International Journal of u- and e- Service, Science and Technology
Vol.8, No.4 (2015), pp.349-358
http://dx.doi.org/10.14257/ijunesst.2015.8.4.32

Detecting Storm Eyes of Advertisement Based on Sina Weibo

Jinhua Xiong[1,2], Fung Wong [1*], Shijun Lee[1] and Xing Yuan[1]

[1](*School of Computer Science，Wuhan University，Wuhan 430072，China)*
[2](*Information Technology Limited Company of Jiu tianyue, Wuhan, China)*
afengcom@163.com

Abstract

With the development of micro-blog advertising industry, micro-blog ads present a new perspective to people. It is observed that the communication effect would be maximized only when the ads spread on optimization path consisted of some crucial nodes. Consequently we define these critical nodes as storm eyes of ads. Referring to the theory of the shortest path and scale-free characteristics of complex networks, we build a model to detect the storm eyes of ads based on the elements closely related to micro-blog ads, combined with the thought of dividing and conquering-merge. A detecting algorithm is proposed based on the model, the time complexity of which is just O (n). The dataset of our experiment obtained from SinaWeibo API has reached the number of 1 billion. And the results show that detection time is much less than artificial statistical approach on the premise of the algorithm with high detection accuracy.

Keywords: *Sina Weibo, advertisements communication flow, storm eyes of ads*

1. Introduction

After Sina micro-blog came into being, the advertisement industry has been perfectly combined with Sina micro-blog, thus giving rise to an emerging industry, the micro-blog advertising industry. This paper focuses on solving the following question: which key customers should the micro-blog ads make use of to be spread in the largest scope, through the shortest communication time and with the maximum communication effect. Due to the low micro-blog ads launch accuracy and the difficulty of real-time update in the current stage, the launching effect of micro-blog ads is far from being satisfying. In order to solve these problems, many scholars both home and abroad have conducted relevant technical researches based on Twitter and Sina micro-blog. David Kempe *et al.* pointed out in their paper that the optimization issue of choosing the most influential nodes is an NP problem [1]. Wei Chen *et al.* put forward two ideas to improve the influence maximization in the social network [2]. Yao Zhang *et al.* adopted a simple but effective evaluation method as the influence function, and put forward greedy algorithm of greedy evaluation expectation [3]. Wei Chen *et al.* put forth a heuristic algorithm, which can easily measure the million-level nodes and sides. Moreover, the algorithm boasts the balance adjustment parameters to control the user operation time and influence communication of the algorithm [4]. Cheng-Te Li *et al.* put forward in their paper issue of influence communication maximization of the diversified social network [5]. However, Kyomin J. *et al.* thought that the influence maximization in the social network refers to the maximization of the influence coverage of these nodes by choosing the number of "k" top nodes [6]. Bo Liu *et al.* put forward the issue of influence maximization with time limit [7].

The structure of this paper is as follows: The first part gives a brief introduction of some new concepts related to micro-blog ads; the second part puts forward a key launching node, or the detecting model of the storm eyes of the ads, to detect the ad communication flow; the third part provides the algorithm related to the detecting model and applies it to support the detecting model; the fourth part proves the feasibility and effectiveness of the detecting mod-

ISSN: 2005-4246 IJUNESST

el, conducts experiments of the detecting model and algorithm and shows the experimental results; the last part is a summary of the whole paper.

2. Ads communication

The ad communication network based on Sina micro-blog puts its advertising on the communication flow formed by the advertising nodes. After several node transmissions, the maximum communication scope (referring to the condition that the ad communication flow covers the most communication nodes) is achieved through the shortest communication path (or the shortest communication time). In this way, the maximum ad influence (ad communication effect) and advertising value can be realized. The research objective of this paper is to find the users for the advertisement launching so as enable the ads to achieve the maximum communication effect through the fewest transmission times. All these users are the storm eyes mentioned in this paper, and they are the central and key element that gives rise to the ad storm. The exposition and experiment in this paper is based on the assumption that communication time between the nodes is the same. Based on the assumption, this paper uses the fewest node transmission times to represent the shortest communication time of the ad communication flow, and uses the maximum covered nodes under the condition of fewest transmission times to represent the maximum communication effect within the shortest communication time. The following part will give a brief introduction of the concept of the micro-blog ad communication flow and the storm eyes of the micro-blog ads. Ad communication network is defined as "G"; the node collection in the network as "V"; and side collection as "W." Then the ad communication network constitutes the following relationship:

