A Novel Method for Identifying Influential Nodes in Social Networks with Interest Groups

Zhenqiang Guo, Hong Gao, Hongzhi Wang, and Jianzhong Li

School of Computer Science and Technology
Harbin Institute of Technology
Harbin, China
{guozhenqiang,honggao,wangzh,lijzh}@hit.edu.cn

Abstract. Social network has been widely studied these years, because people are more likely to share their thoughts and ideas in social networks, these studies mainly focus on influence maximization. Influence maximization is the problem of identifying some influential nodes in a social network that could maximize the spread of information. However, existing works ignore individual interests, they have irrationalities about the model and the method. Since men may have several interests, they may have different sensitivity to them. In addition, they ignore the particularity of information, but the influence maximization is different for different information propagation. In this paper, we extend LT model, and redefine the influence maximization. We propose a novel method based on these problems, and design an algorithm to solve them. We conduct experiments to verify our idea on real social networks, the experimental results demonstrate that our method is better than the existing methods.

Keywords: social network; influence maximization; interest group; information propagation; algorithm

1 Introduction

Social networks have been widely studied these years. They are composed of nodes and edges, and such complex structure play vital roles in the spread of information, ideas and opinions. Nowadays, people are more likely to share their thoughts and ideas in social networks. When a person accepts a new idea, he may recommend the idea to his friends. Some of his friends will take his recommendation, and further will recommend it to their friends, and more and more people may take this idea, and this is the spread of information. Social networks have many applications, for example, if a company wants to promote a new product, they can use social networks to promote it. They initially gives free product samples to a small number of people in the social network, and hope that the initially selected users will recommend the new product to their friends, and their friends will recommend it to their friends' friends and so on. Eventually, there will be a large number of people to buy this product. In this way, the cost is lower and the effect is better. This problem can be abstracted as the influence maximization. Influence maximization is the problem of finding a small subset of seed nodes in a social network that could maximize the spread of information.

The influence maximization problem is firstly proposed by Domingos[2] and Richardson[13] et al. Later, Kempe[8, 9] et al. also propose a $Top-k$ influence maximization problem. Researchers have proposed many information propagation models, Independent Cascade(IC)[3] and Linear Threshold(LT)[5] are the most popular. In these models, each man in the social network is regarded as a node, and the relationships between people are regarded as edges between nodes. Each node has two possible states : active and inactive. Active nodes denote ones adopt the idea, and inactive nodes denote oppositely. Assume that the state of a node can only be changed from "inactive" to "active". When an inactive node adopts ideas from active nodes, its state becomes active. In IC and LT models, the influence maximization problem has been proved to be NP hard[8], and the accurate evaluation of the influence of given k active nodes has been proved to be $\#P$ hard[1].

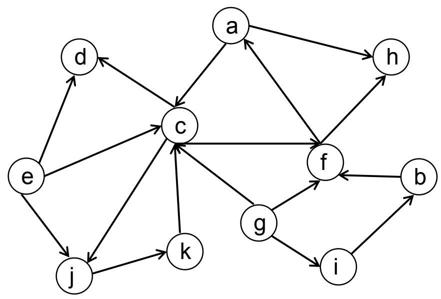


Fig. 1. A Social Network : Teachers

In fact, there are many interest groups in a social network, and there are many nodes in each interest group, a person can exist in some interest groups. For example, Fig 1 is a social network of all science teachers in a high school, and according to their subjects, we can divide the social network of Fig 1 into three sub social networks of Fig 2. A teacher in a network may also exist in other networks, and this is a common situation in real social networks. Previous studies don't take the problem into consideration. Also, a node has different information influence on different interests groups during the spread of different information. For example, if we post mathematical news on Fig 1, nodes a, c, f, h will be more concerned about this news than other nodes, and other math teachers are more willing to accept the news forwarded by them. In other words, a person's influence is related to the spreading information. Existing works don't distinguish different types of information on computing influence maximization, so that the initial active nodes don't have the influence maximization.

