

Time sequential influence maximization algorithm based on neighbor node influence^①

CHEN Jing (陈晶)^{* ****}, QI Ziyi^{** ***}, LIU Mingxin^{② ****}

(^{*} School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, P. R. China)

(^{**} Hebei Key Laboratory of Virtual Technology and System Integration, Qinhuangdao 066004, P. R. China)

(^{***} Hebei Key Laboratory of Software Engineering, Qinhuangdao 066004, P. R. China)

(^{****} College of Electronic and Information Engineering, Guangdong Ocean University, Zhanjiang 524088, P. R. China)

Abstract

In view of the forwarding microblogging, secondhand smoke, happiness, and many other phenomena in real life, the spread characteristic of the secondary neighbor nodes in this kind of phenomenon and network scheduling is extracted, and sequence influence maximization problem based on the influence of neighbor nodes is proposed in this paper. That is, in the time sequential social network, the propagation characteristics of the second-level neighbor nodes are considered emphatically, and k nodes are found to maximize the information propagation. Firstly, the propagation probability between nodes is calculated by the improved degree estimation algorithm. Secondly, the weighted cascade model (WCM) based on static social network is not suitable for temporal social network. Therefore, an improved weighted cascade model (IWCM) is proposed, and a second-level neighbors time sequential maximizing influence algorithm (STIM) is put forward based on node degree. It combines the consideration of neighbor nodes and the problem of overlap of influence scope between nodes, and makes it chronological. Finally, the experiment verifies that STIM algorithm has stronger practicability, superiority in influence range and running time compared with similar algorithms, and is able to solve the problem of maximizing the timing influence based on the influence of neighbor nodes.

Key words: neighbor node influence, time sequential social network, influence maximization (IM), information propagation model

0 Introduction

With the rapid development of Internet, more and more people like to spread their ideas and information through social media to influence other users on the platform. How to make the shared information spread quickly and influence the widest range has become a hot issue in the field of social network analysis. For this kind of problem, Ref. [1] proposed the influence maximization problem for the first time. At present, the widely used problem of influence maximization based on static social network is to find k users in the social network as seed nodes, so that the information can influence other users in the network as much as possible through k users under a specific propagation model.

At present, most studies usually abstract social

network as static graph, which simplifies the study of influence maximization problem, while neglecting some important characteristics existing in the actual network, thus causing the problem of inaccurate scope of influence. To solve this problem, in view of the timing of network and the propagation characteristics of neighbor nodes, taking the timing social network and second-level neighbor nodes as the research objects, and a timing influence maximization algorithm based on the influence of neighbor nodes and the improved weighted cascade model (IWCM) is proposed in this paper. This algorithm has strong pertinence and practicability, and can efficiently solve the problem of maximizing the time sequential influence based on the influence of neighbor nodes.

Because time sequential effect maximization research based on the influence of the neighbor node is

① Supported by the National Natural Science Foundation of China (No. 62172352, 61871465, 42002138), the Natural Science Foundation of Hebei Province (No. F2019203157) and the Science and Technology Research Project of Hebei (No. ZD2019004).

② To whom correspondence should be addressed. E-mail: mingxinliu001@sina.com.

Received on Apr. 20, 2021

very few, and most studies focused on only considering the temporal characteristics of the network or only considering the influence of the neighbor nodes, without considering a variety of characteristics of network, so the timing effect maximization problem based on the influence of the neighbor node challenges are as follows. (1) The traditional information dissemination model does not consider the timing characteristics of the network, it cannot be applied to the timing social network; (2) In the process of seed node selection, it is necessary to comprehensively consider the propagation characteristics of the second-level neighbor nodes and the timing of the network.

In order to solve the above problems, taking the time sequential social network as the research object, timing the traditional information transmission model, the time sequential influence maximization algorithm is designed based on the influence of neighbor nodes. Firstly, the influence measurement method of second-level neighbor nodes and the selection method of seed nodes are given. Secondly, the second-level neighbor node influence measurement method is time-serialized so that it can be applied to time-series social networks. Second-level neighbors time sequential maximizing influence algorithm (STIM) takes full account of various characteristics of the network, and is able to efficiently solve the problem of maximizing the time sequential influence based on the influence of neighbor nodes, and provides a foundation for the modeling of related problems, the selection of seed nodes and how to reduce the time complexity.

The major contributions of this paper are as follows.

(1) On the basis of the traditional weighted cascade propagation model, a new method of transmission probability between computing nodes is introduced, and an improved weighted cascade model is proposed. It enables information to be transmitted in the social network graph based on time sequential relationship.

(2) In view of the propagation characteristics of secondary neighbor nodes in the network, the influence measurement method of secondary neighbor nodes is designed, and the timing characteristics of the network are integrated into the selection process of seed nodes.

(3) It is verified that STIM algorithm can efficiently solve the problem of time sequential influence maximization based on the influence of neighbor nodes, and guarantee a high influence range under the premise of short running time.

The structure of this paper is as follows. In Section 1, the related work is introduced. In Section 2, the main definition and propagation model of the time

sequential influence maximization problem based on the influence of neighbor nodes are discussed. In Section 3, the correlation algorithm of the time sequential influence maximization problem based on the influence of neighbor nodes is described. In Section 4, the experiment results on different data sets are analyzed and compared. In Section 5, the work of this paper is summarized and prospected.

