International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.10 (2016), pp.67-82
http://dx.doi.org/10.14257/ijmue.2016.11.10.06

# Influence Maximization Social Networks Hybrid Algorithm Based on Linear Threshold Model

Dou Ruofei[1]

[1]*College of Computer Science and Information Engineering Tianjin University of Science & Technology, Tianjin, China*
*E-mail: teacherdou@163.com*

***Abstract***

*Aiming at the problems such as the HIM algorithm has a very high time complexity and cannot be applied in the large social networks and so on, and this paper puts forward the hybrid influence maximization algorithm. The algorithm carries out the detailed experiments on 6 real data sets with different characteristics. The experiments show that compared with the HIM algorithm, the algorithm in this paper has great improvement in the scope of influence.*

***Keywords****: Propagation model; Linear threshold; Node selection; Potential-influence*

## 1. Introduction

In the real world, many systems exist in the form of complex network, such as the Internet, the social system, the computing network, the biological network and the social network, *etc.* To understand the topology structure and function of the complex network is the current international research hot spot. Many areas of science all use the Internet to show the relationship between the members in the system, for example, the social network use the nodes to represent people, and the edges to represent the connections between people [1-3]. Social network refers to a complex network that is composed of the individuals and the relationship between the individuals. This kind of complex social structure plays a vital role in the information dissemination and diffusion. When a person is to adopt a new idea or to accept a product, he would recommend it to his friends or colleagues. Some people may accept or follow his recommendation, and further recommend the ideal or product to their friends or colleagues. This process is called the propagation or spreading. A person's behavior depends largely on the decisions of the surrounding friends or colleagues.

The influence maximization problem in social network, which is about how to choose k seeds to achieve the maximum range of influence after their propagation process, has been proven to be a question of NP2hard. Kemple and Mr. Kleinberg proposed a natural climbing-hill greedy algorithm that chooses the current nodes which are "most influential" as the initial propagation targets to carry out the propagation in every step [6-7]. The so-called "most influential" nodes are namely the current nodes that can activate the nodes at most. However, the process of selecting the "most influential" nodes is very time-consuming, and this kind of local optimization cannot guarantee the final optimal results of the propagation. For the large social networks, this greedy algorithm is more inapplicable due to its high time-consuming. Except for the greedy algorithm, there are some common heuristic strategies for node selection, including the point-based out-degree, the degree centrality and so on. Nevertheless, in terms of transmission range, the effect of the heuristic rule that is completely based on the out-degree is not ideal. That is because this method obviously fails to take into account the propagation characteristic of the social network. Of course, the information diffusion has its own rules or models. At present, all the researches on the influence maximization problem of the social network

ISSN: 1975-0080 IJMUE
Copyright ⓒ 2016 SERSC

are based on the following two basic propagation models: Linear Threshold Model and Independent Cascade Model.

At present, the main research work of the influence maximization problem is focused on using the sub-modularity to improve the operating efficiency of the greedy algorithm, under the IC model. Thus, this paper considers the influence maximization problem under the IC model. Through investigation, it is found that the LT model has the property of "influence accumulation". And by using this property, a hybrid algorithm for influence maximization is proposed in this paper, which improves the final range of influence and reduces the running time of the algorithm through the comprehensive consideration of the structure characteristic and the transmission characteristic of network.

## 2. Influence Maximization

The social network is a complex network composed of the relationships between the individuals, and the influence of the nodes in the social network not only refers to the mutual influence between the nodes directly connected, but also includes the inner link between the nodes. For the problem of the influence of the nodes, the simply analysis of the node's certain single attribute often cannot get the most influential individual nodes. For this kind of problem, the influence maximization analysis is usually adopted to deal with it. The influence maximization analysis is to purposefully select k users in a specified network as the initial set, making the users in this collection can eventually affect the most individual users in the network. The traditional local optimal selection strategy is a very time-consuming process, and this kind of local optimum cannot guarantee the final optimal results. And the influence maximization analysis treats the selection of the "most influential" nodes as the goal, and the so-called "most influential" nodes are not limited to the current most influential nodes, but the most influential nodes in the whole process.

Combine the properties of the nodes in the influence process, and the linear threshold is adopted as the analysis model of the influence of the individual nodes. In the linear threshold model, the nodes have two states: the active state and the inactive state. The nodes in the active state have influence on the nodes in the inactive state, and if this effect results in a certain node change from the inactive state to the active state, then this process is known as the activation. The more the neighbor nodes of the certain node are activated, the greater the probability that the node is activated will be. And the newly activated nodes will affect other neighbor nodes in the inactive state, thus with the passage of time, a growing number of nodes will change from the inactive state to the active state. The diffusion process of the linear threshold model is as follows: given the initial set of the transmission nodes $s_0$ , the specific threshold of all the nodes $\theta_v$ , and the influence between nodes $u_{uv}$ . In the step t of spreading, based on the modes that arrive at the threshold by the activation of the set $s_{t-1}$ , the activated node are added to the set $s_{t-1}$ to form the set $s_t$ , and repeat this process, until no new nodes are activated.