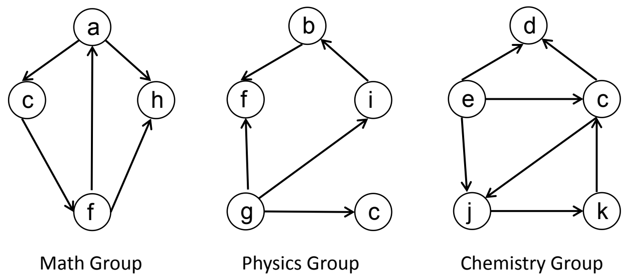


Fig. 2. Three Sub Social Network : Math , Physics , Chemistry

To solve these problems above, in this paper, we extend LT model and propose a novel method to solve then. We conduct experiments to verify our idea on real social networks. Experimental results show that our method is better than the existing methods, and our method is more suitable for the real social network. Our contributions can be summarized as follows:

- We consider that nodes have different sensitivity about different interests, so we extend the LT model. We set different thresholds for different interests of each node, and make the LT model more reasonable.
- We observe that the same active seeds have different influence about different information. We propose the problem that how to identify influential active nodes in social networks with interest groups about different information.
- Since the influence maximization has been proved to be NP hard, we design a novel method to solve the above problem, and experiments show that our method is better than the existing methods on real social networks.

2 Related Work

There are a large number of methods proposed for influence maximization, which focus on the information propagation and analysis of influence in social networks.

For the information propagation, several models have been proposed. Jocb et al. propose Independent Cascade(IC)[3], it is a probabilistic model. Since the active process is uncertain, and the final results of the same network can be significantly different from the same seed nodes, so we don't use it. Mark et al. propose Linear Threshold(LT)[5], an inactive node is affected by all neighbor active nodes, and it will accept the historical influence of other active nodes. The active process is certain, we will introduce it in detail later. There are some other models, but the two above are the most popular.

For the analysis of influence, a number of algorithms have been proposed. These algorithms focus on the running time and influence. Kemple et al. propose a greedy algorithm with an approximation ratio of $(1-1/e)$ to solve the influence maximization. The worst time complexity of the greedy algorithm is $O(n^2(m+n))$, but the real social network is very large, so this algorithm does not work. Leskovec et al. propose the CELF[11] algorithm by mining the submodular of the influence function. This algorithm greatly reduces times of evaluating the impact of seed, and their experiments show CELF improves the performance about 700 times than the greedy algorithm. But for real social networks, it still takes a long time. Goyal et al. proposed CELF++[4] algorithm on the basis of CELF, and the performance is improved about 17% $\sim$ 61%. Kimura et al. proposed the SPM and SP1M model[10], which assume that each node only uses the shortest path for information propagation, and the influence range of each node can be calculated precisely. But this model ignores the importance of the propagation probability between nodes in information propagation. Wang et al.[14] firstly propose that the nodes of the social network can be divided into three states : active, inactive, pseudo active, and consider the value of the pseudo active nodes in information propagation, but the diversity of nodes is not considered. Zhang et al.[15] propose a new method to identify influential nodes based on complex networks with community structure. This method considers the community structure of complex network. However, each node may not only in the same community, but also may exist in some other communities, and this approach does not take it into account.

In addition, some heuristic greedy algorithms are proposed. Tian et al. propose the HPG[7] algorithm based on the heuristic idea. This algorithm is a hybrid algorithm, it uses the $c*k$ seed node heuristic algorithm, and $k-c*k$ seed nodes use greedy algorithm. But in LT model, it does not consider the threshold value of the node in the heuristic stage, so it is unreasonable. Chen et al. propose TBH[6] algorithm based on HPG. TBH considers the threshold value problem of different nodes, it is more reasonable than HPG. But it does not consider the diversity of nodes, because a person may have several interests, the sensitivity of each interest is not the same, so TBH is not very reasonable.

3 Preliminaries and Problem Definition

In this section, we firstly introduce some background knowledge about influence maximization. Then we give a formal statement about our problem. Although we can use different information propagation models to describe our problem, we use the most popular Linear Threshold(LT)[5] model to study our problem in this paper.

3.1 Preliminaries

In the studies of social networks, we usually abstract it into a directed(undirected) network graph $G(V, E)$, the individuals in the social network are regarded as the nodes, and the relationships between the individuals are regarded as the edges. V denotes the set of all the nodes, and E denotes the set of all the edges. Each node has two states, active(buy a product or accept a thought) and inactive(not buy a product or not accept a thought). Active nodes will affect the inactive nodes, the active nodes have no effect on each other. Identifying influential n-odes in social network is to find a small subset of seed nodes in a social network that could maximize the spread of information, namely influence maximization in social network.