$$G(.)=(V,W),(1)$$

In which i stands for node item; n stands for the total number of nodes; m stands for the total number of sides; Vi stands for the node whose node item is i; W_j stands for the side whose node item is j; G(.) stands for the ad communication network made of n nodes and m sides. Diagram 1 is a typical ad communication network. The following part will give the definition of three major concepts in the communication network of the micro-blog ads, namely ad communication flow, storm eyes of the ads and step length.

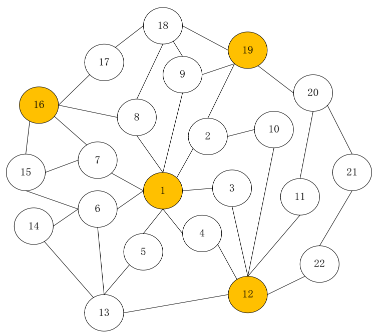

Diagram. 1 Typical Ad Communication Network

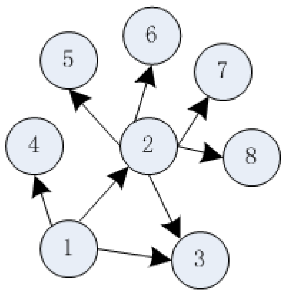

Diagram. 2 Schematic Diagram of Micro-blog Ad Communication Flow

Definition 1 Ad Communication Flow: it is an ads communication path. The source of ad communication is called the ad communication source, which originally comes from the publisher of the ads. The ad communication flow is shown in Diagram 2.

In diagram 2, nodes stand for the registered users of Sina micro-blog. If there exist sides between the nodes, it suggests there is ad communication flow between them. The arrow direction stands for the communication direction of the ad communication flow. For example, 1->2->5 in the diagram stands for the ad communication flow from 1 to 2 and then to 5.

Definition 2 Storm Eyes of the Ads: in the communication path formed by the nodes on the micro-blog ad communication flow, if there is a key path, which can enable the ad communication flow to arrive at the largest communication scope within the shortest time, the nodes on the key path are called as the center nodes, or the storm eyes of the ads. See Diagram 3.

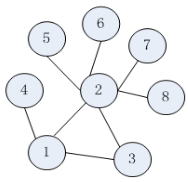

Diagram 3. Schematic Diagram of the Storm Eyes of Micro-blog Ads

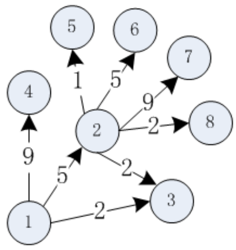

Diagram. 4 Model Schematic Diagram of Weighted Communication Effect Index

It can be seen from Diagram 3 that node 2 connects the other six nodes in total. If the ad is launched in node 2, the ad information can be transmitted to the other six nodes at most at once. Since the total number of the nodes in the network is 8, node 2 only needs to transmit the ad information once to achieve 75% of the ad communication scope, which is larger than any random node in the network. In this way, node 2 can be defined as the center node, or the storm eye of the ad.

Definition 3 Step Length: The distance between nodes on the ad communication flow starting from the node of the ad communication source. The step length will be increased by 1 for each extra communication of one node. In this paper, the step length is represented by “s.” Then, the step length between two nodes, namely I and j, can be represented by s (i, j). Just as Diagram 3 shows, 1——>2——>5 is an ad communication flow. The step length between node 1 and node 2 is s (1, 2)=1; the step length between node 2 and node 5 is s (2, 5)=1; and the step length between 1 and 5 is s (1, 5)=2.Based on Diagram 3, the micro-blog ads communication effectindex η is further defined as follows:

$$\eta = \frac{\mathrm{F}(i)}{n}(i = 1 \ldots n) \quad ,(2)$$

in which, i stands for node item, n stands for the total number of nodes, F (i) stands for the fans’ number of node I, and the micro-blog ads communication effect index η is used to measure the communication influence dimension of Sina micro-blog ads.