1 Related work

In recent years, researchers at home and abroad have done a lot of research work on impact maximization. Kempe et al.^[2] proposed the greedy algorithm for the problem of influence maximization (IM) and proved that its operation result could reach an approximate optimal of 63%. However, the greedy algorithm still has a high time complexity and is not suitable for large-scale social networks. Leskovec et al.^[3] optimized the traditional greedy algorithm for the submodularity and monotonicity of the influence maximization problem, and proposed a cost-effective lazy-forward (CELFG) algorithm, which is about hundreds of times faster than the greedy algorithm. Goyal et al.^[4] proposed CELFG++ by optimizing CELFG algorithm, and proved that it was 35% – 55% faster than CELFG.

All the above algorithms are greedy algorithms or improved greedy algorithms. In recent years, many researchers have studied heuristic algorithms with lower time complexity. Chen et al.^[5] proposed the DegreeDiscount algorithm to solve the overlap problem of influence range of traditional degree estimation algorithms. The node with the highest degree is selected as the seed node, and then the degree of the neighbors of the selected node is discounted until k nodes are selected. Zhou and Gao^[6] used PageRank algorithm to estimate the influence of nodes, and selected nodes with greater influence as alternative nodes to calculate the combined collision probability of alternative nodes. Li et al.^[7] proposed an influence maximization algorithm based on k -kernel filtering in combination with the heuristic algorithm. It has been verified that the algorithm has a wider scope of influence compared with the existing heuristic algorithm.

At present, more and more researchers begin to study the extension deformation with maximum effect. Qiu et al.^[8] proposed the influence maximization algorithm of overlapping communities, which could increase the running time by up to 90% and could be applied to large social networks. Ren et al.^[9] proposed the problem of influence maximization under multi-topic consciousness. Cross-social network model of influ-

ence maximization for topic perception is designed by improving the linear threshold model, and seed nodes are selected by means of heuristic algorithm. Li et al.^[10] took price and other factors into consideration, and studied how to price a product under multiple conditions to maximize income. Zhao et al.^[11] took into account the situation that there are multiple relationships among users in social networks and that multiple relationships affect information transmission together, MR-RRSET algorithm was proposed to solve the problem of influence maximization of multi-relationship social networks.

Kim et al.^[12] transferred the research object of maximum impact to the dynamic graph and designed an algorithm to deal with the update operation on the dynamic graph. Wang et al.^[13] defined novel influence maximization (IM) queries in social networks and used a window sliding model to solve the problem of real-time impact maximization on dynamic graphs. Zhang et al.^[14] studied the problem of influence maximization when there are different relationships in the network that mutually promote communication and inhibit communication. Guo and Lu^[15] improved the sketch-based design influence maximization algorithm and applied it to the influence maximization problem of dynamic graphs, then recalculated the seed node set by calculating the influence of the deletion or addition of nodes on the current sampling set. Cao et al.^[16] set a time window and regarded the connection between nodes as an action. The window slides down with time, and then the new action enters the window while the old action exits. According to the entry and exit of nodes, it is judged whether it is necessary to recalculate the seed nodes in the window obtained in the previous period to solve the problem of maximizing the influence in the dynamic graph. Wu et al.^[17] studied the problem of influence maximization with the object of time sequence diagram, improved the traditional independent cascade model, and proposed two algorithms AIMT and IMIT to solve the problem of influence maximization of time sequence diagram. Wei et al.^[18] studied the influence maximization problem of dynamic social networks with time change as the main feature, and proposed the influence maximization algorithm of dynamic social networks based on the feature. Yang et al.^[19] proposed a method to measure the influence of complex network nodes with three levels of neighbors according to the propagation characteristics of neighbor nodes. This method regarded the neighbors of node level 2 and 3 with propagation attenuation characteristics as a whole to measure the influence ability of nodes.

2 Problem definition

2.1 Basic definition

Definition 1 (time sequential social network) Given network $G_T(V, E, T_E)$ based on the time sequential relationship between social network diagram, V represents node set, E represents the edge of the collection, including $|V| = n$, $|E| = m$, T_E represents the set of moments with connection between nodes in the network, and $T_{(u,v)}$ represents the set of moments with connection between nodes u and v .

Take Fig. 1 as an example to illustrate the social network graph based on time sequential relationship. Compared with static social network graph G (edge weights for the node transmission probability), time sequential social network graph G_T are given the concept of timeline, each node only at specific time points, the weight on the edge represents the moment when there is a connection between the two nodes. For example, in Fig. 1(b), node a and node b are only connected at two moments 3 and 6, there is no connection at other moments. Now calculate the influence range of node a in Fig. 1(a) and Fig. 1(b) respectively. The calculation process is as follows.

For convenience of calculation, it can be supposed that the propagation probability between nodes in Fig. 1(a) and Fig. 1(b) is the same, and is the weight values of each edge in Fig. 1(a). Therefore, in Fig. 1(a), node a activates node b and node c with probabilities of 0.1 and 0.03, respectively. If node c is activated at this time, then node c activates node e with a probability of 0.1. If node e is activated at this time, then node a successfully activates node c and e , with an influence range of 2. In Fig. 1(b), node a activates node b and node c at moments 3 and 2 with probabilities of 0.1 and 0.03, respectively. If node c is activated, then node c will be activated after time 2. Since node c and node e are only connected at time 1, and node c is not activated at time 1, node c will no longer be able to activate node e . Therefore, node a will only successfully activate node c and its influence range is 1.

Therefore, it can be seen that if only the static graph-based influence maximization algorithm is applied to the time sequential social network graph, the correct result cannot be obtained. Therefore, it is necessary to study the time sequential relations based social network influence maximization problem.