The process of information propagation in social networks is that when an inactive node is activated, it will try to activate all its inactive neighbors, when a neighbor is successfully activated, it will try to activate all its inactive neighbors. The process will always go on, until no new node is activated. And the whole process is irreversible, that is to say, the state of a node can only be changed from "inactive" to "active".

At present, researchers have proposed many information propagation models, in these models, Independent Cascade(IC)[3] and Linear Threshold(LT)[5] are the most popular models. Since the active process of LT model is certain, we use LT to solve our problem in this paper, here we will introduce it.

Linear Threshold Model(LT): Linear threshold(LT) is a value accumulation model, each node v has an active threshold $\theta_v \in [0, 1]$, v uses b_{vw} to activate its neighbor w, b_{vw} meet:

$$\sum_{v \in in(w)} b_{vw} \leq 1 \tag{1}$$

Here $in(w)$ is a set of incoming neighbors of w, and node w will be activated on the condition that:

$$\sum_{v \in in(w), active(v) \neq 0} b_{vw} \geq \theta_w \tag{2}$$

That is to say, the cumulative influence of w on the activation of the neighbor is greater than the activation threshold of w.

The characteristic of the LT model is that active process is certain. When we activate the same network with the same seed nodes, the number of activated nodes is the same. When a node v is activated, it will try to activate all its out-inactive neighbor w, and if v does not activate the w, b_{vw} will be accumulated to help other active nodes to activate w. This shows that the active process of LT model is a cooperative active process.

3.2 Problem Definition

We firstly introduce several definitions before our problem defined.

Definition 1. (Influence maximization) *The social network is abstracted as a directed(undirected) graph G(V,E,T,B), where V is a set of nodes, E is a set of edges, T is a set of nodes' active thresholds, and B is a set of influence of each pair of nodes. Given an integer k, influence maximization is to find a set* $S(|S| =$

k), S contains k nodes, so that any set $K(|K|=k)$ of containing k nodes, always have $\sigma(S) \geq \sigma(K)$. S is called the seed set, each node in the S is a seed node.

Kemple[8, 9] et al. propose a greedy algorithm to find the seed nodes. Assuming that S denotes the seed set, $I(S)$ denotes active nodes activated by S, $|I(S)|$ denotes the number of $I(S)$, the algorithm selects a new seed node s to satisfy:

$$argmax \quad m(s|S) = |I(S \cup \{s\})| - |I(S)| \tag{3}$$

The main idea of the algorithm is to find the seed node s with the maximum increment of activation range every time. This greedy algorithm can find an approximate optimal solution of the ratio of $(1 - 1/e)$. The author has proved that the influence maximization is NP hard, and the evaluation of the influence range is proved to be $\#P$ hard. Seen from the process of Kemple greedy algorithm we can, each step of the algorithm needs to calculate every inactive nodes as a seed node to bring the increment $m(s|S)$ of activation range, this process is very time-consuming, so it is not suitable for large-scale social networks.

Interest Groups: The nodes of social network are independent, and these nodes are connected through relationships, such as, family relationships, friend relationships, colleague relationships, etc.

According to the different aspects we can put the nodes of social networks into different relation groups. We use Fig 1 and Fig 2 to illustrate a teacher in a network may also exist in several other networks. For example, a only teaches mathematics, f teaches mathematics and physics, c teaches mathematics, physics and chemistry. This is a common situation in real social networks. However, the LT model does not consider this issue, we extend the LT model and propose a new definition in the following:

Definition 2. (Social Network with Interest Groups) *Given a social network G(V,E,T,B), according to different rules, if it can be divided into multiple social networks $G_1(V_1,E_1,T_1,B_1)$, $G_2(V_2,E_2,T_2,B_2)$,...,$G_n(V_n,E_n,T_n,B_n)$, inside:*

$$V = V_1 \cup V_2 \cup ... \cup V_n, E = E_1 \cup E_2 \cup ... \cup E_n, B = B_1 \cup B_2 \cup ... \cup B_n \tag{4}$$

each social network G_i corresponds to an interest group of network G, the number of social networks depend on the application environment, each node of G can exist in multiple sub networks, the same node may have different active thresholds in different sub networks. G is called Social Network with Interest Groups.