In Diagram 4, the numbers 1-8 with circles stand for the eight communication nodes in ad communication network. The arrows stand for the communication direction of the ad communication path. The numbers on the connecting lines with the arrows stand for weight, which is the fans’ number of the node that the arrow points at. The algorithm model of the micro-blog ads communication effect index is worked out based on the practical case provided in Diagram 3. The communication effect index of every communication path worked out according to formula 2 is as follows: (1)The communication effect index of node 1 is: （9+5+2）/n=16/n; (2)The communication effect index of node 2 is: 5/ n;(3)The communication effect index of node 3 is: 2/ n;(4)The communication effect index of node 4 is: 9/ n; (5)The communication effect index of node 5 is: 1/ n; (6)The communication effect index of node 6 is: 5/ n; (7)The communication effect index of node 1 is: 9/ n;(8)The communication effect index of node 8 is: 2/ n. Based on the above result, it is known the center node of the network in the area is node1. Furthermore, the concept of the communication effect index of the ad communication flow can be extended to overall communication effect index and partial communication effect index. Therefore, the two concepts can be defined as follows: when the whole Sina micro-blog is regarded as the ad communication network, the calculation formula of the overall communication effect index of a certain node σ in the network is $\eta(\sigma)=\frac{F_{out}(\sigma)}{N(\cdot)}$, which means the ratio of the number of degrees of node σ to the total number of nodes in the communication network. When a certain influence area on the Sina micro-blog is regarded as the ad communication network, the calculation formula of the partial communication effect index of node σ in the network is $\eta(\sigma)=\frac{F_{out}(\sigma)}{N(\cdot)}$, which means the ratio of the total number of nodes in the communication network to the output number of degrees of node σ.

Definition 4 Detecting Precision: The connecting number of the storm eyes of the ads and their output number of degrees or input number of degrees. The corresponding practical quantized calculation formula is:

$$p = \frac{\sum_{i=1}^{n} Bf_i}{\sum_{i=1}^{n} Fo_i} \quad (3)$$

in which i stands for the item of the storm eyes of the ads, n stands for the total number of the storm eyes, Bf_i stands for the shared fans' number of the storm eyes of the ads, Fo_i stands for the number of fans, and the detecting precision P is used to measure the detecting precision of the storm eyes of the ads.

In the following part, this paper will focus on researching into the detecting model of the storm eyes of the Sina micro-blog ads.

3. Detecting Models

Due to the complex network characteristics of the ad communication flow network on Sina micro-blog, this paper thus puts forward an assumption about the rule of the shortest path in the ad communication flow network:

Assumption 1 In Sina micro-blog ad communication network with the complex network characteristics, most of the nodes are connected through the center node, and there is a comparatively large probability that the shortest path between any nodes passes through the center node. The center node here refers to micro-blog user with highly converged fans' number. The area with several center nodes as the core can divide the ad communication network on the Sina micro-blog into several partial areas. As long as the strategies are adopted to find the center node in the area, the center node of each partial area can be found out. If all these nodes become the storm eyes of each area, all the storm eyes of the areas can be connected. Then it is possible for us to find a similar virtual key path, or the shortest path, so as to maximize the ad communication effect. In the complex network, if a node in a certain area is the center node, then most of the nodes within the area will pass the center node. Similarly, in the ad communication network, most of the nodes will pass through the center node, thus suggesting that the center node is definitely connected with a majority of the other nodes, which directly shows the large exit number of degrees of the center node. Then, the center node collection is defined as C. On the original state, it is shown as follows:

