

## Identifying Opinion Leader in the Internet Forum

Chao Wu<sup>1,2</sup>, Chunlin Li<sup>1,2</sup>, Wei Yan<sup>1</sup>, Youlong Luo<sup>3</sup>, Xijun Mao<sup>1</sup>, Shumeng Du<sup>1</sup>  
and Mingming Li<sup>1</sup>

<sup>1</sup>*School of Computer Science, Wuhan University of Technology, Wuhan, P.R. China*

<sup>2</sup>*The Key Laboratory of Embedded System and Service Computing, Ministry of Education; Tongji University*

<sup>3</sup>*Management School, Wuhan University of Technology, Wuhan, P.R. China*  
*124528721@qq.com, chunlin74@aliyun.com, 364560925@qq.com,*  
*1069685498@qq.com, 404027587@qq.com, 1650172505@qq.com,*  
*610087798@qq.com*

### Abstract

*Opinion leader is an authority person who has great influence in BBS. Their linguistic behavior has a huge impact on net citizen's behavior and thought. In this paper, we propose an algorithm called OLRA (Opinion Leader PageRank Algorithm) based on topic-field to identify opinion leaders in the Internet forum. In the algorithm the closeness degree factor and sentiment analyses are taken into consideration. Meanwhile, these two authority values are defined as the weight of links among users. The data is collected from a number of posts on Tianya forum which is a famous forum in China. In the experiment, the algorithm is compared with Interest-based PageRank algorithm, online time Algorithm, and Experience-based Algorithm, the result shows that the OLRA algorithm can identify opinion leaders than others in the Internet forum effectively.*

**Keywords:** *The Internet forum, opinion leader, sentiment analysis*

## 1. Introduction

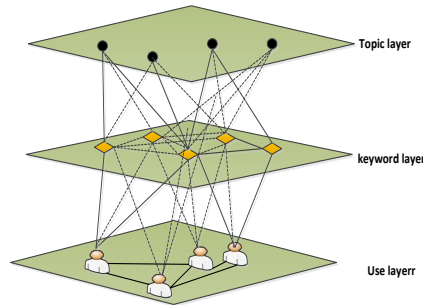
With the rapid development of modern information technology, network acting as the fourth media which follows newspaper, broadcast and television, becomes one of the major carriers that expression consensus of the public opinion [1]. Over the past 20 years, more and more people express their viewpoints online by using blogs, microblogging, Facebook, Twitter, Internet forum and other social medium.

Opinion leaders play a crucial role in controlling public opinion for the government, so it is significant to discover the opinion leaders from the Internet forum. Some scholars focused on opinion leaders' identification and mainly studied the problem as the following three aspects: a) Build the social networks by regarding netizen as a node and the replying relationship between netizen as an edge. Then sort the nodes by using out-degree, in-degree and closeness degree of graph theory to identify opinion leaders [2-3]; b) Put the PageRank [4] and HITS [5] as network hyperlink structure analysis algorithm application to identify opinion leaders in the field of the social network [6-7]; c) Information of users such as posts generated and replied, online time and vitality were recorded to construct the model of social networks and recognize the opinion leaders [8-9]. However, they did not consider the sentiment of the followers leading to discover opinion leaders difficultly.

The Internet forum community which belongs to network research domain can be considered as a form of social network [10]. Users in the Internet forum community tend to browse their interested posts and reply them. There is no doubt that many users likely interact with each other about interesting topics. So, we can classify the Internet forum

community to different topics according to the content of the posts, and similar posts are divided into the same topic. The user can browse post and communicate with others according to their interesting topic. The characteristics of BBS (A bulletin board system) are analyzed and the model of the Internet forum is constructed.

The model consists of three layers: the Topic-Layer, the Keyword-Layer and the User-Layer as shown in Figure 1. Users generate a post which may consist of several topics. So a topic maps server keywords.



**Figure 1. Structure of the Internet Forum**

To solve problems mentioned above, this paper explores the ideas of PageRank algorithm, and propose an OLRA algorithm which is based on topic-field to identify opinion leaders of the specific topic in BBS: posts are crawled from the Internet forum and keywords are extracted from those posts; Clustering method is used to analyze the posts according to the content of posts to form diverse topics, and finally a topic is spilt to several keywords; The OLRA algorithm is proposed by calculating the weight between users with the method of combining the emotion when the post is replied and the author's closeness degree in single topic to improve the accuracy of identifying opinion leaders. The algorithm can find opinion leaders quickly and accurately.