In the linear threshold model, $u_{uv}$ represents the existed influence of the activated node u on the adjacent point $u$ . Aiming at the social networks with different features, and considering that the individuals in the actual social network all have their own characters and status, the intermediate centrality and the calculation close to the centrality are taken into account in the influence of the nodes. The estimation formula of the improved influence of the node $u_{uv}$ is follows:

$$u_{uv} = \frac{outDeg(u)}{\sum_{w \in N(v)} outDeg(w)} + \frac{w_{AB}(u)}{n^2 - 3n + 2} + \frac{w_{AP}^{-1}(u)}{n-1} \text{(Without right)} \quad (1)$$

 Copyright ⓒ 2016 SERSC

$$u_{uv} = \frac{W(u,v)}{\sum_{w \in N(v)} W(w,v)} + y_{AB}(u) + y_{AP}^{-1}(u)(\text{With right}) \tag{2}$$

In the formula, $u_{uv}$ is the existed influence of the individual node u on the adjacent point v; W represents the other individual nodes around node v; n is the number of individual nodes in the network of node $u$; $W(u,v)$ is the weight on the edge $(u,v)$; $N(v)$ is the set of the neighbor nodes around node $v$; CAB(u) is the intermediate centrality for calculating one node u; CAP(u) is the derivative of the node close to the centrality.

According to the basic framework of the influence maximization algorithm, and combine with the centricity property of the individual node, the influence maximization algorithm based on the node position is put forward. The algorithm divides the selection of $h$ target nodes into two stages: the inspiring stage and the target phase. The inspiring stage chooses the odes with maximum value of PI, and the target phase selects the most influential individual nodes as the ultimate goal set, according to the centricity property of the node.

## 3. Hybrid Algorithm for Influence Maximization

### 3.1. Presentation of Framework and the HPG Algorithm

The natural climbing-hill greedy algorithm which is proposed by Kemple and Kleinberg (herein after referred to as HIM algorithm) is very costly, and the local optimization cannot guarantee the final optimal range of influence, which here refers to the number of nodes that can be activated. Therefore, by making use of the “influence accumulation” property of the LT model and synthesizing the structure characteristic and the transmission characteristic of network, the framework to solve the problem is put forward, and based on this framework, a new selection strategy for initial nodes, the hybrid potential-influence greedy algorithm (HPG), is proposed, which is a hybrid algorithm for influence maximization. According to the “influence accumulation” property of the LT model, it can be known that although a node u fails to activate node $v$, its influence $u_{uv}$ is accumulated. Therefore, the starting point is to quickly find some nodes with potential influence, rather than spending a lot of time looking for the "most influential" nodes. Although these nodes cannot activate the most nodes at present, they can accumulate a large number of “potential influence”, which makes there are more nodes being activated in the latter stage. Thus, the running time can be greatly saved, for it is very costly to look for the “most influential” nodes. It is needed to calculate the diffusion scope of all the current inactivated nodes, which is particularly time-consuming in the initial stage of the algorithm, namely when most of the nodes are inactivated. Based on this starting point, the framework to solve the problem is put forward, whose core is the selection process of the nodes, and it is divided into two parts. 1) The enlightening stage. Select the nodes with the most potential influence, which cannot bring the biggest range of influence in the current, but contain a huge potential influence. 2) The greedy stage. Select the most influential nodes.

The nodes with a big out-degree are often at the center position of the social network. And the HIM algorithm considers the spreading process of the influence, thus can achieve the 63% of the optimal value. Comprehensively consider the advantages of the degree centrality and the greedy algorithm and combined with the characteristic of the LT model, firstly choose the nodes with the most potential influence to be activated. Although there are cannot be most nodes to be activated in the current, in the next greedy stage, a big part of the nodes that are not activated but have accumulated a lot of potential influence are explosive. This framework is not a local optimum, but its ultimate scope of influence is bigger.

In order to find the nodes with the most potential influence, the analysis and experiment of the network structure are carried out. And it is found that there are two factors that can affect the selection of the nodes with the most potential influence: the out-degree of the node and the sum of the influence of an activated node on all of its inactivated neighbor nodes, and the effects of the two factors are not the same. According to the experimental results, it is known that the out-degree of the nodes has a more obvious contribution to the final scope of influence. Therefore, synthesize the above reasons, the following formula is given, which makes the selection of the nodes with the most potential influence gives full priority to the out-degree of the node, and comprehensively consider the $\inf$ factor. The potential influence (PI) is defined as follows:

$$\inf(u) = \sum_{v \in N(u), v \notin A(u)} u_{uv} \tag{3}$$

$$PI(u) = outDegree(u) + (1 - e^{-\inf(u)}) \tag{4}$$

Among which, N (u) shows the set of the outside neighbor nodes of node u, A (u) represents the activated nodes in (u), and $u_{uv}$ stands for the influence of node u on node v. Therefore, $\inf$ (u) presents the sum of the influence of node u on all of the inactivated neighbor nodes, which is called the influence of the node u, and is determined by two factors: the number of the inactivated neighbor nodes of node u and the size of $u_{uv}$. Out Degree (u) shows the out-degree of node u. When the out-degrees of the nodes are the same, the nodes with a relatively big influence are selected, rather than blindly and randomly selecting a node with the biggest out-degree as the current node with the most potential influence. Thus, the node with “the most potential influence” is the node with the largest value of PI. Obviously, the calculation of the marginal influence range of each node is simplified to the calculation of the value of PI of each node. Here, the value of PI can be calculated in constant time.