At this time, a book company wants to sell a book in the high school, in order to save cost, the company decides to sell books through the school's social network. They need to choose some influential teachers and hand out these books to them, and if these teachers feel these books are good, they may recommend them to other teachers.

The results are not accurate if we use existing methods to compute seed nodes in Fig 1, since we have to pay attention to the types of books. The physical teacher b may have little interests in a mathematical book, and teacher f is more likely to enjoy it, but he may still have different tastes on mathematical and physical books. It turns out that if we want to sell a book about mathematics and physics, we should try to hand them out to mathematical and physical teachers, chemistry teacher to promote appropriate. So, there will be different results to different information, in order to achieve better results, we must distinguish the information contents.

According to Definition 2, the objects of our study needs to be changed properly. In this paper, we don't discuss how to convert a social network into multiple sub social networks according to the specific application environment, we will study this problem in the future.

Now, we begin to define our problem based on the above description, our ultimate goal is to solve the influence maximization, we introduce a formal definition about influence maximization with interest groups as follows:

Definition 3. (Influence Maximization with Interest Groups) *Given a social network with interest groups G(V,E,T,B), G can be divided into a number of social networks $G_1(V_1,E_1,T_1,B_1)$, $G_2(V_2,E_2,T_2,B_2)$,...,$G_n(V_n,E_n,T_n,B_n)$. Given an integer k, the purpose is to find a seed set $S(S=S_1 \cup S_2 \cup S_3 \cup ... \cup S_n, |S| = k)$ containing k nodes, set $S_i(|S_i| \leq k)$ is from sub social network $G_i(V_i,E_i,T_i,B_i)$, so that for any set $K(K=K_1 \cup K_2 \cup K_3 \cup ... \cup K_n, |K| = k)$ containing k nodes, set $K_i(|K_i| \leq k)$ is from sub social network $G_i(V_i,E_i,T_i,B_i)$, always have $\sigma(S) \geq \sigma(K)$. S is called the seed set, each node in the S is a seed node.*

4 IING Algorithm

In this section, we propose an algorithm to solve influence maximization with interest groups. At first, we illustrate our greedy criteria. Secondly, we discuss that how to evaluate the influence of nodes about different social networks. Finally, we describe the detailed steps of the algorithm based on the previous two steps.

4.1 Greedy Criteria

So far, there have been some algorithms to identify influential nodes based on LT model, and most of them are based on Kemple greedy algorithm, each algorithm has its application environment according to Section 2.

In real social networks, the number of nodes and edges are large. When dealing with real social networks, some of the existing algorithms have poor scalability, so we can reduce the accuracy in order to improve efficiency. Based on the THB[6] algorithm, combined with our model environment, we make some improvements in the greedy algorithm, and propose Definition 4 about the node's potential influence.

Definition 4. (Node's Influence in Single Network) *The value of an edge (v,w) depends on the contribution of v to the activation of w, it is related to the active threshold of w:*

(1) t_w denotes the initial active threshold of node w.
(2) c_w denotes the cumulative value of w from the its in neighbor active nodes.
(3) $b_{v,w}$ denotes the influence of node v on w.
(4) w is an inactive node, $\theta_w(\theta_w = t_w - c_w)$ denotes that the node w becomes active state also needs threshold.

We define the potential influence of edge (v, w) as:

$$p(v,w) = \begin{cases} \frac{b_{v,w}}{\theta_w}, \frac{b_{v,w}}{\theta_w} < 1 \\ \\ 1, \frac{b_{v,w}}{\theta_w} \geq 1 \end{cases} \tag{5}$$

p(v,w) denotes the contribution of edge (v,w) to the activation of node w. Then we define the potential influence PIN(v)of node v:

$$PIN(v) = \sum_{w \in out(v), active(w)=0} p(v,w) \tag{6}$$

PIN(v) is the sum of the effects of v on its own inactive neighbor nodes.

4.2 Influence Evaluation

In section 4.1, we discuss the influence of node in single social network. Now, we discuss how to evaluate the influence of nodes in multi-social networks. In order to get the best spread results, we need to find the most suitable seed nodes. In this paper, the influence maximization is refined, thus we can not use the existing methods to evaluate the nodes' influence. Here we use an example to explain.