$$c = \begin{cases} v_i, i = 1 \ldots n \\ \varphi \end{cases} \quad (4)$$

According to the node collection V defined in formula 1 and combining with the scale-free characteristics and the thought of dividing and conquering-merge, the collection of the nodes can be divided into several sub-collections, A0,A1,…,Al, then

$$A = \bigcup_{i=0}^{k} A_l = G(\cdot) \quad (5)$$

In which l stands for the collection item, k stands for the total number of the sub-collections, and Al stands for the sub-collections, whose item is 1. Formula 1 and f suggests that the ad communication network made up of node collection and side collection can be divided into the number of k sub-networks. Then, in the above sub-networks composed of the above node sub-collections, the maximum value of node fans' number contained in each sub-collection is worked out in turn. $Max\{F(V_i|A_k)\}$(l=0…k),(i=1…n), in which I stands for the node item, n stands for the total number of the nodes in the sub-collection, Vi stands for the nodes, whose node item is i, 1 stands for the collection item, A_k stands for the node sub-collection, whose collection item is k, and then $F(V_i|A_k)$ stands for the fans' number of the node, whose item is I in the k sub-collection and $Max\{F(V_i|A_k)\}$ stands for the maximum value of the fans' number of the node. In particular, the maximum value of the fans' number of each node in the sub-collection A0 can be represented as $Max\{F(Vi|A_0)\}$, (i=1…n), in which i stands for the node item of each node in the sub-collection A0 and n stands for the total number of the nodes in the sub-collection A_0.

In the following part, the value of each node from the source node is compared. The steps are as follows: Initialize the node item as I; Compare the value of F(Vi) and F(Vi+1) and the ads communication effect Y; If F(Vi)>F(Vi+1), put the node Vi into the center node collection C, and move down the source node as Vi+1, in which the node item is added with a, and then repeat Step 1; When F(Vi)≤F(Vi+1), the node Vi flows toVi+1 and the node item is added by 1, thus forming the ad communication flow, then repeats Step 1; When the node item is i=n, the cycle ends;

1. The center node collection of the ad communication flow is acquired as $C = \bigcup_{j=0}^{m} A_j$ in which j stands for the center node item, m stands for the total number of the nodes in the center node collection, A_j stands for the center node whose node item is j and C is the collection of the center nodes. After the collection of the center nodes is gained, a virtual shortest path can be constituted by connecting all the acquired center nodes, and the path is definitely quite close to the practical shortest path. Then node 1 and node n on the path is exactly the storm eyes of the ads. After extending the algorithm method of the micro-blog ad index in formula 2, the area node communication effect algorithm in the micro-blog ad network is shown in formula 6:

$$\eta(i) = S \times P \quad (7)$$

in which, matrix S is a1×n matrix, $S=(\eta_1, \eta_2, \dots, \eta_n)$, matrix P is a n×1 matrix, and $\eta_i = \frac{F(i)-B_i(i)}{F(i-1)-B_i(i-1)}$ (i=0, 1…, n). In particular, η_0 stands for the ad effect index of the source node. Therefore, we make an assumption here η_0=0.

In the above formula 7, η stands for the ads communication effect index in the Sina micro-blog add communication network and stands for the node dimension. When the source node is regarded as the root node of the tree, the node dimension can be regarded as the depth of the tree. In other words, when i=0 represents the source node, i=1 stands for 1°sub-node, which is 1°fans' nodes; i=2 stands for 2°sub-node, which is the 2° fans' node of the source node, and the follow-up sub-node can be deduced likewise. η_i stands for the ad effect index of i° sub-node of the source node, which is the ads communication effect index of the i° sub-node of the source node. η(i) stands for the ad effect of the i° fans of the source node. F(i) stands for the fans' number of the i° fans, and $Bi(i)$ stands for the shared fans' number of the i° fans of the source node. In fact, it is meaningless to launch the ads in the acquired communication flow node of the ad information, because the ad information quantity will be zero to the node if it receives the ad information the same to the acquired one. In this way, the value of the ad communication will be lost. Therefore, while establishing the detecting mode, it is necessary to avoid closed loop circuit on the weighted directed graph (Diagram 4) and closed rings on the nodes as much as possible. When the idea is applied to the Sino micro-blog, it means that the shared fans' relationship should avoided in the users' ad communication flow. Though, the whole social network of Sina micro-blog is in itself a closed loop, the number of the ads are too numerous to count. Therefore, it is assumed that the communication flow an ad cannot be spread in the whole micro-blog, but each user knows the ad information.