What calls for special attention is that when dealing with social networks with signs, the value of out Degree (u) in formula (2) needs for special treatment. Theoretically, the current node with the biggest out-degree is selected, but due to the existence of the negative side, the value of out Degree (u) is redefined as follows:

$$outDegree(u) = outDegree_{+}(u) - outDegree_{-}(u) \tag{5}$$

Among which, $outDegree_{+}(u)$ represents the number of positive edges in out-edges of $u$, and $outDegree-(u)$ shows the number of negative edges in out-edges of u. This is because the out-degree of u contains the positive edge and the negative edge, in which the positive edge has a positive influence on the spreading of influence, while the negative edge has a negative influence on the spreading of influence, and the degree centrality of node u is represented by using the positive effects to minus the negative effects.

In the given framework, the selection of the $h$ initial nodes is divided into two stages: the enlightening stage and the greedy stage. In the enlightening stage, select the node with the maximum value of PI, and in the greedy stage, select the node with the "most influential" node. The inspiring factor $c(c \in 0,1)$ is introduced in the framework, $[ck]$ shows the step number of the greedy stage, and $k-[ck]$ shows the step number of the enlightening stage. Obviously, when there is $c=1$, the method in the framework is the HIM algorithm.

 Copyright ⓒ 2016 SERSC

### 3.2. Improvement of the $u_{uv}$ Estimation Formula

In the linear threshold model, $u_{uv}$ represents the existed impact of the activated node u on its adjacent node v, which is estimated by $1/d(v)$ shows the out-degree of node v) and means that the influences of all the neighbor nodes on node v are the same. Obviously, this assumption ignores the differences between nodes, and does not conform to the reality. Here, according to the different characteristics of the social network, different $u_{uv}$ estimation formulas are given.

#### 3.2.1. $u_{uv}$ Estimation Formula on the Un-Weighted Graph

The size of $u_{uv}$ is a feeling of the node v itself, this feeling is a reflection of the authority of node v to the node that is pointed to node v, and has nothing to do with other nodes. Thus, there is only need to consider the structure relationship of the neighbor nodes of node v. The Neighbor Graph (NG) is made up of the node v, the neighbor nodes that are pointed to node $v$, and the relationship between the nodes. The degree of the nodes in the $u_{uv}$ estimation formula is obtained according to the degree of the nodes in the neighbor graph. Its definition is given in formula (4).

$$\begin{aligned} &NG(v) = G'(V', E'), V' = \{v\} \cup N(v), \\ &E' = \{(x, y) \mid x, y \in V', (x, y) \in E\}, \\ &u_{uv} = \frac{outDeg(u)}{\sum_{w \in N(v)} outDeg(w)} \end{aligned} \tag{6}$$

Among which, $outDeg(u)$ represents the out-degree of node $u$ in the neighbor graph, and $N(v)$ shows the set of the into-side neighbor nodes of node v in the neighbor graph. The influence of node u on node $v$ is mainly decided by the neighbor structure of node $v$.

Figure 1 shows a simple example of the neighbor graph, in which dark node $v$, node $u_1$, $u_2$ and $u_3$ form a neighbor sub-graph. The out-degree of $u_1, u_2$ and $u_3$ in NG (v) is respectively 1, 1 and 2. Thus, the value of $u_{u1v}$, $u_{u2v}$ and $u_{u3v}$ is respectively 0.28, 0.28 and 0.7.

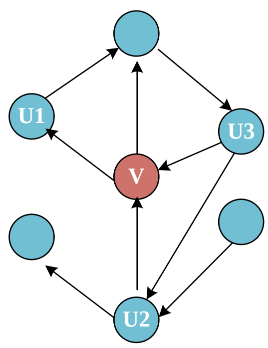



**Figure 1. The Neighbor Graph of Node V**

### 3.2.2. $u_{uv}$ Estimation Formula on the Weighted Graph

Under the condition of considering the weighted edge, the influence of node u on node v is mainly determined by the weight of the edge. The definition of $u_{uv}$ is specified as follows:

$$u_{uv} = \frac{s(u,v)}{\sum_{w \in N(v)} s(s,v)} \tag{7}$$

Among which, $s(u,v)$ represents the weight of the edge $(u,v)$, and $N(v)$ shows the set of the into-side neighbor nodes of node $v$ in the neighbor graph.

Figure 2 shows a simple example, in which the weight of edge $(u_1,v)$, $(u_2,v)$ and $(u_3,v)$ is respectively 2, 5 and 3. According to formula (3), the calculated value of $u_{u3v}$, $u_{u2v}$ and $u_{u1v}$ is respectively 0.3, 0.6 and 0.4.