2.2 Calculation of propagation probability

Definition 2 (propagation probability) The probability that the active node u successfully activates its

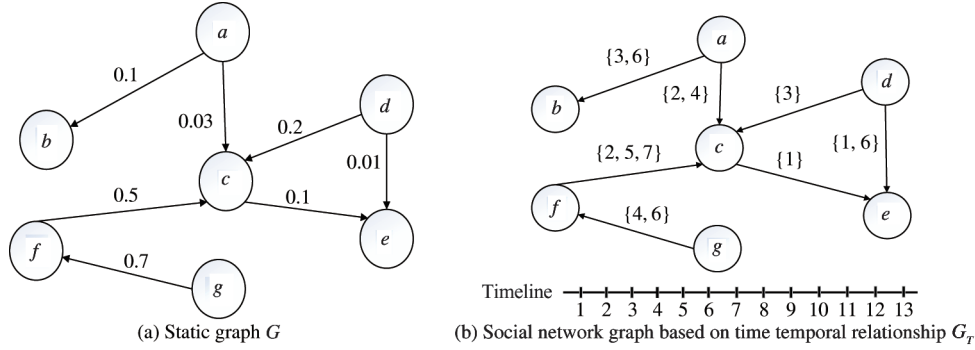


Fig. 1 Static graph and time sequential social network graph

neighbor node v through the edge (u, v) is the propagation probability, expressed as $P_{u,v} \in [0, 1]$.

In the traditional research of influence maximization algorithm, the degree estimation method is usually used to calculate the probability of inter-node propagation, that is, the reciprocal of in-degree of the node is used to estimate the probability of the node being activated by the upper node, as shown in Eq. (1).

$$P_{u,v} = 1/\text{InDegree}(v) \quad (1)$$

In Eq. (1), $\text{InDegree}(v)$ represents the in-degree of node v .

This method has been well proved and applied in the traditional research of influence maximization algorithm. However, the method does not take into account the problem of different contact times between nodes in the social network graph based on time sequential relationship. Now, an example is given to illustrate this problem, as shown in Fig. 2. In the static graph G , the in-degree of node c is 2, the probability that node c is affected by both nodes a and d is $1/2$. But in figure G_T , when considering the number of connections, it can be found that the contact number of nodes c and a is less than the contact number of nodes c and d , and the more times a node is contacted, the greater the probability that it will be affected. Therefore, in Fig. 2(b), the probability of node c being affected by node a should be less than the probability of node c being affected by node d . The calculation result is inconsistent with that in Fig. 2(a), that is, it is inaccurate to use traditional degree estimation methods to calculate the probability of node influence in a time sequential social network graph. Therefore, the traditional calculation method needs to be improved.

In the traditional degree estimation algorithm, the probability of node v being activated by node u is: the proportion of edges (u, v) in all edges of node v , namely $P_{u,v} = 1/\text{InDegree}(v)$. But in the degree estimation algorithm based on time sequential social networks, because of the addition of time sequential relations, the weight of each edge is no longer the same,

and it is related to the number of connections between two nodes $|T(u, v)|$, that is, the more connected the edge, the greater the weight, and is proportional to the relationship between them. Therefore, the weight of the edge (u, v) is set to $|T(u, v)|$, then the proportion of edges (u, v) in all income of node v is $\frac{|T(u, v)|}{\sum_{v_k \in \text{In}(v)} |T(v_k, v)|}$, and the activation probability of node v by node u is shown in Eq. (2).

$$P_{u,v} = \frac{|T(u, v)|}{\sum_{v_k \in \text{In}(v)} |T(v_k, v)|} \quad (2)$$

In the Eq. (2), $|T(u, v)|$ represents the number of connections between node u and node v , and v_k represents all in-degree nodes of node v .

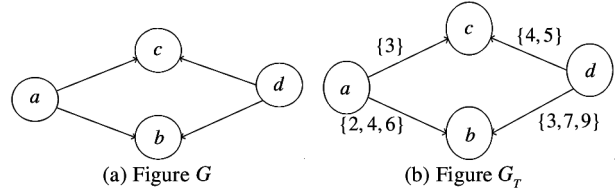


Fig. 2 Schematic diagram of propagation probability calculation

2.3 Improvement of weighted cascading model

2.3.1 Traditional weighted cascading model

In the traditional weighted cascade model (WCM), each directed edge (u, v) is given a real value $P_{u,v} = 1/\text{InDegree}(v)$ and $P_{u,v} \in [0, 1]$, where $P_{u,v}$ represents the probability that node u successfully influences node v through the directed edge (u, v) . The propagation process of WCM model is as follows: at the initial time t , the active node u has only one chance to activate each of its inactive neighbor nodes v , and the activation probability is $P_{u,v}$. If node v has multiple active parents at time t , then the active parent node activates node v in any order at time t . If v is successfully activated, it becomes active node at time $t + 1$ and tries to activate its next level of inactive neighbor node in the same way. This cycle continues until no new nodes in

the network are activated.

2.3.2 Improve the weighted cascade model

Definition 3 (node active initial time) The time when node v is successfully activated by its active parent node u is its active initial time, which is expressed as Act_v and $Act_v = \min \{ t \mid (t \in T(u, v) \text{ and } t \geq Act_u) \}$.

Take Fig. 2(b) as an example, the set node d is the seed node (the initial active time of the seed node is 0), and if it successfully activates node c , $Act_c = \min \{ 4, 5 \} = 4$.

The traditional influence maximization algorithm does not consider the initial time when nodes are activated, while the initial time when nodes are successfully activated in the time sequential social network graph needs to be considered. Therefore, the traditional weighted cascade model is improved, and a new propagation model is obtained based on time sequential social network graph—improved weighted cascade model in this paper.

Take Fig. 1(b) as an example, let node d be the seed node and successfully activate its neighbor node c , then the initial active time of node c , $Act_c = 2$, that is, node c is active after time 2. And because node c and node e are only connected at time 1, at this time c is still in an inactive state, so node c must not activate node e .