The rest of the paper is organized as follows: Section 2 gives the related works. The OLRA algorithm is introduced in Section 3. In Section 4 we evaluate our results and end with some conclusions and future works in Section 5.

## 2. Related Works

In recent years, lots of scholars have done some researches about identifying opinion leaders based on Social Network Analysis, Cluster Analysis, Analytic Hierarchy Process and Matrix Method Analytic. One of the most commonly used methods is the Social Network Analysis Method.

Huang [11], *et al.*, proposed a dynamic opinion leader rank algorithm to find the opinion leaders in the comment of Chinese news. Ding [12], *et al.*, presented a synthetic evaluation and sort-order algorithm of leaders in the model to find opinion leaders with a specific topic and through simulation experiments to verify the feasibility of the algorithm. Based on user influence and user activity two factors, Liu [13], *et al.*, used the rough theory for the modeling the identify opinion leaders, and used related experiments verifying the opinion leaders from dependent objects. Jiang L [14], *et al.*, realized the PageRank algorithm in the environment of MapReduce. Compared with the traditional algorithm, this algorithm improves the accuracy and efficiency of identify opinion leaders. Zhai [9], *et al.*, studied the BBS network leaders to identify problems. They proposed an interest-based page rank algorithm, and measured Tsinghua Forum data. They compared the applicability of number of algorithms. Eventually it was found that the interest-field based algorithms can identify opinion leaders most accurately. However, they did not consider the emotional of followers. Gao [15], *et al.*, constructed a social

network, and verified that the network has "small world" network features according to the relationship between post replies. They used the Average Path Length algorithm which they applied the theory to identify opinion leaders in online forums. But the algorithm has a high complexity, and only analyzed the largest connected graph but ignored the other branches of connected components, which may affect the accuracy of the results.

However, there are still some existing problems: Many researchers set the value 0 or 1 when considering the weights between the users and set the response times as the value; Lots of researchers had made some improvements based on the first point. But they did not consider the personal emotional factors of the repliers. These factors reflect the replier's attitude to the author's post. Sometimes the response is positive, but sometimes is negative, and negative responses will reduce the authority of the author. When most of the response is negative, the author is not a real opinion leader obviously. So we need to keep abreast of the follower's emotional position and attribute and take these into account; An important characteristic of opinion leaders which is that they would exchange opinions with follower frequently, so the relative position of opinion leaders is considered necessarily in the forums.

### 3. The Identification of Opinion Leaders Algorithm based on Specific Topic

Massive posts published by users in the BBS are distributed in different sections. So how to recognize posts which express the same topic from the messy and diverse content is a hot topic. As mentioned above, the next work focuses on the recognition of opinion leaders in a certain topic.

#### 3.1. Topic Clustering

The forum users browse and post threads based on their interested topic online. Specifically, a user can concern section of the forum for some help when he seek for help in the professional field BBS. So, it is necessary for us to classify those diverse posts into topics when analyzing the opinion leaders. The most important is to cluster and to classify analysis on the content of the advocate post, because the advocate post is usually more subject - oriented and have obvious purpose. Some preprocessing is performed on the advocate post, providing conditions for the subsequent clustering classification as shown in Figure 2.

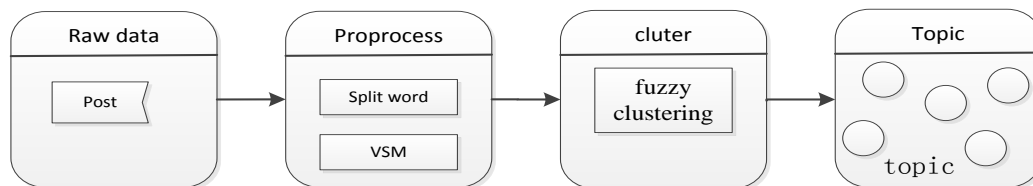


Figure 2. Process of Interest Topic Identifying

In the pretreatment, splitting the title and body of posts into words by using the Chinese Academy of Sciences Chinese Lexical Analysis System ICTCLAS [16] whose splitting accuracy is up to 97% and unlisted words identifying recall ratio is as high as 97%. Then extract three keywords with high weight as text entry feature vectors which indicate the post. The vector  $post = \{kw_1 : w_1; kw_2 : w_2; kw_3 : w_3\}$  denotes the post in the Internet forum and  $kw_i$  indicates the key words and  $w_i$  indicates the weight of the

keyword. The paper uses the ATF-PDF from paper [8], the weight of each word is calculated by formula (1).

$$w_i = \frac{tf_{ji} \times e^{n_i/N}}{N \times \sqrt{\sum_{i=1}^n tf_{ji}^2}} \quad (1)$$

$N$  is the number of documents which contained in the document set,  $n_i$  indicates the count of the word  $i$  in the document set,  $n$  is the length of the  $j$  word in the document,  $tf_{ji}$  denotes a word  $i$  frequency in the document  $j$ .

