [15], many graph representation learning algorithms [16] have been developed, such as GraphSage [17], GAT [18] and some dynamic graph representation learning algorithms [19,20] (e.g., EvolveGCN [21]). Therefore, some scholars have proposed propose an overlapping user group detection algorithm based on representation learning. For example, Zhang et al. [22] designed a MOEA based on mixed representation for overlapping community detection. Many methods achieves performances; however, few of them integrate the local and global influences of users for overlapping community analysis.

For the dynamics of influence evolution, most scholars partially investigated the similarity of influence evolution. Wu et al. [23] predicted user preferences in topics by integrating user's direct and potential behaviors into the learning model. Zhang et al. [24] combined the vocabulary co-occurrence model to construct a topic evolution model suitable for short texts in social media to effectively monitor the evolution of topics in short texts of social media. Deng et al. [25] proposed an interactive topic modeling method to dig public expression dynamics after a major explosion by visualizing the evolution of topics through social software. Although the above methods conducted some studies on the evolution characteristics of hot topics by establishing a reasonable model, they are used only for simple key user discovery. Additionally, they do not analyze the information of key elements in the hot topic propagation process in the entire topic life cycle, which lacks dynamism.

3. Problem formulation

3.1. Related definitions

This study mainly solves the problem of the influence discovery of hot topics. Through the comprehensive analysis of hot topic network N, user relational networks N_U, we develop the "path–user–user group" ternary association graph for determining the influence of each type of element in the hot topic network. Additionally, the evolution of key elements in the entire life cycle is dynamically mined. The basic concepts used in this paper are defined as follows:

Definition 1. Topic network communication time stages $T = \{T_1, T_1, \ldots, T_L\}$.

For a complete hot topic network, the total period T of topic communication is divided into several stages $\{T_1, T_1, \ldots, T_L\}$.

Definition 2. User relation network $N_U = \{U, E\}$.

Where, U is the user collection, there are $u_i \in U, u_j \in U, e_{ij} \in E$. If the user u_i follows the user u_j, $e_{ij} = 1$, otherwise $e_{ij} = 0$.

Definition 3. Path $p_i^{T_l} = \{u_k\}$.

In the topic network, at T_l stage, the link relationship that the topic forwards from the starting user node to the leaf user node is a propagation path in the topic network, this paper uses u_k to represent the users set in path $p_i^{T_l}$, where k indicates the propagation user number. The set of paths at T_l stage can be written as $P^{T_l} = \{p_i^{T_l}\}$.

Definition 4. User group $g_k^{T_l} = \{u_m\}$.

In the topic network, at T_l stage, the set of participating users is called the propagation user group, this paper uses $g_k^{T_l} = \{u_m\}$ to represent. Where, m indicates the user number. The set of user groups at T_l stage can be written as $G^{T_l} = \{g_k^{T_l}\}$.

3.2. Formalizing the problem

To describe our solved problems, the life cycle of topic networks N is divided into different stages, the paths set P^{T_l} is extracted, and the "path–user" binary association graph model $M_{P,U} = \left\{P^{T_l} \cup U^{T_l}, A\right\}$ is constructed. Here, A is the relationship matrix between paths and users. According to the user relationship network $N_U = \{U, E\}$, combined with the overlapping community discovery algorithm, the "user–user group" binary association graph model $M_{U,G} = \left\{U^{T_l} \cup G^{T_l}, B\right\}$ is constructed. Here, B is the relationship matrix between users and user groups. Then, with topic users as the bridge, the topic ternary association graph model $M_{P,U,G} = \left\{P^{T_l} \cup U^{T_l} \cup G^{T_l}, A \cup B\right\}$ is constructed. We initialized the influence of paths, users and user groups to get X_0, Y_0, Z_0 and scored iteratively according to the transfer matrix between different type of elements to mine the influence score vector of key elements X', Y', Z' in the topic propagation process. Finally, by synthesizing the influence of key elements and the multi-stage characteristics of topic propagation, we dynamically evolve how hot topics link from one key user to the next key user (i.e., L_{key_U}), and how hot topics link from one key user group to the next key user group (i.e., L_{key_G}). A clearer definition of the problem is as follows.

$$
\left.\begin{array}{c} \left.\begin{array}{l} N \to P^{T_l} \to M_{P,U} \\ N_U \to G^{T_l} \to M_{U,G} \end{array}\right\} M_{P,U,G} \\ T \end{array}\right\} \xrightarrow{X', Y', Z'} \Rightarrow \begin{array}{l} L_{key\_U} = \left\{u_1 \to u_2 \to \ldots \to u_j, T_l\right\} \\ L_{key\_G} = \{g_1 \to g_2 \to \ldots \to g_m, T_l\} \end{array} \tag{1}
$$

3.2.1. Problem input

Based on the above definition, the input of the proposed method is described as follows.

1. Topic network communication time stages $T = \{T_1, T_1, \ldots, T_L\}$;
2. Hot topic network N;
3. User relationship network $N_U = \{U, E\}$.

3.2.2. Problem output

Based on the above description, we used the model proposed in this paper to obtain the following output:

1. User groups set G^{T_l}.
2. Score vector of element influence in the topic communication process X', Y', Z'.
3. The link relationship from one key user to the next key user (L_{key_U}), and how to spread hot topics from one key user group to another (L_{key_G}) during the entire communication life cycle.

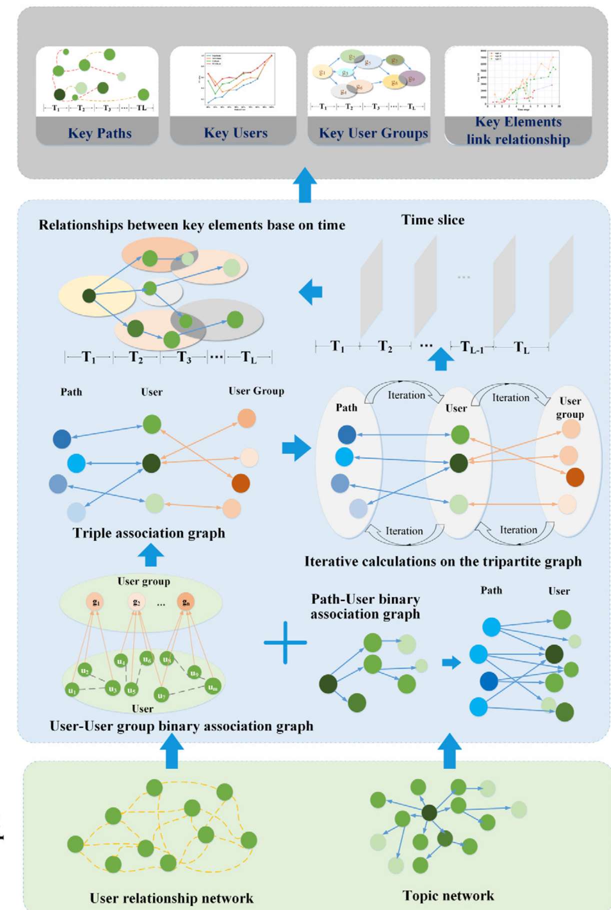


Fig. 2. Model framework.

4. Proposed method

To solve the problems raised in Chapter 3, we proposes a new influence discovery scheme of different elements based on a ternary association graph and representation learning based on the path, user, and user group of the entire life cycle in the topic network. This scheme is implemented in three steps. In the first step, we use an overlapping user group discovery algorithm to obtain user group and build "user–user group" association model. In the second stage, we construct a graph model based on "path–user–user group". In the third stage, we propose a cross iteration scoring algorithm for determining the influence of different elements based on a ternary association graph and identify the influence score vector of each element. Simultaneously, the dynamic analysis of the communication situation of key elements in the entire topic life cycle is conducted. Fig. 2 shows the specific flow.