The IWCM propagation model is proposed based on the WCM propagation model, so that information can be transmitted in the social network graph based on the time sequential relationship. The propagation process of information in the time sequential social network diagram through the IWCM model is described as follows.

(1) In the initial network, the initial active time of all nodes is set to $Act_v = -1$, indicating that all nodes are in an inactive state. Set the initial active time of seed node u , $Act_u = 0$, indicating that the seed node is in active state at the time of 0. At this point, seed node u activates its neighbor node v with a certain probability, and node u has only one chance to try to activate node v .

(2) When node u tries to activate the node v , it first determines whether it meet $Act_u \leq \max(T_{(u,v)})$. If $Act_u > \max(T_{(u,v)})$, skip the node and start trying to activate the next neighbor node; if meeting the $Act_u \leq \max(T_{(u,v)})$, then node u activates node v with probability $p_{u,v} = \frac{|T(u, v)|}{\sum_{v_k \in \ln(v)} |T(v_k, v)|}$.

(3) No matter whether the node can activate node v or not, u will not try to activate node v in the future propagation process.

(4) If node v is successfully activated, its initial active time Act_v is recorded, where $Act_v \in T(u, v)$, $Act_u \leq Act_v \leq \max(T_{(u,v)})$.

(5) In the whole network, information is transmitted from new active nodes to inactive neighbor nodes until no new nodes in the network are activated.

2.4 Problem definition

This section defines and explains the problem of time sequential influence maximizing based on the description of the above discussions. And the concept of node influence and marginal benefit in time sequential social network is introduced.

Definition 4 (node influence) Node influence refers to the set of all nodes that can be successfully activated by node v in the network, expressed as $\sigma(v)$.

Definition 5 (marginal benefit) The marginal benefit of node v refers to the increase in revenue that can be brought by adding a node v to the seed set S . The calculation formula is shown in Eq. (3).

$$\sigma_v(S) = \sigma(S \cup \{v\}) - \sigma(S) \quad (3)$$

where, $\sigma(S)$ represents the influence range of seed set S .

Problem definition (time sequential influence maximization problem based on neighbor node influence) Given time sequential social network graph $G_T = (V, E, T_E)$, and specific propagation model, in the time sequential social network to find a node set S , the number of nodes in the set $S \mid S \mid = k$, make the effect of the set S most widely, set S is the seed node set of G_T .

3 Time sequential influence maximization algorithm based on neighbor node influence

3.1 Influence measurement method of second-level neighbor nodes

The influence of nodes on second-order neighbor nodes can be explained by independent probability events in probability statistics. Independent probabilistic events are defined as: if event A and event B are independent, then $P(AB) = P(A) \times P(B)$, that is, the probability of event A and event B occurring at the same time is the probability of event A multiplied by the probability of event B occurring.

Similarly, in the time sequential social network as shown in Fig. 3, the probability that node 1 activates node 4 through node 2 is the probability that node 1 activates node 2 multiplied the probability that node 2 activates node 4, that is, $P_{(\text{node 1 activates node 4 through node 2})} = 0.2 \times 0.3 = 0.06$.

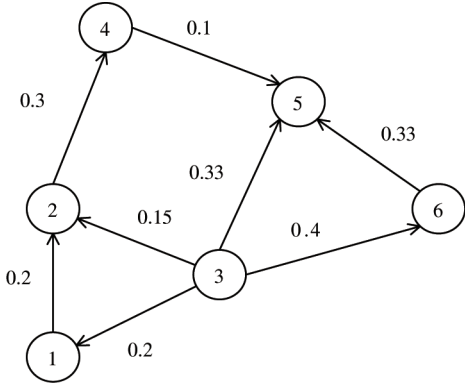


Fig. 3 Social network diagram (second-level neighbor)

Given the time sequential social network $G_T(V, E, T_E)$, V and E represent network node set and edge set respectively, and T_E represents the set at the time when there is a connection between nodes. According to the improved weighted cascading model, the probability that node u activates its neighbor node v is $P_{u,v}$, and the probability that node v activates its next-level node v_1 is P_{v,v_1} , then the probability that node u activates its neighbor node and second-level neighbor node along the path (u, v, v_1) is shown in Eq. (4).

$$P = P_{u,v} + P_{u,v} \times P_{v,v_1} \quad (4)$$

wherein, P represents the sum of the probability that node u activates its neighbor nodes and second-level neighbor nodes on the path (u, v, v_1) .

Therefore, node u is the sum of activation probabilities of all its neighbor nodes and second-level neighbor nodes in the whole network, as shown in Eq. (5).

$$P_u = \sum_{v \in O(u)} (P_{u,v} + \sum_{v_1} P_{u,v} P_{v,v_1}) \quad (5)$$

wherein, $O(u)$ is the outgoing node of node u , and P_u represents the influence range of node u in the whole temporal social network.

Since the influence range of nodes is directly affected by the activation probability between nodes, Eq. (5) can be used to estimate the influence range of nodes. However, when information is transmitted in the actual social network as shown in Fig. 3, node 3 is set as the seed node, and the information transmission process is as follows.

Time step 0 Activate node 3.

Time step 1 Node 3 tries to activate node 2 and node 1 with probabilities of 0.15 and 0.2 respectively, assuming that both node 2 and node 1 are successfully activated.

Time step 2 Since node 2 is already active at this time, node 1 will not try to activate node 2 again. However, when calculating the influence P_1 of node 1, the influence probability of node 1 on node 2 has been included, which is inconsistent with the actual trans-

mission process.