The post of forum can be treated as a three-dimensional vector by the above process. Following work clusters and classifies the post from the forum. We use the fuzzy clustering algorithm and assume that  $D = \{d_j | j = 1, 2, 3, \dots, n\}$  is a given text vector collection,  $m$  is the fuzzy weighting parameter,  $\varepsilon$  is threshold, the maximum number of iterations is  $L$ . The basic flow of the algorithm as following:

**Step1** Clustering number directly determines the number of discovered topics. The number of clusters is calculated as formula (2):

$$C = \begin{cases} n/2, & n > 2 \\ n/4, & 10 < n \leq 100 \\ n/5, & 100 < n \leq 1000 \\ n/8, & n > 1000 \end{cases} \quad (2)$$

**Step2** Randomly select count  $C$  of text vectors as the initial cluster center,

$V^0 = \{v_i | i = 1, 2, 3, \dots, C\}$ , the count of iterations  $L = 0$ .

**Step3** Calculating membership degree matrix  $U^L = \{\mu_{ij}\}_{C \times n}$  using formula (2),

$\mu_{ij}$  represents membership of vector  $d_j$  for the class  $v_i$ .

$$\mu_{ij} = \begin{cases} \frac{\sum_{k=1}^c \left( \frac{sim_{ij}}{sim_{kj}} \right)^{\frac{2}{m-1}}}{\sum_{k=1}^c \left( \frac{sim_{ij}}{sim_{kj}} \right)^{\frac{2}{m-1}}}, & sim_{ij} \neq 1 \\ 1, & sim_{ij} = 1 \end{cases} \quad (3)$$

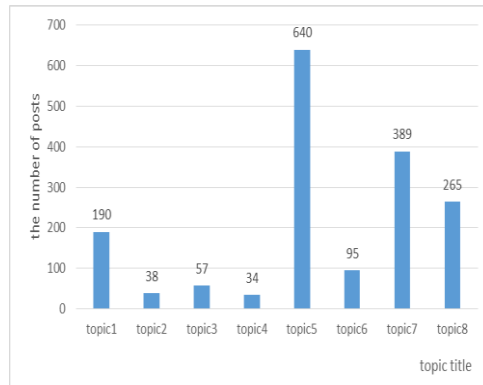
Where,  $1 \leq i \leq C, 1 \leq j \leq n$ ,  $sim_{ij}$  indicates the similarity between vector  $d_j$  and the cluster center.  $s$  denotes dimension of a vector.

$$sim_{ij} = \frac{\sum_{l=1}^s d_{jl} v_{il}}{\sqrt{\left( \sum_{l=1}^s d_{jl}^2 \right) \left( \sum_{l=1}^s v_{il}^2 \right)}} \quad (4)$$

**Step4** Check the conditions which the final iteration is met. If  $\|V_{L+1} - V_L\| \leq \varepsilon$ , then the algorithm stops and output divided matrix  $U$  and cluster centers  $V$ ; Otherwise, go to step (3).

The posts of BBS forum are divided into different topics after the above treatment. In this paper, Livelihood module in TianYa forum which is famous in China is data source. Posts in Livelihood module from 20th July to 21th July are collected and the fuzzy

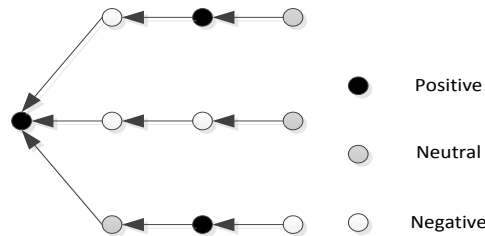
clustering to classify post according to the content of post is used. Topics distribution is shown in Figure 3. The topic5 is the most popular in the Internet forum from the Figure 3.