4.1. Two binary association graphs

4.1.1. Path–user binary association graph

According to the timeliness of hot topic communication, this paper takes the time sequence of users' forwarding behavior as the starting node in the hot topic communication process. It also extracts and analyzes the path in the entire life cycle of the topic. When participating users propagate the topic, the initiator of the topic can be regarded as the root node, each forwarding user can be regarded as a leaf node, and each forwarding behavior is regarded as an edge between the participants. Then, all the

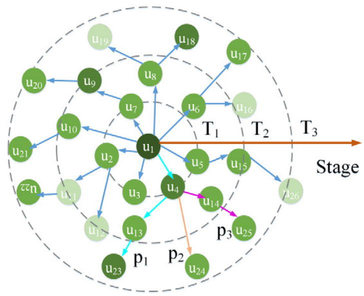


Fig. 3. Propagation path analysis graph.

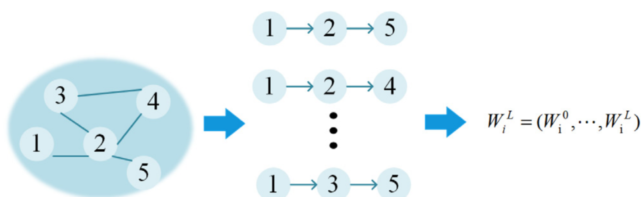


Fig. 4. Deepwalk algorithm.

forwarding behaviors of the participating users can form a topic propagation tree model. The root and leaf nodes are the starting and end points, respectively. Each specific path of topic communication in the whole life cycle can be determined. As shown in Fig. 3. After getting the extracted paths set, the "path–user" binary association graph model is constructed.

4.1.2. User–user group binary association graph

Aiming at the problem of user group discovery in the hot topic propagation process, we propose an overlapping user group discovery algorithm based on DeepWalk. We obtain the overlapping user groups from the perspective of user similarity measurement. The user similarity measurement is divided into global structure feature-based measurement and local structure feature-based measurement.

First, we introduce the measurement based on the global structure feature: Since the users in a user group are closely connected, the distance between users in a user group is generally small. Therefore, we can learn the embedding of nodes in the user relationship network using the DeepWalk algorithm [26] to calculate the distance between users and help to better discover user groups.

DeepWalk learns the social representation of a network through truncated random walk. The random walk here refers to taking the starting point randomly on the graph network; then, fixing the length of the path and selecting the random walk path repeatedly. As shown in Fig. 4, first, the initial node is selected, and the next hop is random and adjacent to the current node. From the selected neighbor node to the next hop neighbor node, the process is repeated until the fixed path length is completed.

For better analysis, it is assumed that the starting node of the generated random walk sequence is a random walk path with the length of node W_i fixed as L, and other nodes on the path can be expressed as $W_i^L = (W_i^0, \dots, W_i^L)$, where $W_i \in N$.

DeepWalk algorithm calculates the probability of the current path downstream to the next hop, namely $\Pr(W_n|W_1, L, W_{n-1})$. In the Word2vec model, the skip-gram model does not use context to predict the missing words but uses missing words to predict context. It can consider the left and right windows simultaneously. It does not consider the order and position of words in the window. Therefore, when skip-gram is applied, the objective function can be changed to:

$$\underset{\phi}{minmize} \ -\log \Pr(\{W_{i-L}, \dots, W_{i-1}, W_{i+1}, \dots, W_{i+L}\}|\phi(W_i)) \tag{2}$$

Where, $\phi(W_i)$ is the given path node W_i.

The embedding of each user $u(i)$ in the user relationship network is obtained from the DeepWalk algorithm is applied to the overlapping user group discovery algorithm. A new user embedding is obtained by splicing the user's in and out degree with the user's embedding $u(i)$:

$$ds(i) = [\deg_{in}(i), \deg_{out}(i), u(i)] \tag{3}$$

Where, $\deg_{in}(i)$ is the in-degree of node i, $\deg_{out}(i)$ is the out-degree of node i, and $u(i)$ is the user vector representation of DeepWalk. Therefore, the user similarity based on global structure features can be calculated as follows:

$$\upsilon(i,j) = \frac{ds(i) \cdot ds(j)}{\|ds(i)\| \|ds(j)\|} \tag{4}$$

The influence of users based on local structure features can be reflected by their neighbors, and the number of common neighbors can reflect the closeness of nodes and neighbors. If the node is closer to its neighbors, the node will have more influence on its neighbors. The user influence based on local features can be calculated as follows:

$$\omega(i,j) = \frac{N(i) \cap N(j)}{N(i) \cup N(j)} \tag{5}$$

Therefore, the comprehensive similarity measure of users i and j can be expressed as follows:

$$sim(i,j) = \upsilon(i,j) + \omega(i,j) \tag{6}$$

The comprehensive influence of users can be calculated as follows:

$$In(i) = \sum_{j \in N(i)} \deg_{in}(i) \cdot \deg_{in}(j) \cdot sim(i,j) \tag{7}$$

Where, $\deg_{in}(i)$ is the in degree of the user i in the user relationship network, $sim(i,j)$ is the similarity between the user i and the user j, and $j \in N(i)$ indicates that the user j is the neighbor node of the user i.

We find the largest Top-k $In(i)$ users as the seed node of overlapping user group discovery. Therefore, the initialization community can be expressed as $G = \{g_1, g_2, \dots, g_k\}$.

For the remaining nodes that are not allocated, the attribution value of user node m for each user group g_k is calculated as follows:

$$cls(m, g_k) = \sum_{t \in g_k} sim(m,t) \tag{8}$$

Where, $t \in g_k$ represents the user t in the user group g_k.

The user joins the user groups whose $cls(m, g_k)$ value is greater than the threshold κ. If the $cls(m, g_k)$ value of node to any user group is not greater than κ, the user randomly selects a user group to join.

Finally, if the repetition rate of the user groups is too high, the user groups need to be merged. User group repetition rate can be defined as follows:

$$\phi = \frac{N(g_i) \cup N(g_j)}{N(g_i) \cap N(g_j)} \tag{9}$$

Where, $N(g_i)$ represents the nodes in g_i. When ϕ is greater than 0.5, the two user groups will be merged.

Through the above overlapping user group discovery algorithm to get the user groups set, as shown in Fig. 5, we can build the "user–user group" binary association graph.

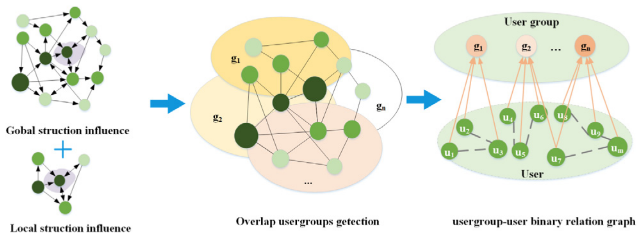


Fig. 5. Construct "user–user group" binary association graph.