We consider a large social network in a school, which can be divided into three social networks: mathematics, physics, and chemistry. A company wants to sell their books with k people. If they sell a mathematical book, the numbers of nodes activated by S_1 in three networks are: 25, 120, 136, and the numbers of nodes activated by S_2 in the three networks are: 100, 15, 21. If we use the existing methods to evaluate the influence, $\sigma(S_1) = 25 + 120 + 136 = 281, \sigma(S_2) = 100 + 15 + 21 = 136$, apparently S_1 is better than S_2. In fact, this is not accurate due to the structure of the social network itself, and from experience we can clearly know: if we sell a mathematical book, although the number of activated nodes in other networks is greater than the mathematical network, activated nodes in mathematics network is more important than any other networks. Similarly, if we sell a physical book, the activated nodes in the physical network are more important than others. If we want to sell a science book, we think that activated nodes in these three networks are equally important.

We have already known the fact, so we have to quantify the importance, the following formula is proposed:

$$\sigma(S) = \lambda_1 \cdot \sigma_1(S) + \lambda_2 \cdot \sigma_2(S) + \lambda_3 \cdot \sigma_3(S) + ... + \lambda_n \cdot \sigma_n(S), 0 \leq \lambda_i \leq 1 \tag{7}$$

parameter λ_i is used to adjust influence σ_i of the sub network G_i, which can be determined according to the application environment and information.

4.3 Algorithm

In this paper, we combine the Formula 7 with $PIN(v)$ to select seed nodes quickly, and then propose Algorithm 1 based on greedy algorithm as follows.

Algorithm 1 IING: Identifying Influential Nodes Greedy Algorithm

Input: Graph $G_1(V_1, E_1, T_1, B_1)$, $G_2(V_2, E_2, T_2, B_2)$,...,$G_n(V_n, E_n, T_n, B_n)$,integer k $\lambda_1, \lambda_1, ..., \lambda_n$
Output: Seed set S

```
1: Initialize set S=Φ;
2: Compute each node Initialize PIN; //PIN needs to combine with Formula 7
3: Build a container sequence Q; //the elements of the container can be automatically
   sorted according to PIN value from large to small
4: Put all nodes to the container Q;
5: while |S| < k do
6:     s is the first node in Q, S = S ∪ {s};
7:     Remove s from Q;
8:     Use node s to active networks;
9:     Update the PIN value of the node in the activation process;
10:    Update the order of elements in the Q;
```

5 Experiment

In this section, we compare our IING algorithm with THB[6] algorithm and perform experiments on real social networks. Then we demonstrate the performance of the proposed algorithm.

5.1 Data Sets

In principle, we need to divide a social network with interest groups into multiple sub social networks, but we haven't studied it in this paper, so we pass it and put it in the future work. In order to ensure that our experiments can be carried out, we select two real social networks: Epinions and Slashdot0811[12], they come from Stanford University's large web data collection site (`http://snap.stanford.edu/data/index.html`). And the two social networks are processed. We think that concerning ourself is meaningless, so we remove these edges. The statistics of the two social networks processed are shown in Table 1.

Because the nodes of the two networks are numbered, there may exists nodes in a same number, and the structure of the two networks is different, we can assume that the two networks come from the same network ES, which are two

No	Data Set	Nodes	Edges	Average Degree
1	Epinions	75879	508837	6.7
2	Slashdot0811	77360	828161	10.7

Table 1. Information of Data Sets Epinions and Slashdot0811

Data Set	Nodes	Edges	Average Degree
ES	77360	1335128	17.3

Table 2. Information of Data Set ES

interest groups of network ES, thus the problem of identifying multiple interest groups in a network is solved. We want to compare with the TBH algorithm, but TBH algorithm is only adapted to influence maximization of single network, for this reason, we combine Epinions and Slashdot0811 into ES. The data statistics of ES are shown in Table 2.

5.2 Experiment Setup

Parameter setting: Our data sets has been specially processed, and it is difficult to accurately obtain the thresholds and influence data. So we use a simple method to obtain the data.