**Figure 3. Topic Distributing**

### 3.2. Emotional Attribute Analysis

The viewpoint of the author's post is expressed when the follower replied to the author and it suggest that the follower stand for the author or not. As shown in Figure 4 there are two kinds of emotion what one is neutral and the other are negative reply to the author and then the authority of the author was impacted. The list of positive negative emotion words come from DataTang [16] which is a professional scientific data sharing service platform and is dedicated to providing scientific data to support domestic and foreign universities, research institutions, enterprises and research and development related researchers ,which are used to quantify the emotional tendency factor when the followers reply.



**Figure 4. Forum User's View of Interaction**

Whether the user supports the author or not is judged with the content of post replied, in which emotional factor  $p_i \in [-1, 1]$  is a parameter. The attitude of replier to the author is neutral when  $p_i = 0$ , positive when  $p_i = 1$  and negative when  $p_i = -1$ .

The authority of evaluation from  $user_i$  to  $user_j$  is an accumulation of emotional factors after they repeatedly communicate. In this article,  $E_{ij}$  represents the authority of evaluation which is the average of the repliers' emotional factors for many times. The average of emotional factor is calculated as formula (5):

$$E_{ij} = \frac{\sum p_{ij}}{t_{ij}} \quad (5)$$

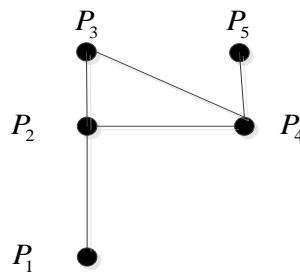
Where  $\sum p_{ij}$  represents the sum of emotional tendencies factor and  $t_{ij}$  indicates the sum of the number of their posts.

### 3.3. Structure Analysis of Network Forum

There is a lot of Internet Water Army who is a group of Internet ghostwriters paid to post online comments with particular content in the Internet forum. There is a significant feature that posts published by Internet Water Army which has mass of replied but rarely reply the followers. Therefore, a method proposed by Ding [14] may select those as opinion leaders. Thus, the closeness degree of the node in the social networks is taken into consideration when identifying opinion leaders in the community.

Network structural analysis is the major strategy of studying and identifying the most influential node. Complex network theory and social network analysis have become the important foundation of network structure analysis.

Closeness degree of complex networks, which reflects center of the nodes in the network, is commonly used in the structure of social indicators analysis of user. Node  $p_1$ ,  $p_3$  and  $p_4$  are connected directly to  $p_2$  as shown in Figure 5. They achieve  $p_5$  after  $p_4$ . As to  $p_1$ , directly connected with only  $p_2$ , it should achieve  $p_3$  and  $p_4$  after  $p_2$ , achieve  $p_5$  after  $p_2$  and  $p_4$ . Node  $p_1$  should go past  $p_2$  for three times and  $p_4$  for one time to achieve every node in graph. Therefore,  $p_2$  is more close to the center of the network than  $p_1$ .



**Figure 5. The Opinion Network based on User Reply and Sentimental Polarity**

Alex Bavelas [18] pointed out that the message which is the center of the network nodes spread over the entire network in the shortest time. Shorter distance network messaging means less time and cost. There are many kinds of definitions about the closeness degree of node in networks. One of the easiest, most natural definitions is proposed by Gert Sabidussi in 1966. Supposing  $d(p_i, p_k)$  is the numbers of edge in which  $p_i$  is the starting point and  $p_k$  is the end point, so the closeness degree of node  $p_k$  is  $c(p_k)$  expressed as a formula that as formula (6):

$$c(p_k) = \left( \sum_{i=1}^n d(p_i, p_k) \right)^{-1} \quad (6)$$

It can be seen that  $c(p_k)$  express a reciprocal of the sum of distance of node  $p_k$  to all other nodes. The greater closeness degree of node is, the closer it locates to the center of the network, which means it is important in the network.

Post are classified to different topics after clustering the post according to the content of posts. Then calculate the sum of followers' emotion factors. In order to prevent the Internet Water Army from being selected as opinion leaders, the closeness degree of the user in the network BBS network structure is introduced. Thus, opinion leaders will be correctly identified in BBS.

### 3.4. The Implementation of Opinion Leader Identifying Algorithm

Opinion leaders who have a huge impact on other users in the Internet forum, and their remarks published would affect public opinion guidance. Therefore, an OLRA algorithm is proposed based the sentiment analysis of the follower and the closeness degree of user to excavate opinion leaders of the Internet forum in this paper. We can obtain the rank of opinion leaders by calculating opinion leader ranking values (OS).