4.2. Ternary association graph

To conveniently identify and analyze the relationship between different elements (path, user and user group), this paper draws on the ternary association graph based on the above "path–user" binary association graph and "user–user group" binary association graph. It divides all elements in the topic network into three disjoint subsets. The relationship between the two different type of elements is regarded as a set of weight matrix, as shown in Fig. 6, which is a ternary association graph of the hot topic network, and it can be represented by $M_{P,U,G} = \{P^{T_l} \cup U^{T_l} \cup G^{T_l}, A \cup B\}$. Among them, P^{T_l} is the set of propagation paths, U^{T_l} is the set of participating users, and G^{T_l} is the set of user groups. A is the weight matrix between paths and users, and B is the weight matrix between users and user groups. If there is an association between path i and user j, transition probability $a_{ij} > 0$ or $a_{ji} > 0$ in the weight matrix A, otherwise, $a_{ij} = 0$ or $a_{ji} = 0$. If there is an association between user j and the user group k, transition probability $b_{jk} > 0$ or $b_{kj} > 0$ in the weight matrix B, otherwise, $b_{jk} = 0$ or $b_{kj} = 0$. Suppose the number of propagation paths is I, the number of participating users is J, and the number of user groups is K. The weight matrix A is divided into $A_{i,j} = \left[a_{ij}\right]_{I\times J}$ and $A_{j,i} = \left[a_{ji}\right]_{J\times I}$, and the weight matrix B is divided into $B_{j,k} = \left[b_{jk}\right]_{J\times K}$ and $B_{k,j} = \left[b_{kj}\right]_{K\times J}$.

If the matrix T is an element transition probability matrix, the value of the element $t_{i,j}$ in the matrix is as follows:

$$t_{i,j} = \begin{cases} \frac{1}{deg_{out}(i)}, \textit{if}\, e_{ij} \in E \\ 0, \textit{otherwise} \end{cases} \tag{10}$$

Where, $\frac{1}{deg_{out}(i)}$ represents the degree of node i, $e_{ij} \in E$ represents there is an association between node i and node j. Similarly, we can get $t_{j,i}$.

4.3. Iterative scoring strategy

4.3.1. Ternary association graph initialization

In this paper, we first assign an initial value to each node of each element on the ternary association graph model. The X vector represents the importance of the path nodes; Y vector, the importance of the users; and Z vector, the driving force of the user groups. We initialize X, Y, Z to X_0, Y_0, Z_0.

1. Calculating the importance of the path i at the T_l stage:

The importance of the propagation path in the social network is determined by the number of participating users. Generally, the more participating users, the more important the propagation path. Therefore, this paper uses $X^{(i)}$ to measure the importance of the propagation path i at the T_l stage of hot topic propagation. The calculation formula is as follows:

$$X^{(i)} = N_{ps\_num} \tag{11}$$

Where, N_{ps_num} represents the total number of participating users in the propagation path i at the T_l stage.

2. Calculating the importance of the user j at the T_l stage:

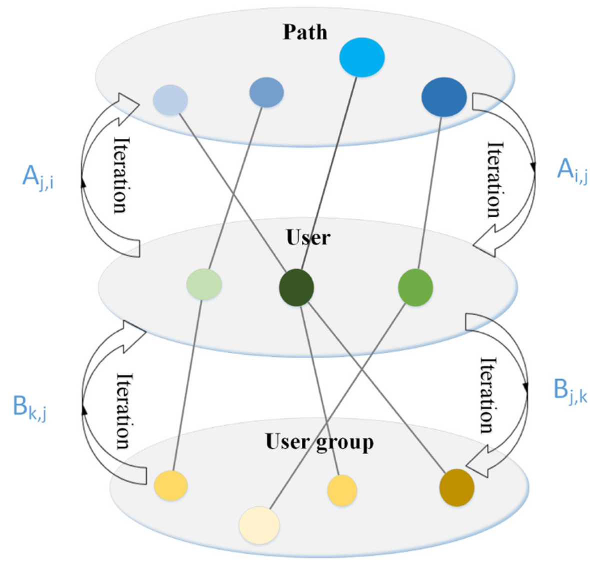


Fig. 6. Construct ternary association graph.

According to Eq. (7):

$$Y^{(j)} = In(j) \tag{12}$$

3. Calculating the importance of the user group k at the T_l stage:

The influence of the users in the hot topic network is calculated as follows:

$$Z^{(k)} = \sum_{u_i \in g_k} In(u_i) \tag{13}$$

Where, $u_i \in g_k$ indicates that the user u_i exists in the user group g_k.

Therefore, according to the above calculation of the importance of paths, users and user groups, we observe that at the T_l stage of hot topic propagation, the calculation formulas of the initial influence score vector of paths, users and user groups are as follows:

$$X_0 = \left[\frac{X_0^{(1)}}{\sum_i^I X_0^{(i)}}, \frac{X_0^{(2)}}{\sum_i^I X_0^{(i)}}, \frac{X_0^{(3)}}{\sum_i^I X_0^{(i)}}, \ldots, \frac{X_0^{(i)}}{\sum_i^I X_0^{(i)}} \right]^T \tag{14}$$

where, I is the number of total paths.

$$Y_0 = \left[\frac{Y_0^{(1)}}{\sum_j^J Y_0^{(j)}}, \frac{Y_0^{(2)}}{\sum_j^J Y_0^{(j)}}, \frac{Y_0^{(3)}}{\sum_j^J Y_0^{(j)}}, \ldots, \frac{Y_0^{(j)}}{\sum_j^J Y_0^{(j)}} \right]^T \tag{15}$$

where, J is the number of total users.

$$Z_0 = \left[\frac{Z_0^{(1)}}{\sum_k^K Z_0^{(k)}}, \frac{Z_0^{(2)}}{\sum_k^K Z_0^{(k)}}, \frac{Z_0^{(3)}}{\sum_k^K Z_0^{(k)}}, \ldots, \frac{Z_0^{(k)}}{\sum_k^K Z_0^{(k)}} \right]^T \tag{16}$$

where, K is the number of total user groups.

4.3.2. Iterative scoring

The idea of the cross-iteration scoring algorithm is that if a node is a hub node, which can be easily accessed by other nodes, then it has a high authority value. If a node has a high authority value, then it is likely to be an important node. Therefore, we can use the idea of the cross-iteration scoring algorithm to consider the interaction relationship on the ternary association graph to determine the final influence score of the three kinds of key element nodes. Using the initial value and transition

Table 1
Experimental data of different topics.

Data item	topic A	topic B	topic C
Number of user	3504	7022	9626
Number of friends	847,500	879,780	534,459
Number of retweeting paths	4027	5479	9568
Number of messages	3476	1948	1369
Propagation time	2013.12.30–2014.1.2	2014.5.13–2014.5.16	2014.9.4–2014.9.7

probability matrices of each type of elements obtained above, the forward scoring and backward scorings are conducted. The specific process is as follows:

In the forward scoring process, first, based on the users influence score vector Y after the last iteration, starting from the score vector X of the influence of the paths, obtain the new influence vector Y' of the participating users by transforming of weight matrix $A_{i,j}$. Then, based on the influence vector Z of user groups influence after the last iteration, starting from the users influence vector Y', obtain the new influence vector of user groups Z' by transforming of weight matrix $B_{j,k}$. The calculation formula is as follows:

$$Y' = \lambda Y + (1-\lambda){A_{i,j}}^T X \tag{17}$$

$$Z' = \lambda Z + (1-\lambda){B_{j,k}}^T Y' \tag{18}$$

In the backward scoring process, the steps are the same as the forward scoring process; thus, we ignore the explanation. The formula is as follows:

$$Y' = \lambda Y + (1-\lambda){B_{k,j}}^T Z' \tag{19}$$

$$X' = \lambda X + (1-\lambda){A_{j,i}}^T Y' \tag{20}$$

Where, λ is the damping coefficient, X, Y, Z represents the influence score vector after the last iteration (At the process of the first forward score, let $X = X_0$, $Y = Y_0$, $Z = Z_0$). Additionally, the forward iterative and the backward iterative voting alternated in this mechanism. When the total variation between the influence score vector of each element after this iteration and the influence score vector of each element after the last iteration is less than the threshold ε, the iterative process is immediately terminated and returned X', Y', Z'. To ensure the final convergence effect of the mechanism, each time a round of backward scoring process is completed, the influence score vectors of paths, users, and user groups in the mechanism are simultaneously normalized.