Considering the influence range overlap between the nodes, when node i is selected as seed nodes, all parent nodes u and ancestor nodes v (the relationship between node i , node u , and node v are shown in Fig. 4) are treated with influence discount. In its estimate, subtract the overlapping part of influence, as shown in Eqs(6) and (7).

$$P_u' = P_u - P_{u,i} - \sum_{i_1 \in O(i)} P_{u,i} P_{i,i_1} \quad (6)$$

wherein, node u is the parent node of node i , and node i_1 is the next-level node of node i .

$$P_v' = P_v - P_{u,i} \quad (7)$$

wherein, node v is the ancestor node of node i , and node u is the parent node of node i .

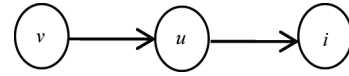


Fig. 4 Schematic diagram of node relationship

3.2 Timing of influence

Due to the addition of time sequential relationship in time sequential social network, Eqs (6) and (7) need to be time-serialized. As shown in Fig. 5, assuming that the initial active time of node 4 is 2, node 4 is already active when node 4 contacts node 2 and node 5. Therefore, node 2 and node 5 have a chance to be activated by node 4 respectively. However, if the initial active time of node 4 is assumed to be 5, node 4 is inactive in the two moments (2 and 3) when node 2 and node 4 are connected, so node 4 has no opportunity to activate node 2. But, node 4 is already active at the time 6 when node 4 is in contact with node 5. Therefore, node 5 has a chance to be activated by node 4 once. It can be seen that, since the initial active time of all nodes in the network is in an unknown state, the larger the maximum value of the contact time between two nodes, the greater the chance that the node has to be tried to activate.

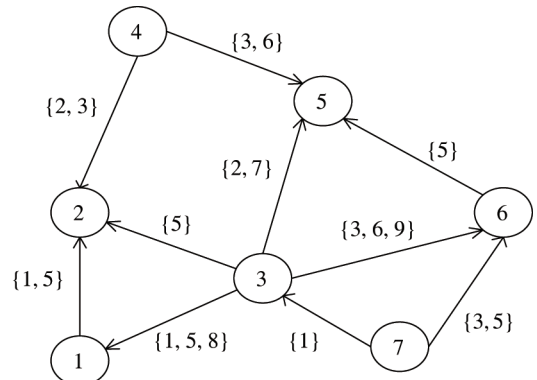


Fig. 5 Time sequential social network (second-level neighbor)

From the above examples, it can be seen that the maximum value of the contact time between two nodes $\max(T(u, v))$ will have an impact on the influence range of nodes, and the larger the maximum value of the contact time between two nodes, the wider the influence range of nodes. Therefore, consideration of $\max(T(u, v))$ is added into the estimation formula of the influence range of nodes, and the time-serialized formula is

$$P_u = \sum_{v \in O(u)} (P_{u,v} \times \max(T(u, v)) + \sum_{v_1 \in O(v)} P_{u,v} P_{v,v_1} \times \min(\max(T(u, v), \max(T(v, v_1)))) \quad (8)$$

After node i is selected into the seed set, the influence of its parent node u is updated, and the influence range of the updated node is shown in Eq. (9).

$$P_u' = P_u - P_{u,i} \times \max(T(u, i) - \sum_{i_1 \in O(i)} P_{u,i} P_{i,i_1} \times \min(\max(T(u, i), \max(T(i, i_1)))) \quad (9)$$

where, P_u' is the estimated value of influence of node u after discount.

At the same time, the influence of its ancestor node v is updated, and the influence range of the updated node is shown in Eq. (10).

$$P_v' = P_v - P_{u,i} \times \max(T(u, i)) \quad (10)$$

where, P_v' is the estimated value of influence of node v after discount.

The above formula is for estimating the influence range of nodes in the time sequential social network. The larger the estimated value of the node's influence, the higher the influence range of the nodes. Therefore, in the heuristic algorithm, the estimated influence of nodes is often used to represent the influence of nodes, and then seed nodes are selected, as shown in Eq. (11).

$$\text{Inf}(v) = P_v' \quad (11)$$

where, $\text{Inf}(v)$ represents the influence range of node v .

3.3 Description of STIM algorithm

The main research strategy of STIM algorithm proposed is heuristic strategy, that is, the influence of non-seed nodes is estimated in each seed node selection process, and the node u with the largest estimated value is selected as the seed node.

The idea of STIM algorithm is as follows: the problem of time sequential social network influence maximization is solved in two steps. Firstly, the node influence of the second-level neighbor is estimated according to the measurement method in subsection 3.1. Then, according to the influence estimation results, the node u with the largest estimated value is selected

as the next seed node, until all k seed nodes are found out. The specific description process of STIM algorithm is Algorithm 1.

Algorithm 1 STIM

Input: social network $G_T(V, E, T_E)$, k ;

Output: seed set S ;

- (1) Calculate propagation probability between nodes in time sequential social network according to the calculation method of propagation probability between nodes in subsection 3.2;
 - (2) Use Eq. (5) to calculate the influence of all nodes in the network, and make it time sequenced;
 - (3) Select the node v with the greatest influence as the first seed node, mark node v as active state and add it to S ;
 - (4) Update the influence estimate of parent node of S and its grandfather node according to Eqs(9) and (10);
 - (5) Select the node u with the largest influence estimation after updating as the second seed node and add it to S ;
 - (6) Repeat Steps (3) to (5) until k nodes are selected.
-

3.4 Pseudo code

$G_T(V, E, T_E)$ is used to represent a social network based on time sequential relationship, where V represents the set of nodes, E represents the set of edges, T_E represents the set of contact moments between nodes, k represents the number of seed nodes required, and S represents the set of seed nodes. The pseudo code of the STIM algorithm is Algorithm 2.

