

# Research Methods on Recognizing Potential Opinion Leaders of Hot Event Based on Social Network

D.C. Qu, H.X. Lian, H. Xie

Computer Science of Beijing Institute of Technology  
Beijing, China

S.L. Guo

Automation School of Beijing Institute of Technology  
Beijing, China

J. Li

Howard University, America

**Abstract-**Recognizing the "opinion leaders" in advance is a new requirement in the context of social network. "From back to front" approaches do not conform to the transmission and evolution of development process of hot events. To address this problem, this paper proposes the concept of "potential opinion leaders", who is a user will become opinion leaders before the common event becoming hot. Our approach intuitively starts from the real development process of hot events, takes advantage of the "from front to back" dynamic research mode according to the propagation characteristics of hot events, and uses fan's number and identity tags as the screening criteria. We then design the analysis algorithm to recognize the "potential opinion leaders". The experimental result is compared with the traditional opinion leaders. It demonstrates that the recognition method by using "potential opinion leaders" is simpler, and guiding significance for network early warning is stronger.

**Keywords-**hot events; potential opinion leaders; number of fans; identity tags

## I. INTRODUCTION

Social network is emerging as one of the major channels to propagate information and share opinion about events, topics, etc. According to the characteristics of social network, the opinion of minority groups has affected the majority drastically in the information propagation process in the social networks such as microblog. The minority group who have greater impact is called the "opinion leaders", which was proposed for the first time in the field of social networks by communication scholar Paul F. Lazarsfeld[1] in the 1940s. Research has found that, as more and more people participating in social network activities, the importance of opinion leaders shouldn't be ignored any more[2]. In the online social network, variety of researches like the new recognition methods or recognition models were already proposed. These approaches are very meaningful, and they can help us recognize opinion leaders in hot events, but most of them take advantage of "from back to front" mode which means they are doing the research after the ordinary events have already become hot events. The "from back to front" mode doesn't match the generation, transmission, and evolution of the dynamic development process of hot events. State-of-the-art approaches are short of foresight and cannot fully satisfy the requirements of the network early warning. These approaches, which based on the fashioned hot events,

are in-effective on detection of emergencies, and are significantly difficult in identifying opinion leaders. To tackle these problems, we propose the concept of "potential opinion leaders". Our proposed approach detects opinion leaders in the mode of "from front to back", which conforms the actual pattern of hot events propagation. By monitoring the "potential opinion leaders", new event which may become a hot topic can be predicted and warned in advance. In this paper, we propose the recognition metrics and design recognition algorithm to recognize the "potential opinion leaders". We evaluate our approach based on real-data extracted from the social network. Compared with the researches of traditional opinion leaders, "potential opinion leaders" method is easier and has an important significance for network information regulation and emergency management.

The rest of this paper is organized as follows. We state the problem of "potential opinion leaders" in Section II. We then elaborate the research method in Section III and present the experimental verification in Section IV. We briefly review related works on the research of opinion leaders in Section V. Section VI give the summary of this paper and outline the future work.

## II. PROBLEM STATEMENT

Network events will experience five stages [3] during its evolutionary process of becoming hot events. During the propagation, the participation of "potential opinion leaders" makes ordinary events attract more attention, so it is not difficult to find that the active time of "potential opinion leaders" is a period between the precursor stage and the outbreak stage. The participation of "potential opinion leaders" makes ordinary events span from precursor stage to outbreak stage quickly. The theory model of "potential opinion leaders" is shown in Fig. 1. The propagation of hot events is similarity to the mode of innovation diffusion. Fig. 2 shows the variation trend of the S curve in the diffusion of innovation, the curve reflects the basic law of diffusion for the innovative things. Diffusion of innovation is always slow at first, and then suddenly speeds up when adopters reaching a certain number, and diffusion velocity slows down gradually, the trajectory of the number of innovators who have adopt innovation changing over time presents like an s-shaped. The participation of

"potential opinion leaders" is a direct cause to make S curve slope become steeper suddenly in the early stage.

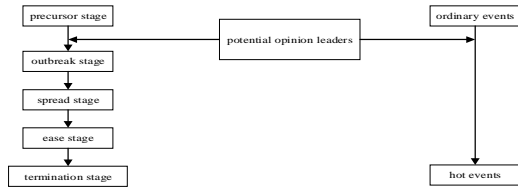


FIGURE 1. THE CONTRAST BETWEEN "POTENTIAL OPINION LEADERS" AND HOT EMERGENCY INFORMATION DISSEMINATION FIVE STAGES MODEL.

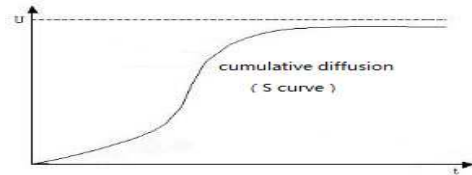


FIGURE 2. S CURVE OF DIFFUSION OF INNOVATION.