4.4. Algorithm

First, based on the user relationship network, the overlapping user group discovery algorithm is implemented. The algorithm flow is described in Algorithm 1. Then, the three types of elements of the path, user and user group, and their relationships are integrated to develop the "path–user–user group" ternary association graph. The influence score vector of elements in the process of hot topic propagation is calculated by cross iteration scoring on the ternary association graph model. Finally, according to the idea of time discretization and the evolution law of key elements in the hot topic propagation process is identified. The algorithm flow is described in Algorithm 2.

In addition, the complexity of the algorithm is also an important factor to be considered in this paper. First, assuming the number of users is n, the complexity required to build the "path–user" binary association graph is $O(n)$, and the complexity required to build the "user–user groups" binary association graph is $O(n^2)$. (Among them the highest complexity is the need for similarity calculation among users is $O(n^2)$, so the complexity of this part of the algorithm can be regarded as $O(n^2)$). Second, when calculating the weight matrix according to the relation of each subset in the model of the ternary association graph, the complexity of the algorithm is $O(kn)$, where k is the number of the weight matrix. Finally, the score vector of the final key elements influence is obtained from the cross-iteration scoring on the ternary association graph model. The complexity required is $O(m(|X|+2|Y|+|Z|))$, where m is the number of iterations, and $|X|$ is the number of paths, $|Y|$ is the number of users, and $|Z|$ is the number of user groups.

The total complexity of the proposed PUGRank is $O(n)+O(n^2)+O(kn)+O(m(|X|+2|Y|+|Z|))$-$O(n^2)$. Since our method aims at social topic networks, the number of elements in the topic network is generally not too many. We will also deal with the invalid element nodes of the topic network properly. Therefore, the complexity of the algorithm is acceptable.

5. Experiments and analysis

5.1. Experimental setting

In this section, we describe the experimental settings. First, we describe the experimental data. Next, the baseline methods compared in the experiments are introduced. Finally, several evaluation metrics are proposed to evaluate the performance of the model and the experimental results are analyzed.

5.1.1. Dataset

The dataset of this paper was obtained from Tencent Weibo dataset compiled by our team [27]. In the second quarter of 2012, Tencent Weibo had 469 million registered users [28]. We selected three hotspots: topic A from Feng Xiaogang's film Personal Tailor; topic B from Dad, Where Are We Going Season 2; and topic C from Rare Blood Type. The specific information is presented in Table 1.

5.1.2. Baseline methods

In order to verify the superiority of our PUGRank algorithm, we selected some classic and advanced influence discovery algorithms as baselines.

- PageRank. This algorithm has been widely used in numerous ranking applications.
- CoRank [9]. This algorithm is an improved algorithm based on the PageRank algorithm.
- MPURank [29]. This algorithm is an improved algorithm based on the HITS algorithm.

PageRank algorithm considers the single type relationship of the same element, while CoRank algorithm considers different types of relationships and different types of elements (including retweet, post, posted, follow and mention relationship). The MPURank algorithm also considers different types of elements and different types of relationships (including retweet, post, posted relationships). This dataset consists of retweet, post, posted, follow relationships. These algorithms do not dynamically mine the influence of elements.

Algorithm 1 Overlapping user group discovery algorithm

Input: Time stages $T = \{T_1, T_2, T_3, ...T_L\}$;user relationship network $G_U = \{U, E\}$
Output: User influence score set $\{In(j)\}$; User groups set G^{T_l};
// Calculating user influence
Using DeepWalk to get the embedding of users in user relationship network;
for user j in users set:
  for other user in users set:
  // User tag initialization
    flag[j] = flase (means unassigned) ;
    According to Eqs. (2)–(6), the comprehensive similarity of users is calculated;
    the comprehensive influence of user j is obtained according to Eq. (7);
// Get the Top-k influential users
The comprehensive influence of users is ranked, and the ranking result of Top-k user is obtained as the seed node;
// User groups initialization
for user s in Top-k user:
  flag[s] = true (means assigned) ;
// Assign unassigned users to the user groups
for user m in unassigned uses:
  for group g_k in user groups G:
    Calculate $cls(m, g_k)$ according to Eq. (8);
    if $cls(m, g_k) > \kappa$:
      append user m to g_k;
      flag[m] =true(means assigned);
for user j in users set:
  if flag[j] = false:
    randomly select a user group to join;
// Merge user groups with high repetition rate
for group g_k in groups G:
  for other group g_q in groups G:
    if $\phi > 0.5$ according to Eq. (9):
      merge two user groups;
return user influence score set $\{In(j)\}$ and user groups G^{T_l};

Algorithm 2 PUGRank algorithm

Input: user groups set G^{T_l} and users influence score set $\{In(j)\}$; paths set P^{T_l}; $T = \{T_1, T_2, T_3, ...T_L\}$
Output: Key element score sequence X', Y', Z'; Key elements link relationship L_{key_U} and L_{key_G};
// Construction and analysis of ternary association diagram
Extract the key elements influence set P^{T_l}, U^{T_l}, G^{T_l};
The triple association graph model $M_{P,U,D}$ is constructed by $M_{P,U}$ model and $M_{U,D}$ model;
The transition probability matrix $A_{i,j}$, $A_{j,i}$, $B_{j,k}$, $B_{k,j}$ is calculated according to Eq. (10);
Calculate the initial score vectors X_0, Y_0 and Z_0 of key elements influence according to Eq. (11)–Eq. (16);
// The initialization of cross iterative scoring mechanism
Initialize X=X_0, Y=Y_0, Z=Z_0, λ and ε;
Do
  // The process of positive scoring
  Calculate the score vector Y' of the users' influence according to Eq. (17);
  Calculate the score vector Z' of the user groups' influence according to Eq. (18);
  // The process of reverse scoring
  Calculating the score vector Y' of the user' influence according to Eq. (19);
  Calculating the score vector X' of the paths' influence according to Eq. (20);
  Normalize X', Y', Z';
  // Save the results of each iteration
  Save the iteration results of this round: $X = X'$, $Y = Y'$, $Z = Z'$;
While $\left\|X' - X + Y' - Y + Z' - Z\right\| \geq \varepsilon$;
Return X', Y', Z', L_{key_U} and L_{key_G};

5.1.3. Evaluation metrics

To confirm the effectiveness of the algorithm proposed in this paper, we used different metrics to evaluate the recognition effect of different key elements:

- Key path evaluation metrics:
  Node coverage of the path(NCR):
  In social networks, the number of participating users is an essential indicator of the importance of the topic's propagation path. Since different propagation paths contain different numbers of participating users, we use the node coverage metrics of the path to evaluate the result of Top-k key paths.
  Node coverage of the path: measuring the importance of the path in hot topic network, which is defined as follows:

$$NCR(p_i) = \frac{\sum_{u_k \in p_i} ret(u_k)}{N} \times 100\% \tag{21}$$

  where, $\sum_{u_k \in p_i} ret(u_k)$ is the count of user u_k who retweet the path p_i, and N is the total number of users in the topic network.
- Key user evaluation metrics: The Precision rate, Recall rate, and F1 value were used to evaluate the experimental results.
  1. Precision, Recall and F1 value:
  We the Precision, Recall, and F1 value detection algorithm to verify the effectiveness of the key users influence discovery results. Suppose that there are given four algorithms A, B, C and D, the ranking sequence of key elements with length N obtained by each algorithm is $I_N(A)$, $I_N(B)$, $I_N(C)$ and $I_N(D)$. Let the key elements of the three algorithms exist as the trusted key element reference results, then the definition of cross-validation sequence set is as follows:

$$\begin{aligned} I_N = {} & (I_N(A) \cap I_N(B) \cap I_N(C)) \cup (I_N(A) \cap I_N(B) \cap I_N(D)) \\ & \cup (I_N(A) \cap I_N(C) \cap I_N(D)) \cup (I_N(B) \cap I_N(C) \cap I_N(D)) \end{aligned} \tag{22}$$