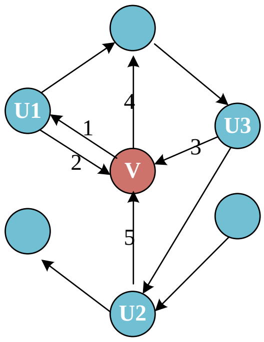



**Figure 2. The Weighted Neighbor Graph of Node V**

### 3.2.3. $u_{uv}$ Estimation Formula on the Signed Graph

In the past, the researches of a large number of social networks only focus on the traditional social network, and there are relatively little researches about the signed social networks. The signed social network refers to that there not only exists the positive relationship between the individuals, but also exists the negative relationship, in which the positive relationship has a positive effect on the spreading of influence, and the negative relationship has a negative effect on the spreading of influence. In the real social networks, it is very important to consider the interaction between the positive relationship and the negative relationship. For example, in the information technology website Slashdot, the users can be labeled as “friend” or “enemy” between each other, in the review site Epinions, the users can express “trust” or “distrust” between each other and so on.

Abstract a signed social network as a directed graph s (e, v) with positive side and negative side, in which the symbol of the edge represents the positive and negative of the influence. As shown in Figure 3, Triangle (a, b, c) is defined as the triangle formed by three directed edges (a, b), (a, c) and (b, c), in which x, y and z show the symbol of the edges, namely the positive and negative of the influence. In the signed graph, when considering the influence of a on c, a has a direct impact on c, and has an indirect effect on c through b, but it's important to note that not all of the indirect effects are effective. According to the Balance principle of the signed graph, the multiplication rule is given: to make the indirect effect of a on c through b be effective, then it must meet that $z = x \times y$. With the multiplication rule, the calculation of $u_{uv}$ on the signed graph can be carried out through the neighbor graph.

 Copyright ⓒ 2016 SERSC

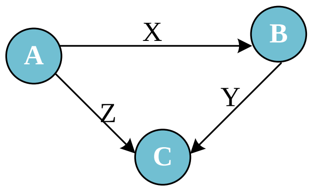



**Figure 3. The Triangle Formed by (A, B, C) in the Signed Graph**

Under the condition of considering the existence of the negative side, the definition of $u_{uv}$ is specified as follows:

$$u_{uv} = \frac{sign(u,v) \times s(u,v)}{\sum_{w \in N(v)} s(s,v)} \tag{8}$$

Among which, $s(u,v)$ refers to the influence weight of node $u$ on node $v$ in the neighbor graph, rather than the weight of the edge, and it is determined by the directly affect of edge $(u,v)$ and the indirect influence of Triangle $(u,u',v)$. At first, investigate the directly affect of edge $(u,v)$, and initialize the value of $s(u,v)$ to 1, then investigate the indirect influence. If there exists Triangle $(u,u',v)$ that satisfies the multiplication rule, then add 1 to the value of $s(u,v)$, or do nothing.

Figure 4 shows a simple example. Triangle $(u_1,u'_2,v)$ satisfies the multiplication rule, while Triangle $(u_1,u'_3,v)$ does not meet the multiplication rule, thus, the influence weight of $u_1$ on $v$ in the neighbor graph is 3.

In the graph, there are $s(u_1,v)=3$, $s(u_2,v)=3$, $s(u_3,v)=2$, $s(u_4,v)=2$ and $s(u_5,v)=1$. Thus, according to formula 3, the calculated value of $u_{u1v}$, $u_{u2v}$, $u_{uv}$, $u_{u4v}$ and $u_{u5v}$ is respectively 0.28, 0.28, 0.127, 0.28 and 0.127.

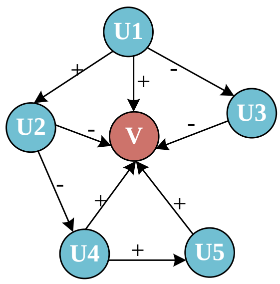



**Figure 4. The Signed Neighbor Graph of Node V**

### 3.3. Time Complexity Analysis of HIM Algorithm and HPG Algorithm

The brief explanation and analysis on the running time of the two algorithms are carried out below. HIM algorithm considers the whole propagation process of the influence, thus, it is need to calculate the marginal influences of all the inactivated nodes in every step. At first, all the nodes in the graph is in the inactivated state, and to calculate the marginal influence of each node, it will need to traversal the entire graph to carry out the propagation of the influence, which is very time- consuming.

HPG algorithm is an algorithm for influence maximization under the proposed framework and when the value of the inspiring factor c is 0.4. The time complexity analysis of HPG algorithm is from two stages.

In the enlightening stage, each step is to choose the node with the largest value of PI in the inactivated nodes. The calculation of the PI value basically does not consume time (its time complexity is an constants, and is expressed with O(C)), it is because the value of outDegree (u) in formula (1) invariable, and the value of inf (u) have been calculated when updating the entire graph after the determination of the last initial seed. When the enlightening stage is over, there will be a lot of potential influences accumulated in the graph and a large number of nodes activated at the same time.