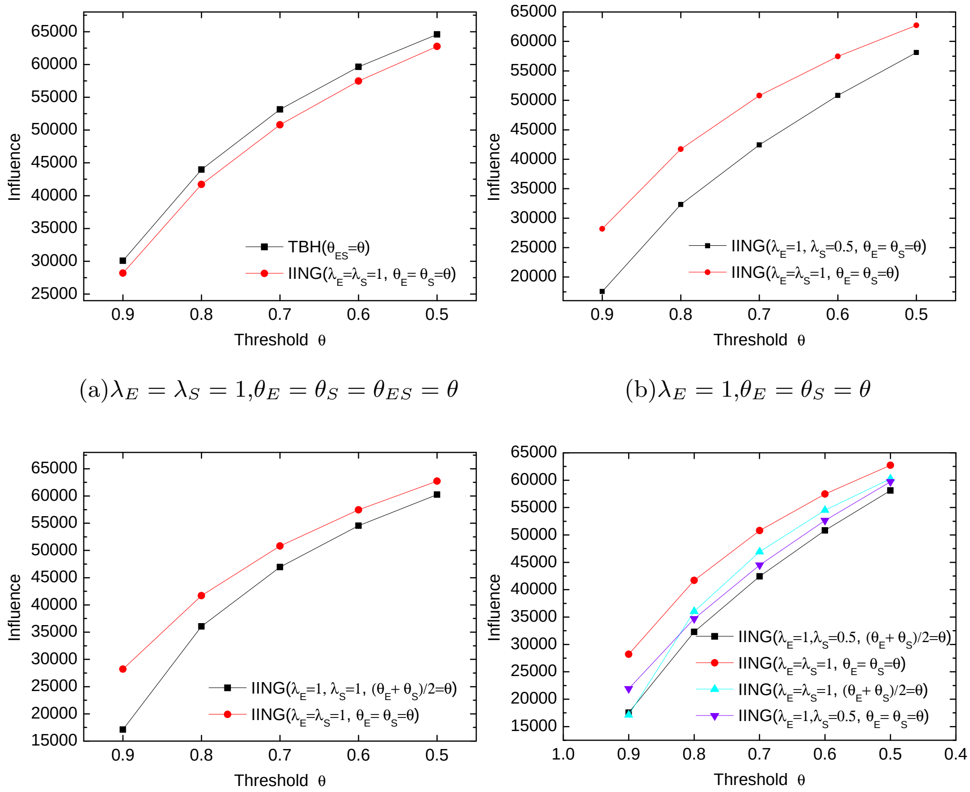
For the network ES, we randomly generate an influence $b_{uv}(0 \leq b_{uv} \leq 1)$ for each edge $(u \rightarrow v)$, and we set the active threshold of all nodes as $\theta_{ES}(0 \leq \theta_{ES} \leq 1)$. For the network Epinions and Slashdot0811, because the network Epinions and network Slashdot0811 are from the network ES, if the edge $(u \rightarrow v)$ in the network Epinions is also present in the network Slashdot0811, we randomly generate an influence $b^E_{uv}(0 \leq b^E_{uv} \leq b_{uv})$ to network Epinions, the $b^S_{uv}(b_{uv} - b^E_{uv})$ to network B; if the edge $(u \rightarrow v)$ in the network Epinions is not in the network Slashdot0811, we give $b^E_{uv}(b^E_{uv} = b_{uv})$ to the network Epinions; if the edge $(u \rightarrow v)$ in the network Slashdot0811 is not in the network Epinions, we give $b^S_{uv}(b^S_{uv} = b_{uv})$ to the network Slashdot0811; similarly, all nodes in Epinions and Slashdot0811 use the same threshold $\theta_E(0 \leq \theta_E \leq 1)$ and $\theta_S(0 \leq \theta_S \leq 1)$. As for the influence parameters λ_E and λ_S of network Epinions and network Slashdot0811, we need to combine the information to be spread, and we change the size of the two parameters to observe the change of influence.

Algorithm for Comparison: Due to the traditional greedy algorithm takes too long time, we only compare THB[6] algorithm with our algorithm. The performance is evaluated from two aspects, the influence scope and the running time.

Experiment Environment: We implement the algorithms in Java and conduct the following experiments on a Linux 3.4GHz Four-Core Intel Core I7-6400 and 16G memory.