### 1 "POTENTIAL OPINION LEADERS" RECOGNITION METHOD

The relationships among social network users could be modeled as a graph. Each user is represented as a node in graph. In general, the importance of a node related to the number of adjacent nodes of the node. The greater the degree of a node, the node becomes more important. If there is "mutual follow" relationship between two users, then there is a path between these two nodes, and these nodes and paths together constitute a network graph of mutual follow relationship. "Potential opinion leaders" is the important node which has lots of adjacent nodes on a topic.

We propose two metrics to identify "potential opinion leaders":

- i. Fans number;
- ii. Identity tags of users.

Firstly, parameter  $I$  represents the fans number of social network users, it also represents the degree in graph we have formed. If  $F_y$  follows  $F_x$ , there is a connection between  $F_y$  and  $F_x$ , and from  $F_y$  to  $F_x$ ,  $I_x = 1$ ; If  $F_z$  follows  $F_x$  at the same time, there is a connection between  $F_z$  and  $F_x$ , and  $F_z$  to  $F_x$ ,  $I_x = 2$ . Secondly, parameter  $D$  represents the weight of user identity tags, which is the keyword such as "official corruption", this shows what kind of events in the social network community will attract their attention. "Potential opinion leaders"  $P$  can be expressed by parameter  $I$  and parameter  $D$  as follows:

$$P = I \cup D \quad (3.1)$$

" $\cup$ " represents  $I$  and  $D$  are a kind of mutual follow relationship, it shows that when the probability of some user becoming opinion leaders is large, which means the value of  $I$  is large, this user will become different kind of opinion leaders based on the difference of  $D$ . Equation 3.1 indicates that "potential opinion leaders" are the users who have accumulated a certain amount of  $I$  while follows  $D$  category events highly and participate in the discussion in the initial

period of event development, then rapidly become the opinion leaders of hot events.  $D$  is used as the weight of user identity tag keyword to distinguish the level of hot events, as shown in Table 1, the larger the value of  $D$ , the greater weight of attention the event has, and the event should be monitored strictly.

TABLE 1. USERS IDENTITY TAGS KEYWORD AND HOT EVENTS LEVEL PARTITION COMPARISON.

Keyword	Weight value
International events (reunification of the motherland, etc.)	0.8-1.0
Domestic events (political, economic, environment etc.)	0.6-0.8
Society and people's livelihood, education, medical, etc.)	0.4-0.6
Entertainment moral (entertainment news, etc.)	0.2-0.4
Others	0-0.2

The dissemination process of sudden hot events information is divided into the following five stages [3]: precursor stage, outbreak stage, spread stage, ease stage and termination stage. When recognizing "potential opinion leaders", we pay more attention to the users who play a promoting role at the beginning of hot event propagation and evolution. "Potential opinion leaders" is the key factor that push ordinary event from precursor stage to outbreak stage.

We use  $I$  to represent the fans number. The intuitionistic manifestation of the change of parameter  $I$  is the change of the node size: the larger the value of parameter  $I$ , the greater the size of node, the larger probability of the user to become "potential opinion leaders". Setting  $m$  as the threshold to represent the fans number of the users, and simplify data via increasing the value of  $m$  gradually and deleting the user whose fans less than  $m$ . In the end, we find these users' real identity, and observe their identity tags to determine what kind of events they will follow. For example, if a user's identity tags indicate he focuses on agriculture, rural areas and farmers, namely the value of  $D$  is "agriculture, rural areas and farmers", so the agriculture, rural areas and farmers events will attract his attention especially, once he participates in the topic discussion, he will drive lots of his fans and enable it to be hot. According to formula 3.1, the user is the "potential opinion leaders" of agriculture, rural areas and farmers. According to Table 1, the event belongs to the society and people's livelihood, so its weight is between 0.4 and 0.6.

The algorithm of identifying "potential opinion leaders" is described as follows:

---

**Algorithm 1** "potential opinion leaders" recognition method:

---

**Begin**  
 Extract the mutual follow data from the social network  
**for**  $i=0$  to 521  
   count  $I$   
   recognize "potential opinion leaders" according to  $I$  and  $D$   
   find the real identity according to  $ID$   
**output** the weight of user identity tags  
**end**

---





### III. EXPERIMENTAL EVALUATION

#### A. Fans' number of "potential opinion leaders"

We use a real data set extracted from Sina microblog to evaluate our approach. For the data set, we extract mutual follow relationship data of 521 social network users from 1st to 20th of February 2013, including the users ID information. The Figure in Original Data column of Table 2 is the network diagram of 521 social network users mutual follow relationship data, we take each social network user as a network node, and edge represents the follow relationship between the two nodes.