In the greedy stage, each step selects the node with the biggest marginal influence, but after the enlightening stage, there are a large number of nodes be activated in the graph. At this moment, the nodes without activation is much less than that of the original data set, and the HIM algorithm will correspondingly traverse the graph with less times, which can be regarded as the running of HIM algorithm on small data sets. Therefore, the time complexity of HIM algorithm is much smaller than that of HPG algorithm.

The contrast of time complexity between HPG algorithm and the two stages of HIM algorithm is shown in Table 1. HIM algorithm is accordingly divided into two stages, each stage selects k / 2 initial nodes. What calls for attention is that with the decreasing of the number of the inactivated nodes, namely with the carrying out of the selection of the initial nodes, the running time of the greedy algorithm will also monotonically reduce. Assume that there are k initial nodes that need to be selected, and there are N nodes and M sides in the graph.

**Table 1. Time Complexity of the Two Stages of Hpg Algorithm and Him Algorithm**

| | HPG algorithm | HIM algorithm |
|---|---|---|
| The enlightening stage | Large amount of inactivated nodes | Large amount of inactivated nodes |
| Time complexity | $w_1 + w_2 = k/3 \times O(s)$ | $w'_{i+w} = k/2 \times O(s)$ |
| The greedy stage | Small amount of activated nodes | Small amount of activated nodes |

In Table 1, shows the meaning of far less. The time complexity of HPG algorithm is $w_1 + w_2 = k/3 \times O(nNm)$ , and the time complexity of HIM algorithm is $w'_{i+w} = k/2 \times O(nNm)$. Therefore, the time complexity of HPG algorithm is much lower than that of HIM algorithm.

## 4. Experimental Simulation and Analysis

### 4.1. Experimental Data

The experiments are carried out on 3 real data sets, and the statistical information are shown in Table 2. The data set 1 is the interaction between the yeast and protein, and it is a indirected and unweighted graph. The data set 2 is a social network for author cooperation, in which the point represents the author, and the side shows that there is a relation of cooperation between the two authors, and it is a indirected and unweighted graph.

 Copyright ⓒ 2016 SERSC

**Table 2. The Statistical Information of the Data Set**

| Sequence number | Data set | Number of nodes | Number of sides | Average degree | Social semantic |
|---|---|---|---|---|---|
| 1 | yeast | 7401 | 12021 | 12.0 | interaction between protein |
| 2 | information technology website | 78031 | 39842/121323 | 14.8 | Wiki vote |
| 3 | Epinions | 131901 | 72134/123874 | 13.0 | review website |

The data set 1 is a weighted social network for author cooperation, in which the point represents the author, and the weight on the side shows the number of cooperation times between the authors, it is a representative of the weighted social networks and an indirected and weighted graph. The data set 2 is a voting history network of Wikipedia, in which the point represents the user, the directed side from u to $v$ means that u votes for $v$, and it is a directed and unweighted graph. The data set 3 is a friend or enemy network from the Slashdot website, in which the point represents the member of the website, the directed side from $u$ to $v$ means that u marks $v$ as a friend or en enemy (determined by the symbol on the side), and it is a signed and directed graph. The data set 3 is a trust network from the Epinions, in which the node represents the member of the website, the directed side from u to v means that u trusts or distrusts $v$ (determined by the symbol on the side), and it is a signed and directed graph. The 3 data sets are the data sets that are openly used for various tests in the field of the social network, which have different characteristics (weighted / unweighted, signed / unsigned). It is noted that for data sets Wiki_vote, Slashdot and Epinions, the backward processing is needed for the original figure, because the research is the influence maximization. Thus, when v votes for u, trusts u or marks u as a friend, it is treated as the influence of u on v.

### 4.2. Experimental Environment

The experiment is carried out based on the linear threshold model, in which the value of buv is calculated according to the estimation formula that is given in Section B. In the literature <Maximizing the spread of influence in a social network> of Kempe $B$, a classic method for the evaluation of the threshold value $\theta_v$ is given, which is to fix the threshold value of each node to 0.6. The inspiring factor c means that the greedy stage has ck steps, and the enlightening stage has k-ck steps. The algorithm given in the literature of Chen Wei reduces the time complexity of the greedy algorithm, but it fails to improve the final range of influence. Therefore, there is no comparison of it.

### 4.3. Results Analysis

#### 4.3.1. Effects of the Algorithm Framework on the Indirected Network

First of all, inspect the joint effect of the inspiring factor c and the size of the target set k, namely the influence of different values of c on the range of influence with the same value of k. The results of the experiment on the data set 1 are shown in Figure 5. It is can be known from the figure that for different values of k, most of the ranges of influence with other values of c are bigger than the range of influence when the value of b is 1, except for the condition of b=0. When the value of b is 0.6 and the value of k is 60, the range of influence of the algorithm under the framework is about 10% higher than that of the greedy algorithm. When the value of c is 0, all the initial nodes are statically chosen from the nodes with the largest value of PI, which fails to consider the propagation process of influence, thus, its range of influence is the worst. In the following experiments, the condition of b = 0 is ignored.