The Precision of algorithm A is defined as follows:

$$P_N(A) = \frac{|I_N(A) \cap I_N|}{|I_N(A)|} \tag{23}$$

The Recall of algorithm A is defined as follows:

$$R_N(A) = \frac{|I_N(A) \cap I_N|}{|I_N|} \tag{24}$$

Using the Precision and Recall rate, the comprehensive measurement of the F1 value of the algorithm A is performed using the following equation:

$$F1_N(A) = 2\frac{P_N(A) \cdot R_N(A)}{P_N(A) + R_N(A)} \tag{25}$$

2. Node coverage of the user(NCR):
Measuring the importance of the user in hot topic network is defined as follows:

$$NCR(u_j) = \frac{\sum ret(ret(u_j))}{N} \times 100\% \tag{26}$$

where, $\sum ret(ret(u_j))$ is all users affected by user u_j, and N is the total number of users in the topic network.

- Key user group evaluation metrics:
1. Overlapping modularity function(OCM):

$$OCM = \sum_{i=1}^{K} \sum_{u \in c_i, v \in c_i} \frac{1}{Q_u Q_v}(A_{uv} - \frac{k_u k_v}{2m}) \tag{27}$$

where, m is the total number of edges in the network, A is the adjacency matrix of the network, c_i is the user group i, k_u is the degree of node u, Q_u is the number of user groups that node u belongs to, k_u is the degree of node u, A_{uv} is whether there is a link relationship between node u and v, if there is a link relationship, $A_{uv} = 1$, otherwise, $A_{uv} = 0$.
2. Node coverage of the user group(NCR):
Measuring the importance of the user group in hot topic network is defined as follows:

$$NCR(g_m) = \frac{\sum_{u_k \in g_m} ret(u_k)}{N} \times 100\% \tag{28}$$

where, $\sum_{u_k \in g_m} ret(u_k)$ is the sum of users affected by all nodes included in the user group g_m, and N is the total number of users in the topic network.

5.1.4. Experimental parameter settings

In the overlapping user group discovery algorithm, get the Top-300 influential users for initializing user groups, we change the threshold value κ in [0.3, 0.7], and the *OCM* value of overlapping communities under each topic is shown in Fig. 7. It can be seen from the figure that when the threshold κ is set at 0.6, the overall performance of the algorithm is better. Thus, we set the threshold $\kappa = 0.6$.

Set $\varepsilon = 1 \times 10^{-5}$ during the iteration. Given the set of reference algorithms as MPURank, CoRank, and PageRank algorithm. Set N = 10, use the Top-10 user result set of each method to get the cross-validation result set I_{10}. By analyzing the number of the same elements of the Top-10 user result set obtained by our method and the cross-validation result set, we can obtain the ideal value of λ. The results are presented in Table 2. From the above table, we obtain that at the beginning, with the increase in λ value, the number of correct results of our algorithm in the Top-10 users of each topic gradually increases; However, as λ approaches 1, the effect starts deteriorating. Considering the influence finding results, the λ is between 0.5 and 0.6, the algorithm performs better. Therefore, we choose $\lambda = 0.6$.

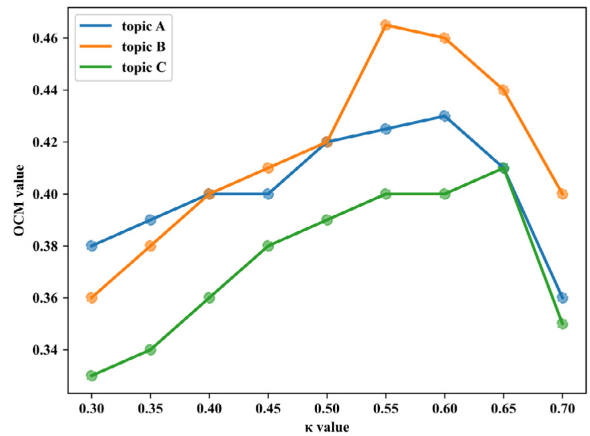


Fig. 7. Parameter setting of κ.

Table 2
Parameter setting of λ.

The value	topic A	topic B	topic C
0	1	0	0
0.1	3	2	1
0.2	6	2	2
0.3	7	4	2
0.4	7	5	4
0.5	9	8	4
0.6	**9**	**9**	**5**
0.7	8	5	4
0.8	7	3	4
0.9	5	2	3
1	1	2	2

5.2. Analysis of the experimental results

In this section, we use three hot topics (topic A, topic B and topic C) to analyze key paths, key users, key user groups and propagation trend in hot topic network. First, we used the proposed algorithm and baseline method to identify the influence results of the different key elements of three hot topics. Then, we used each type of element evaluation index to evaluate the results of elements influence discovery. Finally, different type of key elements can be analyzed according to the influence of elements in each stage of topic communication.

5.2.1. Key paths analysis

We used the PUGRank algorithm is used to identify key paths in the hot topic propagation process. The ranking results of Top-10 paths are shown in Table 3. The relationship between the ranking of the propagation path in the topic and the number of nodes is not exactly proportional. The importance of a propagation path is not only determined by the number of nodes included in the path but also related to the influence of other node affected by the node in the path. It is related to the cross-iterative scoring mechanism of the PUGRank algorithm, and consistent with the real situation of information dissemination in social networks.

Then, the NCR of the path in the three hot topics is calculated, and the performance of the PUGRank algorithm is verified. The result is shown in Fig. 8.

As shown in Fig. 8, as the path ranking increases, the node coverage value of the propagation path shows a downward trend. This shows that the higher the ranking of the path in hot topic network, the higher the node coverage, the more important it is, confirming the effectiveness of the proposed algorithm for path influence identification.

Table 3
Ranking result of key paths on topic A, B and C.

Rank	topic A-PathID	topic A-Nodenum	topic B-PathID	topic B-Nodenum	topic C-PathID	topic C-Nodenum
1	p4009	9	p390	6	p783	10
2	p4008	7	p337	11	p202	12
3	p3805	9	p2743	12	p515	8
4	p2109	8	p336	9	P1364	10
5	p3884	6	p3894	7	p961	10
6	p4016	6	p434	9	p3278	7
7	p3960	9	p278	6	p76	5
8	p2094	5	p2861	5	p633	3
9	p2082	4	p4012	3	p1994	7
10	p4007	4	p995	5	p3573	3

Table 4
Ranking result of key users on topic A.

Rank	PageRank	MPURank	CoRank	PUGRank
1	u317	u240	u265	u240
2	u240	u265	u240	u317
3	u915	u1048	u322	u265
4	u322	u322	u317	u322
5	u650	u915	u1048	u915
6	u1048	u565	u650	u1048
7	u614	u650	u614	u650
8	u655	u1765	u915	u614
9	u1328	u1048	u565	u1765
10	u1765	u614	u1765	u565

Table 5
Ranking result of key users on topic B.