Algorithm 2 Time sequential influence maximization algorithm based on neighbor node influence

Input: social network $G_T(V, E, T_E)$, k ;

Output: seed set S ;

- (1) Initialize $S = \emptyset$;
 - (2) For any node u in graph G_T do;
 - (3) Calculate $P_{u,v}$; // v indicates the next-level neighbor node of node u ;
 - (4) End for;
 - (5) For any node u in graph G_T do;
 - (6) $P_u' = P_u - P_{u,i} \times \max(T(u, i) - \sum_{i_1 \in O(i)} P_{u,i} P_{i,i_1} \times \min(\max(T(u, i), \max(T(i, i_1))))$;
 - (7) End for;
 - (8) For $i = 1$ to k do;
 - (9) $v = \text{argmax}_u \{P_u' \mid u \in V/S\}$;
 - (10) $S = S \cup \{v\}$;
 - (11) End for.
-

In Algorithm 2, Step (1) initializes the seed set S to an empty set; In Steps (2) – (4), the propagation probability between nodes in time sequential social net-

work is obtained based on in subsection 2. 2; Steps (5) – (7) estimate the influence range of all nodes; Steps (8) – (11) find the first k nodes with a larger estimated influence range and incorporate them into the seed set.

The time complexity analysis of STIM algorithm is as follows; let the number of nodes of network $G_T(V, E, T_E)$ be n , the number of edges be m , the size of seed set be k , and the number of contact moments between nodes in the network be t . In Algorithm 1, Steps (2) – (4), the time complexity generated when calculating the propagation probability between nodes is $O(m)$; Steps (5) – (7) estimate the influence scope of each node in the time sequential social network, and the time complexity generated is $O(n)$; Steps (8) – (11) select k nodes with the greatest influence and iterate k rounds, so the time complexity is $O(k)$; To sum up, the time complexity of STIM algorithm is $O(n + m + k)$. Since the value of k is generally less than or equal to 50, that is, far less than n and m , so $O(k)$ can be ignored, then the time complexity of the STIM algorithm is $O(n + m)$.

4 Experiment and result analysis

Three real datasets of different scales are selected as input data to realize seed node selection and seed influence calculation in time sequential social network graph based on neighbor node influence.

4.1 Experimental data and parameter setting

Dataset 1 used in the experiment is derived from the online social network of University of California, composed of private messages. Edges (u, v, t) indicate that user u sends a private message to user v at time t ^[20]. Dataset 2 is the E-mail data of a large European research institution, and the directed edges (u, v, t) indicate that user u communicates with user v through E-mail at time t ^[21]. Dataset 3 is a time sequential network that edits each other's 'conversation' pages on behalf of Wikipedia users. Edges (u, v, t) indicate that user u has edited the conversation page of user v at time t ^[22]. The number of experimental datasets is shown in Table 1.

Table 1 Adoption number of experimental datasets

Dataset	Nodes of dataset	Number of sequential edges	Number of static edges	Time span
Wiki-Talk	1 140 149	7 833 140	3 309 592	2320
CollegeMsg	1899	59 835	20 296	193
Email-Eu-Core	986	332 334	24 929	803

In this paper, the STIM algorithm is divided into two steps. The first step is to calculate the propagation probability between all nodes. The second step is to estimate the influence range of all nodes and select k nodes with the largest influence range as the seed node set. While realizing the algorithm, some classical algorithms and new algorithms in recent years with better running effect are reproduced. The advantages and disadvantages of each algorithm are compared and analyzed from two aspects of the influence range and running time of the algorithm.

Coverage threshold maximum influence (CTMD) is a degree maximum heuristic algorithm based on coverage threshold. This algorithm uses the improved k-shell algorithm to calculate the influence of nodes in the network and select the initial seed node set. Meanwhile, the activation probability of nodes within two degrees is considered.

Influence estimation influence ranking (IEIR) is an algorithm based on influence estimation and influence ranking, and it has good comprehensive ability among traditional influence maximization algorithms at present.

DegreeDiscount is the representative of the heuristic algorithm. The node with the highest degree is selected as the seed node, and then the degree of the neighbors of the selected node is discounted until k nodes are selected.

Random, as a benchmark comparison method, simply randomly selects k non-repeating nodes from the time sequential social network graph as seed nodes.

In the stage of seed node selection, CTMD algorithm, IEIR algorithm, DegreeDiscount algorithm, Random algorithm and STIM algorithm select the size of seed node set k as 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50 respectively.

4.2 Influence of seed nodes in different algorithms

The correlation algorithm is tested on three different datasets. Scope of influence refers to the number of nodes ultimately affected by calculating the seed set through the algorithm in the initial stage of the network to spread the seed set in the network. The wider the influence range of the seed set, the higher the accuracy of the algorithm.

Fig. 6 shows the seed nodes impact on the Wiki-

Talk dataset. As can be seen from Fig. 6, among many comparison algorithms, IEIR algorithm and STIM algorithm have a wide influence range, while the influence range of other algorithms is at a low level. When selecting different number of seed nodes, with the increase of the number of seed nodes, the influence range of IEIR algorithm is always in a stable state. As the number of selected seed nodes increases, the influence range of STIM algorithm expands. That is, when the number of seed nodes is less than 25, the influence range of IEIR algorithm is higher than that of STIM algorithm; but when the number of seed nodes is greater than 25, the influence range of STIM algorithm is higher than that of IEIR algorithm.