5.3 Experiment Results

We mainly test the influence scope and the running time. Our algorithm is carried out on the network Epinions and network Slashdot0811, and the TBH algorithm is carried out on the network ES.



(a)$\lambda_E = \lambda_S = 1$,$\theta_E = \theta_S = \theta_{ES} = \theta$ (b)$\lambda_E = 1$,$\theta_E = \theta_S = \theta$

(c)$\lambda_E = \lambda_S = 1$,$(\theta_E + \theta_S)/2 = \theta$ (d)$(\theta_E + \theta_S)/2 = \theta$,

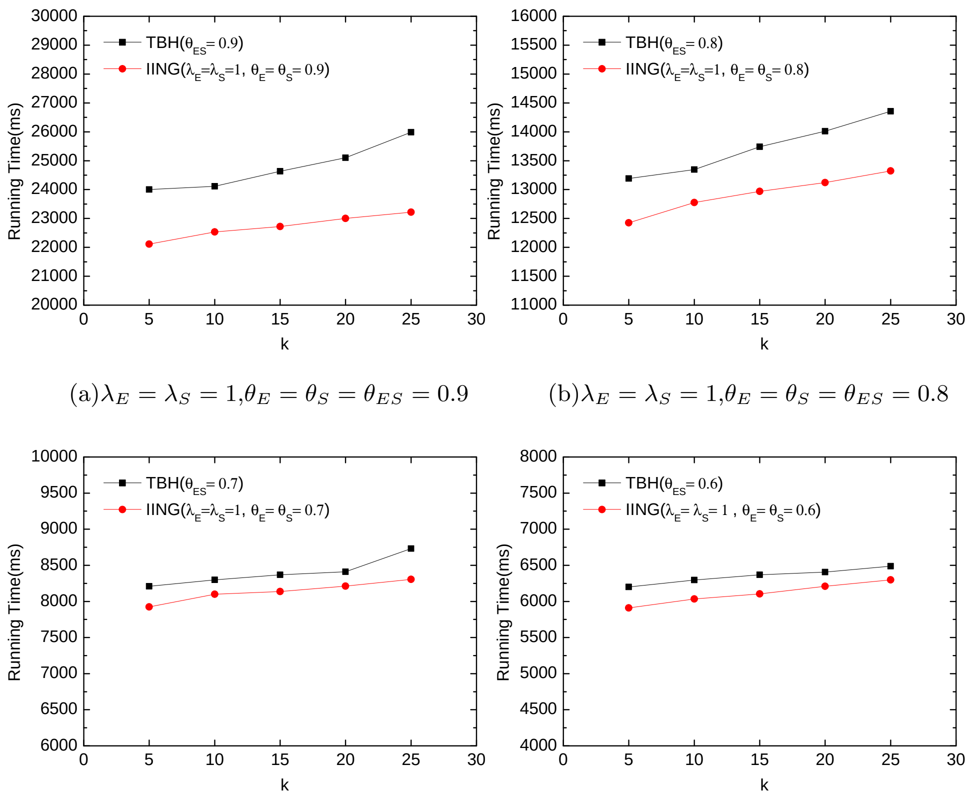
Fig. 3. Influence of fixed $k = 5$, fixed influence factors λ, and different thresholds θ

Influence Scope: The influence of the fixed $k = 5$, different active thresholds and influence factors are compared. Experiments are conducted from the following four aspects:

(1)We set $\lambda_E = \lambda_S = 1$, $\theta_{ES} = \theta_E = \theta_S = \theta$, Fig 3(a) shows the influence of the two methods is very close. The set of parameters of our method is equivalent to not considering interest groups, and the experimental result is also consistent with our experience.

(2)We set $\lambda_E = 1$, $\theta_E = \theta_S = \theta$, Fig 3(b) shows the influence of the our method at different information. We think that the influence of different information on network Slashdot0811 is different, the experimental result has great changes.

(3)We set $\lambda_E = \lambda_S = 1$, $(\theta_{ES} + \theta_E)/2 = \theta$, Fig 3(c) shows the influence of the our method at different active thresholds. We changed the active thresholds of the network Epinions and Slashdot0811, because a person has different tastes to his different interests, we observe that the experimental result are quite different.