Rank	PageRank	MPURank	CoRank	PUGRank
1	u3703	u2546	u3703	u3703
2	u1445	u3027	u3027	u3027
3	u3289	u4703	u2546	u2546
4	u2204	u151	u614	u1445
5	u151	u3289	u3289	u3289
6	u2546	u1553	u650	u151
7	u35	u1445	u4525	u1553
8	u1553	u1765	u151	u2204
9	u1720	u2204	u3400	u4525
10	u4525	u237	u19	u3400

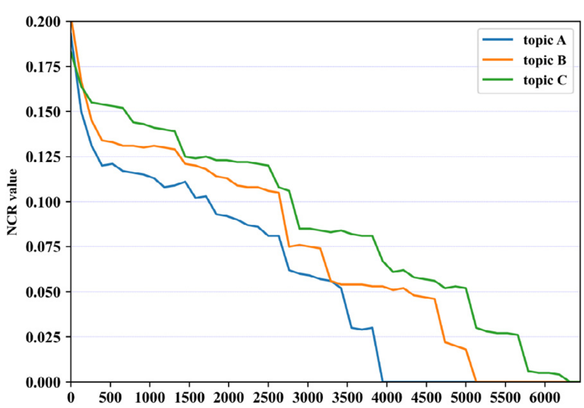


Fig. 8. The NCR value of paths on topic A, B and C.

5.2.2. Key users analysis

First, compare the sequence results obtained by our algorithm and the baseline method, Tables 4, 5, 6 present the Top-10 user ranking results for topic A, B and C respectively.

Then we analyze the performance of the PUGRank algorithm by comparing it with the baseline algorithm, as shown in Fig. 9.

It can be seen that the key user discovery results of our algorithm perform better in most indicators than the baseline algorithm. The PageRank algorithm has the worst performance,

Table 6
Ranking result of key users on topic C.

Rank	PageRank	MPURank	CoRank	PUGRank
1	u989	u151	u151	u151
2	u1546	u1800	u1800	u989
3	u1800	u4729	u989	u1800
4	u2726	u989	u2726	u2334
5	u151	u1546	u4729	u1546
6	u4729	u428	u2334	u4729
7	u2334	u2334	u428	u2726
8	u1703	u108	u2291	u2291
9	u428	u1703	u1546	u428
10	u108	u2291	u108	u1703

and other algorithms can better consider multiple types of elements and relationships; thus, they have better performance. Our PUGRank and CoRank algorithms can better find the influence of key users, because they take into account the user's follow relationship more fully than the MPURank algorithm.

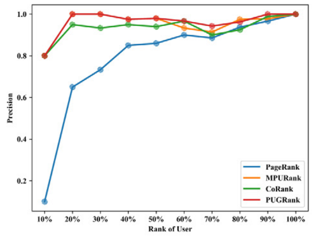
As shown in Fig. 10, among the *Top* − 2% of the key users, the NCR value of a user of the PUGRank and the baseline method are almost the same. It means that for users with greater influence, several methods can effectively identify. However, among the key users after *Top* − 2%, the NCR value of our algorithm is significantly higher than other baseline methods, and finally stabilizes at around 0.2. Since some users in hot topic network participate in multiple paths to spread the topic information, the baseline methods ignore the overlap of user participation; thus, the calculation of NCR values for the most key user is lower than our algorithm. Therefore, we can prove that our algorithm has better performance than the baseline method, and has a good effect on the identification of key users in hot topics.

5.2.3. Key user groups analysis

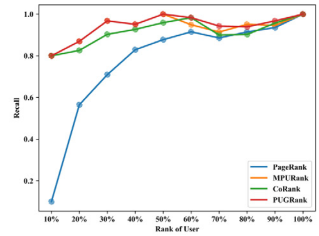
Key user groups refer to certain user groups of high popularity during the topic propagation life cycle. Here, the proposed PUGRank is compared with PsoRank [30] and PageRank algorithms, and the results are shown in Tables 7, 8, 9.

For a better understanding of how to apply the PageRank to the key user groups discovery algorithm, a simple diagram is drawn here, as shown in Fig. 11, user group 1(g1) contains five users. Among them, g1 contains two users following users in user group 2(g2), g1 contains three users following users in user group 3(g3), and g1 contains two users following users in user group 4(g4). The transfer probability values of g1 assigned to g2, g3, and g4 are 2 / (2 + 3 + 2), 3 / (2 + 3 + 2), and 2 / (2 + 3 + 2), respectively. Then we use the PageRank to calculate the influence of user groups.

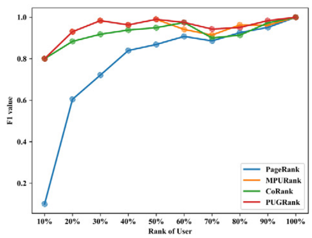
Finally, from Fig. 12, it is not difficult to find that there is little difference between the recognition results of these methods. However, in the key user groups before *Top* − 5%, the user groups' node coverage of our algorithm is significantly higher than that of other methods. Because considering friend relationship between users and considering the cross-user groups characteristics of the



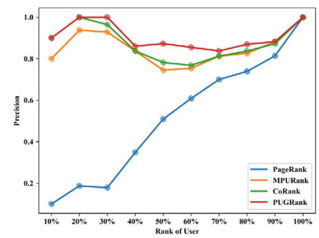
(a) Precision analysis under topic A



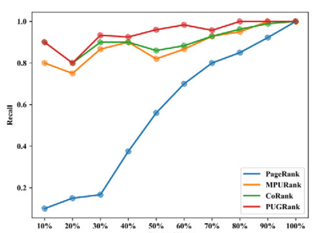
(b) Recall analysis under topic A



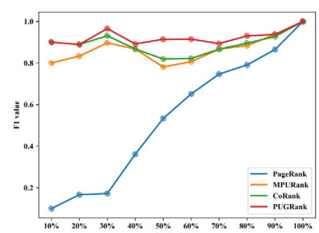
(c) F1 value analysis under topic A



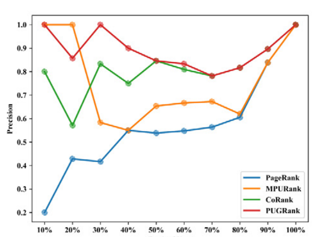
(d) Precision analysis under topic B



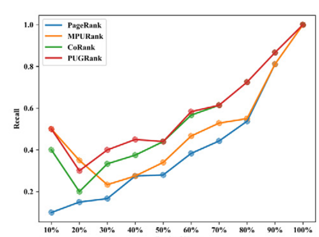
(e) Recall analysis under topic B



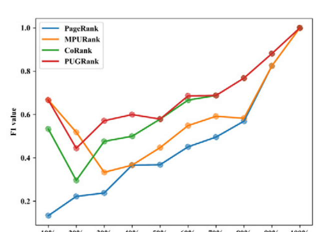
(f) F1 value analysis under topic B



(g) Precision analysis under topic C

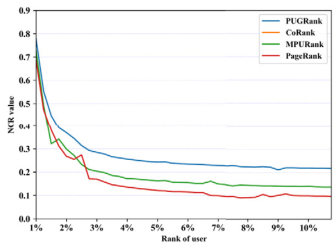


(h) Recall analysis under topic C

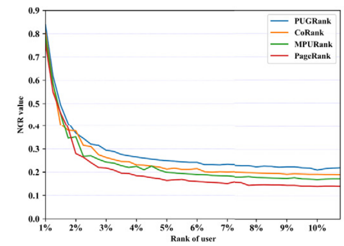


(i) F1 value analysis under topic C

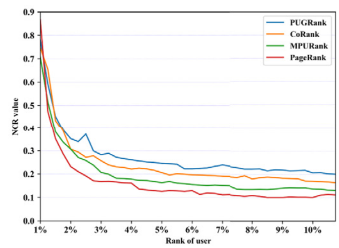
Fig. 9. Key user discovery performance analysis of each algorithm(1).