The openness of Social network makes any internet users can become social network users once registered. In consideration of some users have only a handful of fans, we need do data preprocessing. We set parameters  $m$  to represent the fans threshold number of social network users and set  $m$  different value to represent the minimum number of fans, and the results are shown in the following table 2.

TABLE II. THREE GROUPS OF EXPERIMENT RESULTS BASED ON DIFFERENT VALUE OF  $M$ .

parameters	primary data	experiment 1	experiment 2	experiment 3
value of $m$	$m=0$	$m=500$	$m=1000$	$m=1500$
nodes number	135864	114228	78045	43331
edges number	286456	221393	128341	60747
mutual follow network graph				

With experiment 1, we can see that setting the value of  $m$  as 500 has nothing to do with the recognition of "potential opinion leaders" when there are thousands of nodes and edges. That is to say, if the fans number is less than 500, then he is less possible to become "potential opinion leaders". Compare experiment 2 with experiment 1, the fans number is less than 1000 still have a tiny influence. It is observed by experiment 3 that when the value of  $m$  is adjusted for 1500, network diagram shows more apparent structure features, and user nodes become clearer. These special nodes are a part of the 521 social network users, the similarity of them is the fans number is very large. That is to say, if the value of  $m$  is large, the users are important. After statistics, these special nodes are 34 totally, and is shown in Fig. 3.



FIGURE 3. MUTUAL FOLLOW NETWORK GRAPH OF 34 SPECIAL SOCIAL NETWORK USERS.

The 34 users shown above are the ones who have a large probability to become "potential opinion leaders". Finding real identity of these users via social network, then we find these users' fans number is really large. Large fans number makes these users have a greater influence in social network community, and they can drive more users to participate in the topic discussion. In addition, Fig. 3 also shows that these special users have "attention - be followed" relationship between each other at the same time. People with greater influence follow on each other, and if they participate in the same topic interaction, the influence is double, this makes the ordinary events warm rapidly. At this moment, these users are more likely to be the "potential opinion leaders" of this event.

#### B. Evaluation of identity tags

Another condition is the identity tags of social network users. The identity tags can help us know what category of event these users who have qualification to become "potential opinion leaders" will follow and then become what kind of "potential opinion leaders". Add ID information of the 34 users and draw mutual follow relationship graph, as shown in the Fig. 4 below.

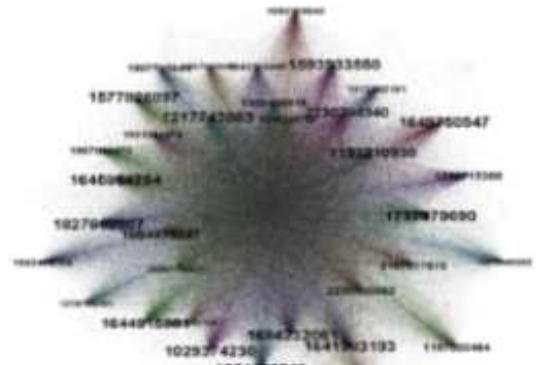


FIGURE 4. MUTUAL FOLLOW NETWORK GRAPH OF 34 USERS CARRYING WITH ID.

Select some nodes randomly from Fig. 4, and find their real identity according to the ID, the results are shown in Table 3 followed.

TABLE III. THE REAL SOCIAL NETWORK IDENTITY OF SELECTED USERS.

Microblog ID	Information of identify tags	Fans number	Identity tags by weight
1813080181	Angel Investor	11572873	0.6-0.8
1827652007	The Chinese academy of social sciences countryside social events research centre director	1832897	0.4-0.6
1737979690	network politics and pay attention to people's livelihood	20644867	0.6-0.8
1216766752	Professor of law school of Peking University	1291720	0.6-0.8
1807715644	The national "top ten official microblog" of 2011 and 2012. The national top ten police officers microblog of 2013.	620496	0.2-0.4

From the fans number of Table 3 we can know these users have a huge fans community, therefore the probability of being "potential opinion leaders" is very large. We can know what kind of events they will follow through their identity tags, and then know what kind opinion leaders they may be. For example, the identity tags of the user whose ID is 1827652007 are "The Chinese academy of social sciences countryside social events research centre director", so the events about agriculture, farmer and rural will be more attractive to him, he will make the event attract more users in a short time and then rapidly become hot.