PageRank random walk algorithm is presented by Google co-founder Larry Page [4], and initially used in the Search Engines. The value of PageRank is calculated based on hyperlinks between pages, a page's votes is determined by importance of all the chain to its page. But then the researchers use the PageRank algorithm in social network, it is a basic algorithm that reflects the strength of the individual effects. In this paper, by using PageRank algorithm, OLRA algorithm is presented based on the user's emotional tendency and the location in the user's network Opinion-Leader-Rank algorithm. In the user node network, set the emotional factors of followers and the closeness degree of author as weight of edge between users. By using OLRA algorithm, each use's OS in the group can be represented as formula (7, 8).

$$OS(u) = (1 - d) + d \sum_{v \in B_u} \frac{OS(v) \times c(p_v) \times E_{uv}}{w(v)} \quad (7)$$

$$w(v) = \sum_{k \in T_v} E_{vk} \quad (8)$$

$OS(u)$  is the opinion leader ranking value of node  $u$ ,  $B_u$  is the set of points to the node,  $T_v$  is the set of points to the node  $v$ ,  $c(p_v)$  is the closeness degree of node in BBS,  $E_{uv}$  is  $v$ 's comprehensive emotional factor value to  $u$ ,  $w(v)$  is the sum of absolute value which is linked by node  $v$ .  $d$  is the damping coefficient, and it is normally set between 0 to 1. The damping coefficient is 0.85 in the references in this paper [9]. The opinion leader ranking values of all nodes is set 0.1, by iterating until convergence. All users'  $OS(u)$  are got, and user who has the highest  $OS(u)$  will be the opinion leader. The Algorithm is Described in Detail as Algorithm 1:

---

**Algorithm1: Opinion-Leader-Rank**

Input:  $G(V,E)$ ; //  $V$  is the set of users,  $E$  is edge,  $\varepsilon$  is the iteration threshold

Output:  $OS[i]$  //  $OS[i]$  is ranking score of opinion leader

---

- (1) Description:
  - (2)     Foreach user  $u$
  - (3)     {
  - (4)         Foreach user  $v$  link  $u$ :
  - (5)         {
  - (6)             compute  $E_{uv}$  using formula (5);
  - (7)             compute  $c(p_k)$  using formula (6);
  - (8)         }
  - (9)     }
  - (10)    Until current  $|OS[i] - OS[last\ one]| \leq \varepsilon$
-

## 4. Experimental Results and Analysis

Experiment analysis is necessary to verify the proposed algorithm. The environment of the experiments is shown as follows: CPU: Inter(R) Inside CORE i5, Cache size: 2048KB, Memory: 4G, Operation System: Windows7.

### 4.1. Experiment process

In this section, the data are collected from the TianYa forum and the start time of the post is from 20th July to 30th July 2014, the methods which have been described above to find opinion leaders are used. Data is crawled such as author and the replies from forum named “Livelihood” in TianYa forum in our experiment.

The datasets by splitting ICTCLAS [16] which is open-source software in Chin is preprocessed. The posts are clustered by their content and formatting different topic. The distribution of topic is shown as Table 1. Eight topics for the livelihood posts are associated with 3 keywords that can represent the content in the post, which provide a meaningful description of the topic. The posts are classified into clusters according to their content. From Table 1, we can see the topic “MH17 crash” is the hot topic. So people will discuss this topic in deeply.

**Table 1. The Eight Topics Extracted from the Livelihood Forum**

Top 3 keywords for each topic	Topics of livelihood
MH17, Ukraine, Ukraine rebels	Top1:MH17 crash
Education; Study; School	Top2:The education
English; Reading; Interest	Top3:English study
Children; Doctor; Hospital	Top4:Children’ health
Famous star; Popular song; Dad	Top5:Entertainment
Tiger; Fly; Xijiping	Top6:Chinese officialdom anti-corruption
Cinema; Box-office; Director	Topic7:Movie
Terrorist; XinJiang; Police	Topic8:The terrorist attacks in XinJiang