(a) NCR value analysis under topic A



(b) NCR value analysis under topic B



(c) NCR value analysis under topic C

Fig. 10. Key user discovery performance analysis of each algorithm(2).

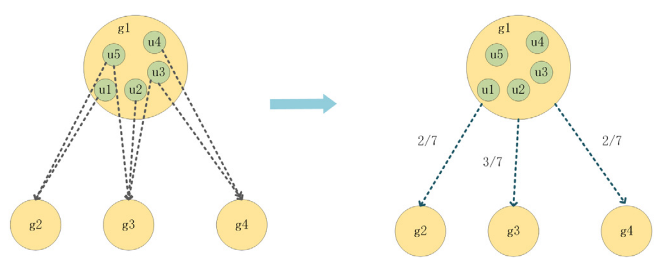


Fig. 11. Application of PageRank algorithm in key user groups discovery algorithm.

Table 7
Ranking result of key user groups on topic A.

Rank	PsoRank	PageRank	PUGRank	Critical score
1	g1	g1	g1	0.387085
2	g35	g35	g35	0.349137
3	g41	g3	g3	0.327770
4	g3	g41	g41	0.308673
5	g52	g88	g52	0.232686
6	g88	g33	g88	0.228077
7	g28	g52	g33	0.218571
8	g33	g28	g28	0.211978
9	g60	g8	g60	0.193128
10	g8	g101	g8	0.192608

Table 8
Ranking result of key user groups on topic B.

Rank	PsoRank	PageRank	PUGRank	Critical score
1	g1	g1	g1	0.385142
2	g14	g2	g2	0.330618
3	g2	g14	g14	0.218179
4	g23	g40	g23	0.164767
5	g26	g23	g40	0.146252
6	g40	g9	g9	0.143243
7	g9	g26	g26	0.133426
8	g10	g10	g10	0.131545
9	g34	g20	g34	0.117894
10	g20	g34	g20	0.100762

Table 9
Ranking result of key user groups on topic C.

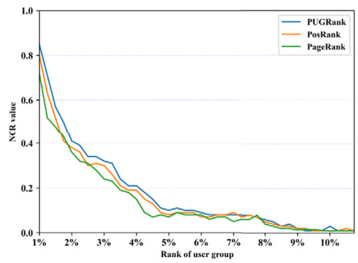
Rank	PsoRank	PageRank	PUGRank	Critical score
1	g5	g5	g5	0.395447
2	g17	g17	g17	0.369574
3	g43	g32	g43	0.217332
4	g1	g43	g32	0.191585
5	g32	g1	g1	0.172056
6	g34	g34	g34	0.112725
7	g25	g10	g25	0.112028
8	g10	g25	g10	0.103109
9	g17	g89	g3	0.092792
10	g3	g3	g17	0.085499

topic propagation process, while the other two methods neglect the characteristics of this user's information dissemination of multi-user groups, so the node coverage rate of the key user groups identified by our algorithm is significantly higher than that of the other two methods, confirming the superiority of the proposed algorithm in effectively identifying the influence of user groups.

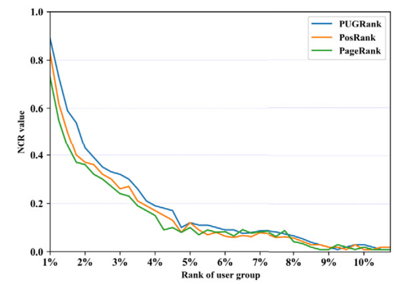
5.2.4. *Propagation trend analysis*

In this paper, topics A, B, and C are divided into ten stages according to the order of topic propagation time, and the number of users and user groups in each stage is counted, as shown in Fig. 13. It is not difficult to find that in topics A, B, and C, with the increase in the propagation stage, the number of users and user groups in each stage increases, the number of users participating in hot topics and the number of communication user groups in turn show three states: first increase and then decrease, fluctuation increase and fluctuation within a certain range. It is consistent with the actual situation of topic communication in social networks and proves the rationality of the experimental dataset selected.

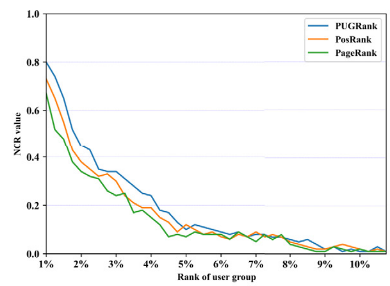
Based on identifying key elements in the topic in the previous section, the participating users and user groups are numbered according to the sequence of occurrence time to realize the statistical analysis of the communication situation of key users and key user groups of the three hot topics in each stage, as shown



(a) NCR value analysis under topic A

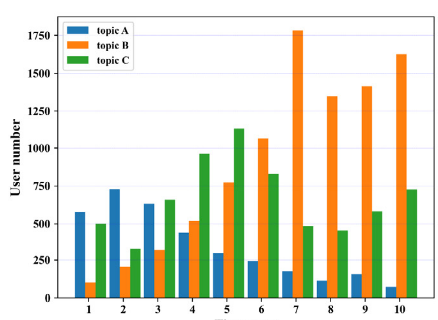


(b) NCR value analysis under topic B

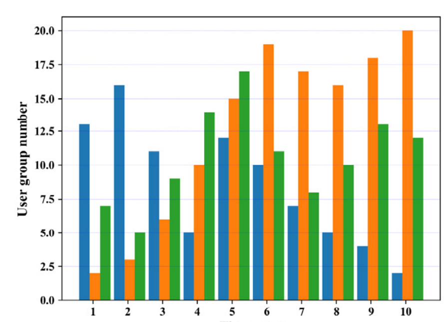


(c) NCR value analysis under topic C

Fig. 12. Key user group discovery performance analysis of each algorithm.

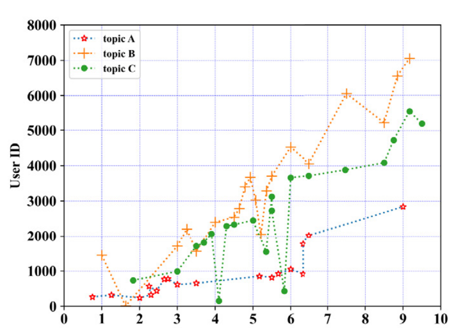


(a) Number of users

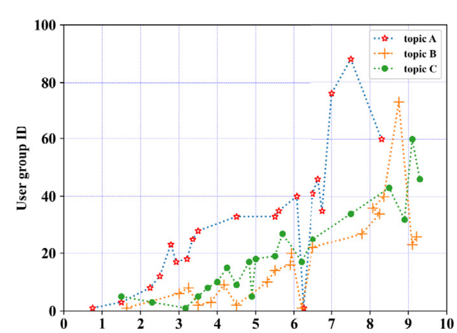


(b) Number of user groups

Fig. 13. The number of different elements on each stage.



(a) Propagation trend of key users



(b) Propagation trend of key user groups

Fig. 14. The propagation trend of key elements on each stage.

in Fig. 14. Through this graph, we can intuitively find the link relationship between the key user and the next key user, the key user group and the next key user group in each stage. We can also intuitively understand the distribution of key users and key user groups in hot topics. For example, in topic A, key users are concentrated in stages 2–3 and 5–6, in topic B, the key user groups are concentrated in stages 3–6 and 8–9. Additionally, it is consistent with the number of users and user groups in Fig. 14, showing that the proposed algorithm has a good effect on the identification of key elements and confirming the effectiveness of the algorithm.