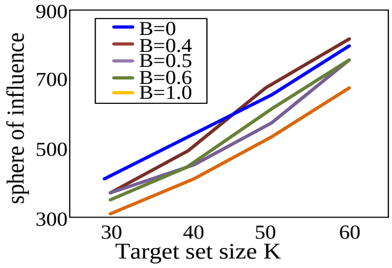



**Figure 5. The Influence Curves with Different Values of K and C on Data Set 1**

The results of the experiment on the data set 2 are shown in Figure 6. The effects of the vast majority of the same algorithm framework with the appropriate value of c are better than that of HIM algorithm, only the effects of the isolated cases are worse than that of HIM algorithm.

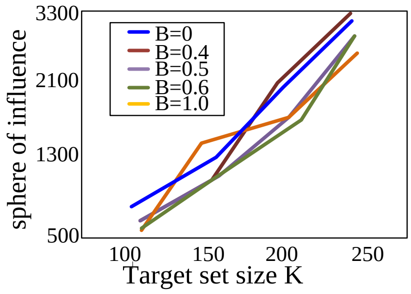



**Figure 6. The Influence Curves with Different Values of K and C on Data Set 2**

### 4.3.2. Effects of the Algorithm Framework on the Weighted Network

In order to verify the effectiveness of the framework on the weighted network, the same experiment is carried out on the data set of the author cooperation network for computation geometry. In the experiment, the value of buv is calculated by adopting the estimation formula in the Section A, and the value of $\theta_v$ is 0.5. The results are shown in Figure 7.

It can be seen from the figure that most of the ranges of influence of the algorithm under the algorithm framework are better than that of HIM algorithm, especially when the value of b is 0.6, and the improvement is about 12% under the best case. Figure 5, Figure 6 and Figure 7 all show that with the increasing of k, the slope of the influence curve of the algorithm under the algorithm framework becomes greater, which means that with the increasing of k, the improvement becomes more obvious. However, the current large social networks all contain millions of nodes, and it does not have practical significance if the value of k is too small. Therefore, the advantage of the proposed algorithm framework in the large social networks becomes more obvious, for the framework needs enough nodes to accumulate the potential influence.

 Copyright ⓒ 2016 SERSC

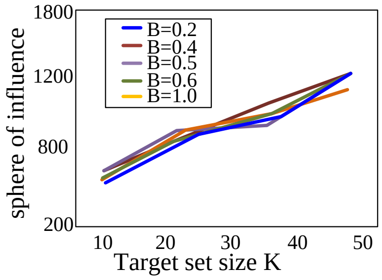



**Figure 7. The Influence Curves with Different Values of K and C on Data Set 3**

#### 4.3.3. Effects of the Algorithm Framework on the Directed Network

In order to verify the effectiveness of the framework on the directed network, the same experiment is carried out on the data set of Wiki-vote. In the experiment, the value of buv is calculated by adopting the estimation formula in the Section A, and the value of $\theta_V$ is 0.6. The results are shown in Figure 8.

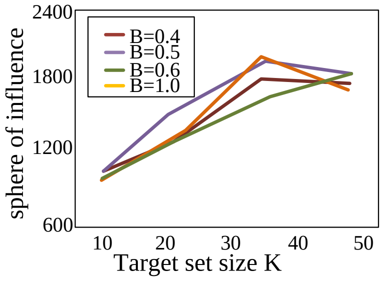



**Figure 8. The Influence Curves with Different Values of K and C on Data Set 4**

It can be seen from the figure that the experimental results are similar to the results on the indirected network and the weighted network, most of the ranges of influence of the algorithm under the algorithm framework are better than that of HIM algorithm, especially when the value of c is 0.6, which means that the algorithm framework also can obtain good effect on the directed network.

Then, for different types of data sets, how to determine the value of b, there cannot be experiment for each value of b to choose the best value. But it can be seen from the experimental results that most of the cases are superior to HIM algorithm when the value of b is 0.6. Therefore, the median value 0.6 is chosen as the parameter of the algorithm, when the value of c is 0.6, it is the nearest to the center whether the value of b is smaller or bigger. Thus, when there is b=0.6, the algorithm for influence maximization under the algorithm framework is determined as the HPG algorithm.

#### 4.3.4. Effects of the HPG Algorithm on the Signed Network

It has been confirmed that HPG algorithm is the algorithm for influence maximization, when the value of b is 0.6, the effects of the HPG algorithm and HIM algorithm on the signed network will be directly verified in the following. Two large signed social networks are selected to carry out the experiment, which is respectively Slashdot data set and Epinions data set. The experimental results are shown in Figure 9 and Figure 10.