In the following part, this paper will introduce the algorithms related to the detecting model.

4. Algorithm Description

This part will introduce the algorithms related to the detecting model of the storm eyes on the Sina micro-blog mentioned in Part 4.

Detecting the eyes of Advertising Storm Algorithm
1、Input：NODE-List V and Edge Set W ；
2、Output：Set C of Center Node and Advertising communication effect of every node in Set C;
3、Initialize：NODE-List V, Edge Set W, Node-ID i, Node-Count n, Source-Node V_0, Fans-Count of every node;
4、 $V_0=V_i$**;**
5、Compare $F(V_i)$ **,** $F(V_{i+1})$ **and Y of every node;**
6、if $F(V_i)>F(V_{i+1})$**, then**
insert V_i **to Set C of Center Node and Source node** $V_0=V_{i+1}$**,**
i++ ， jump to step 4;
7、if $F(V_i)\le F(V_{i+1})$**, then**
8、 $V_0=V_{i+1}$**, i++ ， jump to step 4;**
9、if i=n then
10、end;
11、return C=$\bigcup_{j=0}^{m} A_j$**, m is Node-Count of Center Node**

From formula 1, it can be seen that Step 1 to Step 4 is the preparation work of the algorithm, including the initialized point collection V, side collection W, node item i, the number of the nodes n, source node V_0 and the fans' number of each node. From Step 5, the iterative process of the algorithm begins, which ends until in Step 10. Step 11 returns to the center node in the ad network, which forms the key path, and the time complexity of the algorithm is only O(n). Through the above algorithm of detecting the storm eyes of the ad, the center nodes in the ad network can be gained, or the storm eyes of the ads. Through the calculation results, we can gain the approximate key path so as to work out the maximum approximate value of the ad communication effect.

In the following part, this paper will discuss the experiment process so as to prove the above model and algorithm.

5. Result and Analysis of the Simulation Experiments

Before introducing the experiment, this paper will firstly describe the experiment data collection. After that, this paper will show the experimental result so as to prove the effectiveness of the experiment model and algorithm.

5.1. Data Collection and Experiment Preparation

The data collection employed by the experiment is the sub-collection of the originally collected data. All the data in the originally collected data are collected through Sina microblog's API before December 2012, whose total number is more than 2 billion. This paper adopts the data in November and December of 2012 as the experiment data. The experiment is conducted on the server equipped with the operation system of Linux operation platform. The server's configuration is an 8-eight CPU with 32G internal storage. After the data are acquired, programming and simulation experiment are conducted on Matlab. The specific steps of the simulation experiment are shown as follows: firstly, repeatedly select all these data gathered through the Sina micro-blog's API; then import them into Matlab and make them into the simulation data collection required by the experiment; later, realize the algorithm models and algorithms mentioned in the previous part in the programming of Matlab; finally, draw the result of the simulation experiment.

5.2. Analysis of the Experimental Result

The experiment draws randomly 1 billion pieces of data from the data collection, and randomly divides them into ten groups, each of which contains 100 million pieces of data. After several iterative comparison and combination through the above mentioned models and algorithms, we gain ten center nodes with the maximum ad communication effect. Then, the statistical method is adopted for comparison and combination to find the genuine center node among the 1 billion pieces of data. Due to the limited article length, this following part just lists the ads communication effect result when the number of the center nodes is 10. Then, the two methods are compared and go through the normalization processing. After that, the acquired experimental result is shown in Chart 5.