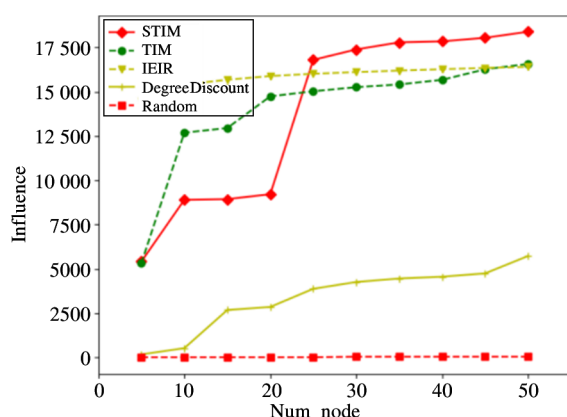


Fig. 6 Effect of seed nodes on Wiki-Talk dataset

Fig. 7 is an effect diagram of seed nodes on CollegeMsg dataset. It can be seen from Fig. 7 that the influence range of STIM algorithm is much higher than that of other algorithms regardless of the size of the seed node set. At $k = 50$, the influence range of STIM algorithm is 7.82 times, 4.56 times, 2.25 times, and 6.92 times that of IEIR algorithm, CTMD algorithm,

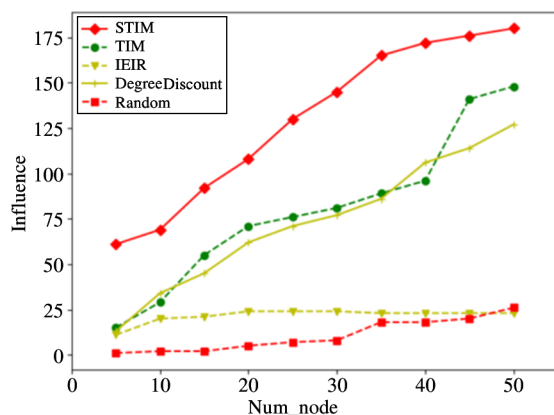


Fig. 7 Effect of seed nodes on CollegeMsg dataset

DegreeDiscount algorithm, and Random algorithm respectively. Among other algorithms, DegreeDiscount

algorithm has better effect, while IEIR algorithm, CTMD algorithm, and Random algorithm have worse effect.

Fig. 8 is an effect diagram of seed nodes on Email-Eu-Core dataset. As can be seen from Fig. 8, STIM algorithm has the widest influence range, followed by IEIR algorithm and DegreeDiscount algorithm, and CTIM algorithm and Random algorithm have the smallest influence range. When $k < 15$, the influence range of IEIR algorithm is higher than DegreeDiscount algorithm, while when $k > 15$, IEIR algorithm is overtaken by DegreeDiscount algorithm. When $k = 15$, the influence ranges of CTMD algorithm and Random algorithm reach the same level.

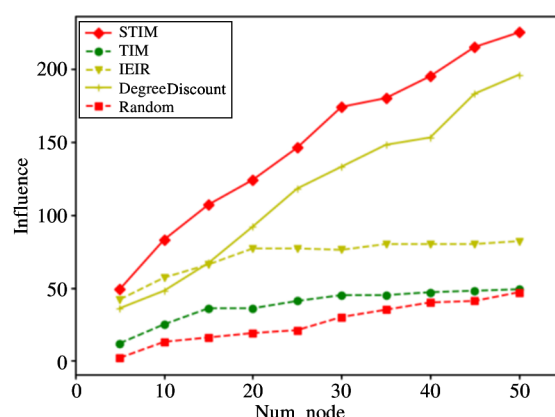


Fig. 8 Effect of seed nodes on Email-Eu-Core dataset

4.3 Selection time of seed nodes in different algorithms

In this subsection, under the IWCM propagation model, the running time of CTMD algorithm, IEIR algorithm, DegreeDiscount algorithm, Random algorithm, and STIM algorithm are respectively counted. The calculated time is the running time of 50 seed nodes selected from three datasets of different scales, as shown in Table 2.

Table 2 Running time of STIM algorithm

Algorithm	Wiki-Talk/s	CollegeMsg/s	Email-Eu-Core/s
CTMD	42.290	1.070	0.310
Degree-Discount	4.800	0.078	0.048
Random	2.900	0.041	0.026
IEIR	209.980	3.560	1.630
STIM	112.450	6.270	0.706

As can be seen from Table 2, the running time of STIM algorithm, CTMD algorithm, DegreeDiscount algorithm, IEIR algorithm, and Random algorithm grows gradually with the gradual increase of the scale of

time sequential social network. IEIR algorithm has the longest running time, followed by STIM algorithm, and other algorithms have relatively short running time.

The analysis of the experimental results shows that the Random algorithm has the shortest running time because of the randomness of seed node selection, but its influence range is much smaller than other algorithms. IEIR algorithm is a traditional influence maximization algorithm with better comprehensive ability. Although its influence range is relatively large, its running time is also the longest. CTMD algorithm is a relatively new influence maximization algorithm in the past two years. Although its running time is short, it does not consider the factors of time sequential and second-level neighbor nodes, which leads to the shrinkage of the influence range. DegreeDiscount algorithm is a heuristic algorithm that only considers first-level neighbor nodes. Compared with STIM algorithm, the running time of DegreeDiscount algorithm is smaller than that of STIM algorithm because the scope of neighbors it considers is smaller, but its influence scope is also much smaller than that of STIM algorithm.