Copyright ⓒ 2016 SERSC 

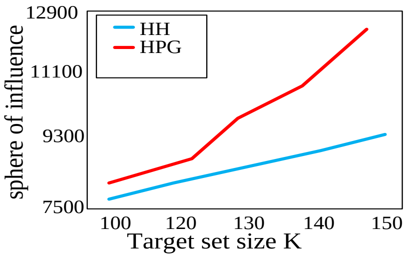



**Figure 9. The Influence Curves of HPG and HIM on Data Set 5**

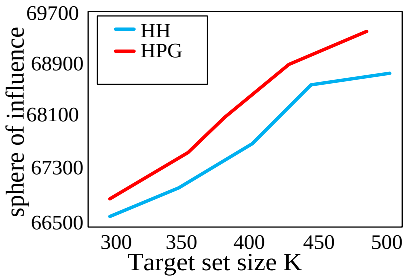



**Figure 10. The Influence Curves of HPG and HIM on Data Set 6**

It can be seen from the figure that relative to HIM algorithm, the range of influence of HPG algorithm also has a big improvement on the signed network, and it verifies the correctness of the value of c at the same time.

### 4.3.5. Minute Comparison between HIM Algorithm and the Algorithm Framework

The differences between the two stages of the algorithm framework will be inspected in the following, which can explain the reason why the range of influence of HPG algorithm is broader than that of HIM algorithm.

The average influence (AI) of the nodes in the two stages of the algorithm framework is listed in Table 3. In order to make it easy for comparison, the HIM algorithm is accordingly divided into two stages, the corresponding stage 1 and the corresponding stage 2. Here, AI is used as the comparison tool. For example, there are 10 nodes that are seleted in the greedy stage, and they activates 60 nodes, then the value of AI is 6.0 (60/10) in the greedy stage.

**Table 3. The Value of AI of Each Stage with K=50 on Data Set (1,2)**

| The average influence | The algorithm framework | | | | The greedy algorithm | | | |
|---|---|---|---|---|---|---|---|---|
| | The enlightening stage | | The greedy stage | | The corresponding stage 1 | | The corresponding stage 2 | |
| $b=0.2$ | 10.29 | 27.01 | 10.30 | 47.35 | 15.73 | 28.20 | 11.39 | 28.78 |
| $b=0.4$ | 10.3 | 19.80 | 10.5 | 47.26 | 13.52 | 24.02 | 10.87 | 30.61 |
| $b=0.6$ | 10.58 | 18.43 | 10.68 | 40.59 | 13.26 | 25.32 | 10.79 | 32.54. |
| $b=0.8$ | 10.62 | 18.67 | 12.01 | 36.83 | 12.49 | 23.98 | 10.12 | 32.31 |
| $b=1.0$ | 12.3 | 18.90 | 12.37 | 34.91 | 11.93 | 22.38 | 10.09 | 34.63 |

It can be seen from Table 3 and Table 4 that the average influence of the enlighting stage is lower than that of the greedy stage, while the average influence of the greedy stage is much higher than that of the corresponding stage 2. Therefore, with the appropriate parameter c and integrate the two stages, the effect of the algorithm under the proposed framework is better than that of HIM algorithm. It is easy to explain the reason

 Copyright ⓒ 2016 SERSC

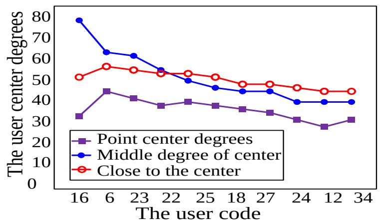
why the average influence of the enlightening stage is lower than that of the corresponding stage 1, because the nodes selected in the enlightening stage are statically selected and with the maximum value of PI, rather than the most influential nodes. The average influence of the greedy stage is much higher than that of the corresponding stage 2, and it is mainly inspired by the characteristics of the nodes selected in the enlightening stage and the "influence accumulation" property of the LT model, although they fails to activate many nodes, they widely add influences to the other nodes in network, and the affected but inactivated nodes is explosive in the greedy stage, this is why there are more nodes can be activated in the greedy stage.

### 4.3.6. Comparison of the Centricity of User

Closeness centrality is to measure the degree of an actor for not being controlled by others. In contrast to the intermediate centricity, the smaller the value of the closeness centrality is, the closer the point to the core position will be. Table 4 shows the closeness centrality of the user. As shown in Figure 11, select some representative users to carry out the analysis and comparison of the various closeness centralities of the user.

**Table 4. The Closeness Centrality of User**

| user | Near the center of penetration | The degree of closeness centrality |
|---|---|---|
| 2 user | 65.043 | 41.607 |
| 6 user | 59.093 | 54.897 |
| 7 user | 60.011 | 68.065 |
| 8 user | 60.012 | 64.987 |
| 14 user | 60.001 | 59.084 |




12900

**Figure 11. Comparison of the Centricity of User**

The experiments of the centricity analysis that combines with the specific node users show the influence of the individual nodes. The centricity analysis is the inspection of the intermediate centricity and closeness centrality of the target nodes, which shows the influence of the nodes from the interaction and the status of the nodes in the network. The closeness centrality of the 6 user node is only 59.093, and the node is located right in the middle of all the nodes, but its centrality is the biggest of all. The large intermediate centrality states that other users in the network are more likely to rely on the node, and the node has larger power in the network, and the influence is reflected in the dependencies of other users. The value of the closeness centrality of the 2 user node is the largest of all nodes, while its intermediate centrality is only 41.607, which shows that when the node actives other nodes, it will be more influenced by other nodes, namely that it will be more controlled, therefore, its independence is relatively weak in the network, and relative to the 14 user node, the influence of this kind of node is relatively weak. The intermediate centrality and the closeness centrality of the 8 user node and the 7 user node are all in the middle level, and this kind of user node not only has relatively larger power in the network, but also is not easy to be controlled by others at the same time, and has a strong independence, thus this kind of node often has greater influence in the network.