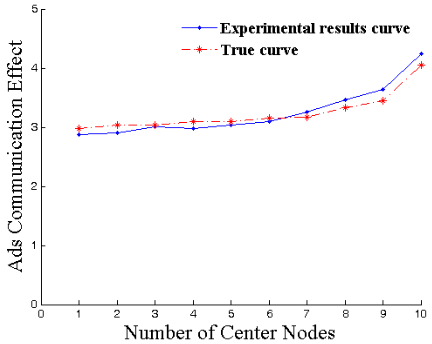

Chart 5. Ads Communication Effect

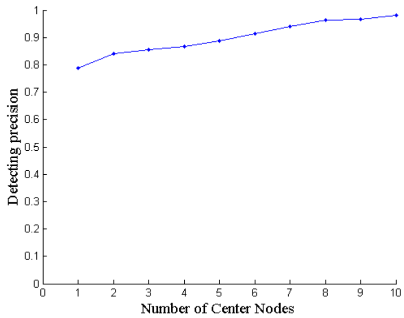

Chart 6. Storm Eyes vs. Detecting Precision

In Chart 5, Number of Center Nodes is the x-coordinate, and Ads Communication Effect is the y-coordinate, standing for the ads communication effect passing through the ads communication path formed by each center node. The red curve stands for the maximum value of the ten nodes' ads communication effect worked out through the statistic method after a comparatively long period of time. The blue curve stands for the center nodes gained in the experiment employing the models and algorithms mentioned in the above. This paper divides the ads communication effect into 0-5 levels, which is the value of the ads communication effect within the scope of 0-5. When the ad is launched on one center node, its ads communication effect is 2.8772. When the number of the center nodes is 2, the ads communication effect will be increased to 2.9063. The number of the center nodes is continuously increased to 10, and then its ads communication effect reaches 4.2468. Table 1 shows the ads communication effect value through manual statistical ad launching method and the experiment method in this paper.

Table 1. Comparison of the Statistical Method and Experiment Algorithm

Number of center nodes	Experiment method	Statistical method
1	2.8772	2.9810
2	2.9063	3.0315
3	3.0063	3.0387
4	2.9810	3.0950

5	3.0387	3.0971
6	3.0950	3.1580
7	3.2580	3.1647
8	3.4647	3.3351
9	3.6412	3.4412
10	4.2468	4.0456

The ideal effect of ads communication is to make all micro-bloggers learn the existence of an ad, that is, the ads communication effect value reaches 5. In fact, due to the offline status of the micro-bloggers and the existence of factors, like weak ties between the micro-bloggers and the data sparity [13], ads communication effect will just get infinitely close to the ideal value. Therefore, in Table 2, the ads communication effect percentage is given to represent the gap between the ideal value, and experiment value and genuine value respectively. Assuming that the ads communication effect percentage of the ideal value is 100%, their gap can be compared through the ratio of the ideal value to the experiment value and to the genuine value respectively.

Table 2. Comparison of the Ads Communication Effect and Ideal Value

	Ads communication effect percentage		
Number of the center nodes	Experiment vale	Statistical value	Ideal value
1	57.544%	59.620%	
2	58.126%	60.630%	
3	60.126%	60.774%	
4	59.620%	61.900%	
5	60.774%	61.942%	100%
6	61.900%	63.160%	
7	65.160%	63.294%	
8	69.294%	66.702%	
9	72.824%	68.824%	
10	84.936%	80.912%	

It can be seen from Table 2 that, when the number of the center nodes reaches 10, the ads communication effect percentage already reaches 84.936%, while that gained through the statistical method is just 80.921%. Compared with the ideal value, the ads communication effect gained through the experiment method is comparatively high, which is higher than that gained through the statistical method by 4.024%.

In Chart 6, the Number of Center Nodes is the x-coordinate; and the Detecting Precision is the y-coordinate. Based on formula 3, it can be known that, as long as the degree of the center node is the largest in the area scope divided by the sub-collection, it can be ensured that it is the center node in the area, thus guaranteeing the detecting precision. It can be seen from Chart 6 that, with the increase of the number of the storm eyes of the ads, the detecting precision is higher and higher. Thus, it is suggested that, due to the similar properties of the storm eyes, as long as one storm eye is found, more storm eyes can be found according to the similar properties, thus achieving the purpose of improving the detecting precision.