6. Conclusion

In this paper, we proposed a method for discovering the influence of different type of elements in the hot topics. First, we obtained the overlapping user groups through the user relationship network, and the "user–user group" binary association graph model is constructed. Second, based on the "path–user" binary association graph and user group binary association graph, the "path–user–user group" ternary association graph model is constructed to describe the potential relationship. Finally, referring to the cross-scoring strategy, we proposed elements influence discovery algorithm based on ternary association graph and cross-iterative scoring mechanism. In this paper, the real data of Tencent microblog for the experiment. The experimental results showed that this method improves the accuracy and convenience of hot topic key elements influence discovery, and provides some help for online public opinion topic monitoring and hot information delivery.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This paper is partially supported by the National Natural Science Foundation of China (Grant No. 62006032, 62072066) and Science and Technology Research Program of Chongqing Municipal Education Commission, China (Grant No. KJZD-K201900603, KJQN201900629) and Chongqing Technology Innovation and Application Development Project, China (Grant No. cstc2020jscx-msxmX0150).

References

[1] M.A. Al-Garadi, K.D. Varathan, S.D. Ravana, E. Ahmed, G. Mujtaba, M.U.S. Khan, S.U. Khan, Analysis of online social network connections for identification of influential users: Survey and open research issues, ACM Comput. Surv. 51 (1) (2018) 1–37.
[2] G. Fanti, P. Kairouz, S. Oh, K. Ramchandran, P. Viswanath, Hiding the rumor source, IEEE Trans. Inform. Theory 63 (10) (2017) 6679–6713.
[3] J. Becker, D. Brackbill, D. Centola, Network dynamics of social influence in the wisdom of crowds, Proc. Natl. Acad. Sci. 114 (26) (2017) 5070–5076.
[4] J. Geng, Y. Li, Z. Zhang, L. Tao, Sentinel nodes identification for infectious disease surveillance on temporal social networks, in: IEEE/WIC/ACM International Conference on Web Intelligence, 2019, pp. 493–499.
[5] L. Liu, J. Tang, J. Han, M. Jiang, S. Yang, Mining topic-level influence in heterogeneous networks, in: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, 2010, pp. 199–208.
[6] J. Wu, Y. Sha, R. Li, Q. Liang, B. Jiang, J. Tan, B. Wang, Identification of influential users based on topic-behavior influence tree in social networks, in: National CCF Conference on Natural Language Processing and Chinese Computing, Springer, 2017, pp. 477–489.
[7] A. Zareie, A. Sheikhahmadi, M. Jalili, M.S.K. Fasaei, Finding influential nodes in social networks based on neighborhood correlation coefficient, Knowl.-Based Syst. 194 (2020) 105580.
[8] W. Zhao, Z. Guan, Y. Huang, T. Xi, H. Sun, Z. Wang, X. He, Discerning influence patterns with beta-poisson factorization in microblogging environments, IEEE Trans. Knowl. Data Eng. 32 (6) (2019) 1092–1103.
[9] P. Li, W. Zhao, J. Yang, Q.Z. Sheng, J. Wu, Let's CoRank: trust of users and tweets on social networks, World Wide Web 23 (5) (2020) 2877–2901.
[10] S. Giannoulakis, N. Tsapatsoulis, Filtering Instagram Hashtags through crowdtagging and the HITS algorithm, IEEE Trans. Comput. Soc. Syst. 6 (3) (2019) 592–603.
[11] J. Cheng, X. Wu, M. Zhou, S. Gao, Z. Huang, C. Liu, A novel method for detecting new overlapping community in complex evolving networks, IEEE Trans. Syst. Man Cybern. 49 (9) (2018) 1832–1844.
[12] M. Lu, Z. Zhang, Z. Qu, Y. Kang, LPANNI: Overlapping community detection using label propagation in large-scale complex networks, IEEE Trans. Knowl. Data Eng. 31 (9) (2018) 1736–1749.
[13] G. Bello-Orgaz, S. Salcedo-Sanz, D. Camacho, A multi-objective genetic algorithm for overlapping community detection based on edge encoding, Inform. Sci. 462 (2018) 290–314.
[14] Y. Li, K. He, K. Kloster, D. Bindel, J. Hopcroft, Local spectral clustering for overlapping community detection, ACM Trans. Knowl. Discov. Data 12 (2) (2018) 1–27.
[15] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, K. Weinberger, Simplifying graph convolutional networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6861–6871.
[16] A. Grover, J. Leskovec, node2vec: Scalable feature learning for networks, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 855–864.
[17] W.L. Hamilton, R. Ying, J. Leskovec, Inductive representation learning on large graphs, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 1025–1035.
[18] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, 2017, arXiv preprint arXiv:1710.10903.
[19] J. Chen, M. Zhong, J. Li, D. Wang, T. Qian, H. Tu, Effective deep attributed network representation learning with topology adapted smoothing, IEEE Trans. Cybern. PP (99) (2021) 1–12, http://dx.doi.org/10.1109/TCYB.2021.3064092.
[20] G. Xue, M. Zhong, J. Li, J. Chen, C. Zhai, R. Kong, Dynamic network embedding survey, 2021, arXiv preprint arXiv:2103.15447.
[21] A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. Schardl, C. Leiserson, EvolveGCN: Evolving graph convolutional networks for dynamic graphs, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 5363–5370.
[22] L. Zhang, H. Pan, Y. Su, X. Zhang, Y. Niu, A mixed representation-based multiobjective evolutionary algorithm for overlapping community detection, IEEE Trans. Cybern. 47 (9) (2017) 2703–2716.
[23] L. Wu, Y. Ge, Q. Liu, E. Chen, R. Hong, J. Du, M. Wang, Modeling the evolution of users' preferences and social links in social networking services, IEEE Trans. Knowl. Data Eng. 29 (6) (2017) 1240–1253.
[24] Y. Zhang, W. Mao, J. Lin, Modeling topic evolution in social media short texts, in: 2017 IEEE International Conference on Big Knowledge (ICBK), IEEE, 2017, pp. 315–319.
[25] Q. Deng, G. Cai, H. Zhang, Y. Liu, L. Huang, F. Sun, Enhancing situation awareness of public safety events by visualizing topic evolution using social media, in: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, 2018, pp. 1–10.
[26] B. Perozzi, R. Al-Rfou, S. Skiena, Deepwalk: Online learning of social representations, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014, pp. 701–710.
[27] Y. Liu, J. Zhao, Y. Xiao, C-RBFNN: A user retweet behavior prediction method for hotspot topics based on improved RBF neural network, Neurocomputing 275 (2018) 733–746.
[28] Wikipedia contributors, Tencent Weibo — Wikipedia, the free Encyclopedia, 2020, https://en.wikipedia.org/w/index.php?title=Tencent_Weibo&oldid=977079887 (Accessed July 1 2021).
[29] Y. Xiao, H. Yu, Q. Li, L. Liu, M. Xu, H. Xiao, MPURank: A social hotspot tracking scheme based on tripartite graph and multimessages iterative driven, IEEE Trans. Comput. Soc. Syst. 6 (4) (2019) 715–725.
[30] Z. Li, C. Xia, T. Wang, X. Liu, An information source identification algorithm based on shortest arborescence of network, in: Proceedings of the 2019 8th International Conference on Software and Computer Applications, 2019, pp. 442–446.