5 Conclusion

In recent years, social phenomena such as secondhand smoke, happiness transmission and microblog forwarding have two main characteristics: (1) Time sequential change. As time goes on, the topology of such networks also changes constantly; (2) Secondary dissemination of information. Information is spread widely in the category of secondary neighbors in such networks. In view of the two characteristics of the network, the STIM algorithm is proposed in this paper. Firstly, the propagation probability between nodes is calculated by the improved degree estimation algorithm; Secondly, the traditional weighted cascade model is improved so that it can be applied to social networks based on time sequential relationships; Finally, STIM algorithm is proposed based on the improved weighted cascade model. Experimental results show that STIM algorithm is more targeted than CTMD, DegreeDiscount, IEIR, Random and other influential maximization algorithms with universal characteristics. In the three networks selected in this section with time sequential characteristics and next-level transmission characteristics of information, STIM algorithm shows its advantages in terms of influence scope and running time, and can better solve the problem of maximizing the influence of next-level neighbor's time sequential social network.

The following in-depth studies will be carried out

in the future work; (1) Research the actual factors of the time sequential influence maximization problem based on the influence of neighbor nodes, such as the different information types, cost, time; (2) Expand the algorithm so that it can be applied to time sequential social networks targeting multi-level neighbor nodes.

Reference

- [1] RICHARDSON M, DOMINGOS P. Mining knowledge-sharing sites for viral marketing[C] // Proceedings of the International Conference on Knowledge Discovery and Data Mining, Alberta, Canada, 2002: 61-67
- [2] KEMPE D, KLEINBERG J, ARDOS É. Maximizing the spread of influence through a social network[C] // Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, USA, 2003: 137-146
- [3] LESKOVEC J, KRAUSE A, GUESTRIN C, et al. Cost-effective outbreak detection in networks[C] // Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, USA, 2007: 420-429
- [4] GOYAL A, LU W, LAKSHMANAN L V S. Celf ++ : optimizing the greedy algorithm for influence maximization in social networks[C] // Proceedings of the 20th International Conference Companion on World Wide Web, Hyderabad, India, 2011: 47-48
- [5] CHEN W, WANG Y, YANG S. Efficient influence maximization in social networks[C] // Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 2009: 199-208
- [6] ZHOU F, GAO M. Influence Maximization Algorithm for Social Network Based on Combined Impact Probability [J]. *Computer Engineering*, 2018, 44 (5) : 188-193, 200 (In Chinese)
- [7] LI Y Z, ZHU Y Y, ZHONG M. Social network impact maximization algorithm based on k -kernel filtering[J]. *Computer Applications*, 2018, 38(2) :464-470
- [8] QIU L Q, JIA W, FAN X. Influence maximization algorithm based on overlapping communities[J]. *Data Analysis and Knowledge Discovery*, 2019, 3(7) :94-102
- [9] REN S Y, SHEN D R, KOU Y, et al. Topic-aware influence maximization across social networks[J]. *Journal of Frontiers of Computer Science and Technology*, 2018, 8 (6) :741-752 (In Chinese)
- [10] LI Y, LI V O K. Pricing strategies with promotion time limitation in online social networks [C] // 2018 IEEE/WIC/ACM International Conference on Web Intelligence, Santiago, Chile, 2018:254-261
- [11] ZHAO Y F, SUN G X, BIN S. Research on influence maximization of multi-relationship social networks based on MRLT model[J]. *Application Research of Computer*, 2020, 37(9) :2679-2683 (In Chinese)
- [12] KIM D, HYEON D, OH J, et al. Influence maximization based on reachability sketches in dynamic graphs[J].

- Information Sciences*, 2017, 394-395:217-231
- [13] WANG Y H, FAN Q, LI Y C, et al. Real-time influence maximization on dynamic social streams[J]. *Proceedings of the VLDB Endowment*, 2017, 10(7): 805-816
- [14] ZHANG Y, LI J, YUE K, et al. Influence maximization methods of correlated information propagation[J]. *Journal of Frontiers of Computer Science and Technology*, 2018, 12(12):1891-1902 (In Chinese)
- [15] GUO J F, LU J G. Research on influence maximization algorithm based on information preference[J]. *Computer Research and Development*, 2015, 52(2): 533-541 (In Chinese)
- [16] CAO J X, MIN H Y, WANG H R, et al. Self-interest information influence maximization method based on theme preference in competitive environment[J]. *Chinese Journal of Computers*, 2019, 42(1): 1495-1510 (In Chinese)
- [17] WU A B, VILEN, QIAO B Y, et al. Research on the algorithm of maximizing the influence of large-scale timing diagrams[J]. *Journal of Computer Science*, 2019, 42(12):2647-2664
- [18] WEI L. Research on Influence Maximization Based on Node Degree and Faction [D]. Lanzhou: Lanzhou University, 2019 (In Chinese)
- [19] YANG S X, LIANG W, ZHU K L. Research on algorithms for maximizing oppositional influence in social networks[J]. *Journal of Electronics and Information Technology*, 2020, 40(7): 1944 – 1949(In Chinese)
- [20] CHRISTAKIS N A, FOWLER J H. Social contagion theory: examining dynamic social networks and human behavior[J]. *Statistics in Medicine*, 2013, 32(4): 556-577
- [21] LESKOVEC J, DANIEL H, KLEINBERG J. Governance in social media: a case study of the wikipedia promotion process[EB/OL]. <https://arxiv.org/pdf/1004.3547.pdf>;arXiv, (2010-04-20), [2021-04-20]
- [22] PANZARASA P, OPSAHL T, CARLEY K M. Patterns and dynamics of users' behavior and interaction: network analysis of an online community[J]. *Journal of the American Society for Information Science and Technology*, 2009, 60(5): 911-932

CHEN Jing, born in 1976. She received her Ph.D degree in Yanshan University in 2007. She received her M. S. degree in Yanshan University in 2002. She also received her B. S. degree in Yanshan University in 1999. Her research interests include peer-to-peer network, social computing, and web service.