The experimental result shows that a shortest path composed of center node sequence can be found through the models and algorithms mentioned in the paper. At the same time, the models and algorithms mentioned in the paper can help us gain the approximate key communication path in the ads communication network within a short time period, and can detect the center node with comparatively high precision. As long as the center nodes are precisely detected, the ad can be launched through the shortest path composed of the central node sequence to achieve the goal of obtaining the maximum ads communication effect through a small number of center nodes.

6. Conclusions

Through analyzing the rules of Sina micro-blog ads launching and communication, this paper puts forward the new concepts of micro-blog ads communication effect and their communication effect index and the storm eyes of the ads, and the model of detecting the storm eyes of the ads on Sina micro-blog through combining the complex network's idea of acquiring the shortest communication path. Based on that, the storm eyes of the ads are abstracted into the center nodes in the complex network. By employing the idea of detecting the center nodes and combining the experiment method, this paper preliminarily solves the problem of detecting the center nodes and approximate communication path for the advertising. The main contributions of the article can be boiled down as follows: (1) define the storm eyes of the ads, ads communication effect and their communication effect index; (2) put forward the detecting model of the storm eyes of the ads based on the Sina micro-blog platform; (3) prove through the experiment that the storm eyes detecting model of the ads based on the Sina micro-blog platform can better solve the problem of ads communication effect maximization and can largely promote the benefits of the advertising. The time complexity of the detecting algorithm is only O (n), so it can largely improve the time to detect the advertising nodes. At the same time, the experimental result proves that the algorithm boasts higher detection precision, thus suggesting the feasibility and effectiveness of the detecting model. Besides, the exposition of the paper provides a novel thinking for the ad launching benefits and increase of the ad return benefits on the Sina micro-blog for the further research work of the researchers in the future.

Acknowledgements

The authors thank National Natural Science Foundation of China (61272109) and the Ph.D. Student Independent Scientific Research Project of Wuhan University (2011102110060).

References

[1] D. Kempe, J. Kleinberg and E. Tardos, "Maximizing the Spread of Influence through a Social Network", SIGKDD 2003, pp. 137-146.
[2] W. Chen, Y. Wang and S. Yang, "Efficient Influence Maximization in Social Networks, KDD'09, (**2009**), June 28-July 1, pp. 199-207.
[3] Y. Zhang, "Estimate on Expectation for Influence Maximization in Social Networks", PAKDD 2010, (**2010**), pp. 99-106.
[4] C. Wei, W. Chi and W. Yajun, "Scalable Influence Maximization for Prevalent Viral Marketing in Large-Scale Social Networks", KDD'10, (**2010**) July 25-28, pp. 1029-1038.
[5] L. Cheng-Te, L. Shou-De and S. Man-Kwan, "Influence Propagation and Maximization for Heterogeneous Social Networks", WWW 2012, (**2012**), April 16-20, pp. 559-560.
[6] J. Kyomin, H. Wooram and W. Chen, "IRIE: Scalable and Robust Influence Maximization in Social Networks", ICDM (**2012**), pp. 918-923.
[7] B. Liu, "Time Constrained Influence Maximization in Social Networks", ICDM (**2012**), pp. 439-448.

Authors

Jinhua Xiong, he was born in 1970. He is a Ph.D. candidate at School of Computer, Wuhan University. His research interests include big data and intelligent water meter.

Fung Wong, he was born in 1984. He is a Ph.D. candidate at School of Computer, Wuhan University, and the student member of CCF. His research interests include social network and social media.

Shijun Lee, he was born in 1964. He is a professor at School of Computer, Wuhan University, and the senior member of CCF. His research interests include Web search and mining, the management and optimization of mass data, data mining, database technology and Android mobile development, etc.

Xing Yuan, she was born in 1992. She is an undergraduate at School of Computer, Wuhan University. Her research interests include micro-blog research and user behavior mining.