## 5. Conclusion

Since HIM algorithm is very costly, it cannot be applied to large social networks. Also, HIM algorithm could not guarantee the best influence spread. In this paper, based on the “influence accumulation” property of the LT model and by synthesizing the advantages of the degree centrality and the greedy algorithm, a framework of the algorithm for influence maximization is proposed, and a new hybrid algorithm for influence maximization is given. The experimental results show that compared with HIM algorithm, the algorithm proposed in this paper has a very good improvement in the aspect of the scope of influence, and it is better than HIM algorithm in the aspect of time complexity.

## Acknowledgement

The research is supported by the project “The on-time monitoring system for hazard in chemical industrial park based on Internet of Things”(14ZXCXGX00409) and “The College students' psychological consultation service platform based cloud computing”( 14ZXCXGX00531).

## References

[1] A. R. M. Teutle, “Twitter: Network Properties Analysis. Electronics, Communications and Computer (C0NIELEC0MP)20th International Conference on, Cholula, **(2010)**, pp. 80-186.
[2] M. Cha, H. Haddadi and F. Benevenuto, “Measuring User Influence in Twitter: The Mi 11 ion Follower Fallaey”, International AMI Conference on weblogs and Social Media (ICWSM), **(2010)**.
[3] K. A. Aaltonen, “Network perspective on response strategies to stakeholder pressures in global projects, In Miia Martinsuo edt Recipes for success in project-based management”, Project Management Association Finland, **(2008)**, pp. 29-59.
[4] K. Andrews and B. L. Delahaye, “Influences on knowledge processes in organizational learning The psychosocial filter”, Journal of Management Studies, no. 7, **(2000)**, pp. 797-810.
[5] B. Ramasamy, K. W. Goh, M. C. H. Yeung, “Is Guanxi (relationship) a bridge to knowledge transfer”, Journal of Business Research, vol. 59, **(2009)**, pp. 130-139.
[6] H. Cho, G. Gay and B. Davidson, “Social networks, communication styles, and learning performance in a CSCL community”, Computers & Education, vol. 49, no, 2, **(2007)**, pp. 309-329.
[7] S. Ferriani, G. Cattani and C. B. Fuller, “The relational antecedents of project-entrepreneurship: Network centrality, team composition and project performance”, Research Policy, vol. 38, no. 10, **(2009)**, pp. 1545-1558.
[8] H. X. Yang, B. H. Wang, W X. Wang and Z. H. Rong, “Spatial games based On pursuing the highest average payoff”, Chin. Phys. Lett., vol. 28, **(2008)**, pp. 3504.
[9] J. Bruneel and P. D'Este, “Ammon Salter. Investigating the factors that diminish the barriers to university-industry collaboration”, Research Policy, no. 39, **(2010)**, pp. 858-868.
[10] L. W. Gar, “New institutional policies for university-industry links in Japan”, Research Policy, no. 36, **(2007)**, pp. 1261-1274.
[11] M. J. Mirza and N. Anjum, “Association of Moving Objects Across Visual Sensor Networks”, Journal of Multimedia, vol. 7, no, 1, **(2012)**, pp. 2-8.
[12] B. Masquefa, “Top management adoption of a locally driven performance measurement and evaluation system: A social network perspective”, Management Accounting Research, vol. 49, no. 2, **(2008)**, pp. 182-207
[13] D. W. Pearson, P. Albert and B. Besombes, “Modeling enterprise networks; A master equation approach”, European Journal of Operational Research, no. 138, **(2012)**, pp. 663-670.
[14] J. Skvoretz, T. J. Fararo, F. Agnessens, “Advances in biased net theory: definitions, derivations, and estimations”, Social Networks, vol. 26, **(2010)**, pp. 113-139.
[15] G. Szulanski, “The process of knowledge transfer. A diachronic analysis of stickiness”, Organizational Behavior and Human Decision Processes, vol. 82, **(2000)**, pp. 9-27.
[16] T. Bjerregaard, “Industry and academia in convergence: Micro-institutional dimensions of R&D collaboration”, Technovation; no. 30, **(2011)**, pp. 100-108.
[17] B. Xianyu and J. M. Yang, “Evolutionary ultimatum game on complex networks under incomplete information”, Physica A, vol. 389, **(2011)**, pp. 1115-1123.

## Author

**Dou Ruofei**, received her M.E. degree in software engineering from Nankai University in Tianjin, China. She is currently a lecturer in the Tianjin University of Science & Technology. Her research interest is mainly in the area of Social Networks, Web Data Mining. She has published several research papers in scholarly journals in the above research areas and has participated in several books.

 Copyright ⓒ 2016 SERSC