(a)$\lambda_E = \lambda_S = 1$,$\theta_E = \theta_S = \theta_{ES} = 0.9$ (b)$\lambda_E = \lambda_S = 1$,$\theta_E = \theta_S = \theta_{ES} = 0.8$

(c)$\lambda_E = \lambda_S = 1$,$\theta_E = \theta_S = \theta_{ES} = 0.7$ (d)$\lambda_E = \lambda_S = 1$,$\theta_E = \theta_S = \theta_{ES} = 0.6$

Fig. 4. Running time of fixed thresholds θ, fixed influence factors λ, and different k

(4)We set $\lambda_E = 1$, Fig 3(d) shows the influence of the our method at different active thresholds and information. This is the real situation, the experimental result shows that there is obvious difference with the previous experiments.

The above experimental results demonstrate that our idea is correct, when we change the influence factor and the active threshold of the network, the influence of the seed nodes we selected is closer to the real results.

Running Time: In addition, our experiments compare the running time of different seed numbers k , fixed thresholds θ and influence factor λ. In this part, we all set $\lambda_E = \lambda_S = 1$. In order to obtain accurate results, for each case, we run algorithms 10 times to find the average time. Fig 4 (a) $\sim$ (d) are the average running time of the two methods.

As it can be seen in Fig 4 (a) $\sim$ (d), the running time of the two methods have very small differences in the same k, we can even ignore these differences. With the increase of k, the growth rate of running time is very slow. Our method and THB have the same efficiency.

6 Conclusion

In this paper, to begin with, we extend the LT model and make the LT model more reasonable. And we propose the problem that how to identify influential active nodes in social networks with interest groups about different information. Furthermore, we propose a novel method based on the problem, and design an algorithm to solve it. Finally, we conduct experiments to verify our idea on real social networks. Experimental results show that our method works very well.

However, our method still needs to be improved. Several problems are still existing: how to divide a network into multiple networks; how to rationally determine the nodes' threshold in different interest groups; how to determine the contribution ratio of each network, and so on. We will focus on these problems in the future.

References

1. W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010*, pages 1029–1038, 2010.
2. P. M. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco, CA, USA, August 26-29, 2001*, pages 57–66, 2001.
3. J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Journal of Computer Research and Development.*, 12(3):211–223, 2001.

4. A. Goyal, W. Lu, and L. V. S. Lakshmanan. CELF++: optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011 (Companion Volume)*, pages 47–48, 2011.
5. M. Granovetter. Threshold models of collective behavior. *American journal of sociology.*, 83(6):1420, 1978.
6. C. Hao and W. Yitong. Threshold-based heuristic algorithm for influence maximization. *Journal of Computer Research and Development.*, 49(10):2181–2188, 2012.
7. T. Jiatang, W. Yitong, and F. Xiaojun. A new hybrid algorithm for influence maximization in social networks. *Chinese Journal of Computers.*, 34(10):1956–1965, 2011.
8. D. Kempe, J. M. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 24 - 27, 2003*, pages 137–146, 2003.
9. D. Kempe, J. M. Kleinberg, and É. Tardos. Influential nodes in a diffusion model for social networks. In *Automata, Languages and Programming, 32nd International Colloquium, ICALP 2005, Lisbon, Portugal, July 11-15, 2005, Proceedings*, pages 1127–1138, 2005.
10. M. Kimura and K. Saito. Approximate solutions for the influence maximization problem in a social network. In *Knowledge-Based Intelligent Information and Engineering Systems, 10th International Conference, KES 2006, Bournemouth, UK, October 9-11, 2006, Proceedings, Part II*, pages 937–944, 2006.
11. J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, San Jose, California, USA, August 12-15, 2007*, pages 420–429, 2007.
12. J. Leskovec and A. Krevl. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, June 2014.
13. M. Richardson and P. M. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23-26, 2002, Edmonton, Alberta, Canada*, pages 61–70, 2002.
14. Z. Wang, E. Chen, Q. Liu, Y. Yang, Y. Ge, and B. Chang. Maximizing the coverage of information propagation in social networks. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 2104–2110, 2015.
15. X. Zhang, J. Zhu, Q. Wang, and H. Zhao. Identifying influential nodes in complex networks with community structure. *Knowl.-Based Syst.*, 42:74–84, 2013.