#### IV. RELATED WORK

Currently, the research works about opinion leaders mainly discuss the opinion leaders whether exist or not, analyse the characteristics and influence of them, and propose recognition model. Reka Albert et al summarized the progress made in the field of complex network in recent years, they introduced the research results of network topology and dynamics, and discussed the random networks, small-world network, scale-free network, the interaction and robustness between network topology[4]. Zhiming Liu built the social network opinion leaders index system by using users influence and users activities, he used chromatography analysis method and rough set decision analysis theory to identify and analyze the opinion leaders[5]. Fusuijing Cheng et al proposed new algorithm to identify opinion leaders in BBS [6]. Shrihari A. Hudli et al presented an approach to identification of opinion leaders using the K-means clustering algorithm which does not require knowledge of the user's opinions or membership in other forums[7]. Yuzhen Li et al took advantage of activity, propagation and coverage to assess the influence of social network opinion leaders, built the evaluation index system[8].Matsumura N et al proposed the influence diffusion model based on text, the model could find influential comments, terminology, and members of network community[9]. Longwen Zhao et al proposed social network topic heat prediction model which based on the participative behavior of opinion leaders to study the development trend and characteristics of social network hot topics[10]. Dacheng Qu et al proposed three new methods to study hot events propagation and diffusion of innovation of them is important for finding opinion leaders[11]. Michael Bendersky et al proposed a existing term dependence model that can be trained by using assigning weights to concepts[12].

These researches have proved the existence and importance of opinion leaders, and have proposed a variety of recognition models and data mining methods, and they have a profound significance. But the researches about opinion leaders are all based on the precondition that hot events have already formed, which doesn't match the generation, transmission, evolution of the dynamic development process of hot events.

#### V. CONCLUSION AND FUTURE WORK

"Potential opinion leaders" means the small group who promote ordinary event develop before it become hot, and it matches the actual development process of hot events even more by starting from the early development stage of the event. In this paper, we use fans number and identity tags to identify

"potential opinion leaders". The former determines the probability the users become "potential opinion leaders", the latter indicates the users will become what kind of event's opinion leaders. Identifying "potential opinion leaders" can realize the network public opinion early warning better, and it has important significance in the management of emergencies.

Although the approach proposed in this paper can identify the "potential opinion leaders" in the social network community quickly and effectively, but it is still needed to improve the recognition criteria in the future study, add other influenced factors gradually, such as geography, nationality and so on to make "potential opinion leaders" enhance its effect gradually in the early regulatory stage of network.

#### ACKNOWLEDGEMENT

This work is supported by the NSFC under Grant No. 61370136 and thanks for the help of professors and students in Beijing Engineering Research Centre of Massive Language Information Processing and Cloud Computing Application.

#### REFERENCES

- [1] Paul F.Lazarsfeld. People's Choice [M]. Beijing: China Renmin University Press, 2012.
- [2] Liyang Huo. The microblog opinion leaders influence research [D].Beijing: Beijing University of Posts and Telecommunications, 2013.
- [3] Zhihong Li. An inherent characteristics and management counter under sudden public crisis information communication mode [C]. Library and Intelligence Service, 10, pp.88-91, 2007.
- [4] Reka Albert, Albert-Laszlo Barabasi. Statistical Mechanics of Complex Networks[ J ] . Reviews of Modern Physics, 74, pp. 47-97,2002.
- [5] Zhiming Liu. The identification and analysis of microblog opinion leaders in online public opinion [A]. Systems Engineering, 6, pp.8-16, 2011.
- [6] Fusuijing Cheng, Chenghui Yang, Yongfeng Huang, Linna Zhou. Algorithm of identifying opinion leaders in BBS[C].2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems, pp.1149-1152, 2012.
- [7] Shrihari A. Hudli, Aditi A. Hudli, Anand V. Hudli. Identifying online opinion leaders using K-means clustering[C].International Conference on Intelligent Systems Design and Applications, 12, pp.416-419, 2012.
- [8] Yuzhen Li, Yong Hu, Xi Xiong, Xiaojuan Ma, Min Zhang. An evaluation model for microblog opinion leaders [A]. Information Security and Communications Privacy, 2, pp.79-81, 2013.
- [9] Matsumura N, Ohsawa Y, et al. Influence Diffusion Model in Text-Based Communication[J]. In International World Wide Web Conferences, 17(3), pp.259-267, 2002.
- [10] Longwen Zhao, Rongtao Gong, Mingyan Chen, Haibo Yao. Microblog topic heat prediction research based on the participative behavior of opinion [A]. Intelligence Magazine, 12, pp.42-46, 2012.
- [11] Dacheng Qu, Wei Xin, Quan Zhang, Shuli Guo. The New Analysis Methods of Hot Events Propagation Based on Interdisciplinary Studies[C]. The fourth session of the National Social Computing, 6, pp.24-30, 2012.
- [12] Michael Bendersky, Donald Metzler, and W. Bruce Croft. Learning concept importance using a weighted dependence model. In International Conference on Web Search and Web Data Mining, pp. 31–40, 2010.