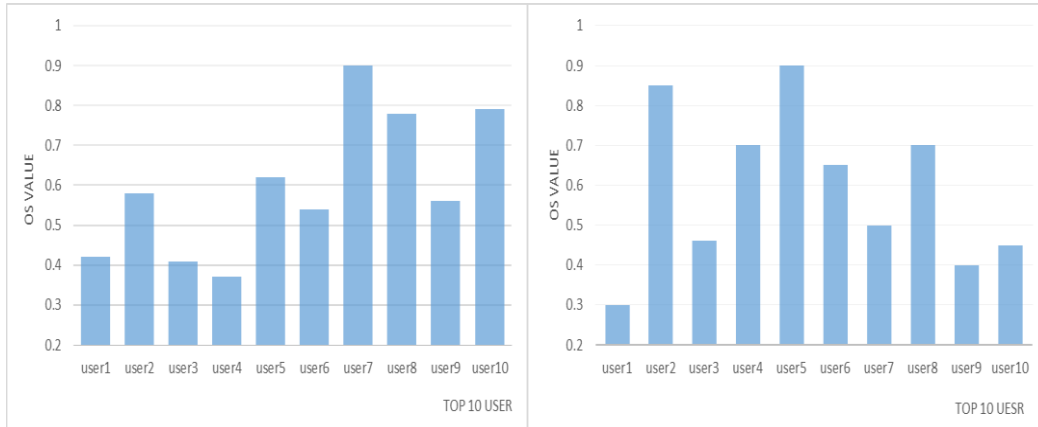
The next is to find opinion leaders for the each topic. The OLRA algorithm to get the opinion leader score of each user in the forum is utilized. Table 2 illustrates the top 10 opinion leaders with getting the highest opinion leader score on two topics in the Livelihood forum. For example, User7:467234 has the most top opinion leader score on the topic “MH17 Crash” and User5: 4356 has the highest opinion leader ranking on the topic “Chinese Officialdom Anti-corruption”.

**Table 2. Top Ten Opinion Leader for Two Topics in the Livelihood of BBS**

Topic1:MH17 Crash		Topic6:Chinese Officialdom Anti-corruption	
User ID	OS(leader value)	User ID	OS(leader value)
User1:89076	0.42	User1: 89076	0.30
User2:34786	0.58	User2:34521	0.85
User3:576	0.41	User3:908778	0.46
User4:3290	0.37	User4:6740900	0.70
User5:4356	0.62	User5: 4356	0.90
User6:8956432	0.54	User6:786753291	0.65
User7:467234	0.90	User7: 467234	0.50
User8:1245278	0.78	User8:56321	0.70
User9:85643	0.56	User9:0956432	0.40
User10:96754	0.79	User10:673328	0.45



From the Table 2, the opinion leader is inclined to be topic-specific. In other words, a person is an opinion leader in one topic but maybe a follower in another topic. For example, the user User7:467234 is the opinion leader for the topic “MH17 Crash” but ranks 5th for the topic “Chinese Officialdom Anti-corruption”, declaring that different roles do people play in different topics. The histogram are applied to describe the top 10 opinion leader of the two kinds of topic as shown in the following Figure 6:

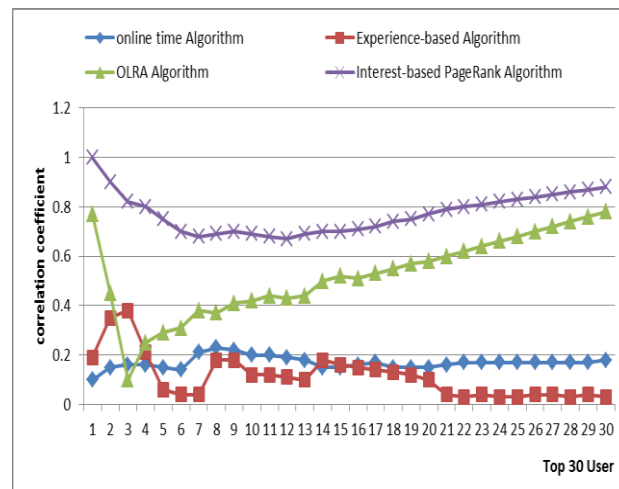


**Figure 6. Opinion Leader for Two Kinds of Topic**

“MH17 Crash” topic is the current hot topic, showed in Table 1. User7 is found to be the opinion leader of this topic through the OLRA algorithm. In order to verify User7 is the opinion leader, comparing the OLRA algorithm and other algorithms which are proposed by the other researchers, and introducing Coverage Ratio evaluation to verify the feasibility of each algorithm. The proportion of Top N users by calculating Coverage Ratio is calculated. This process can be expressed as formula (9):

$$CoverageRatio = \frac{\text{the posts number of topN}}{\text{the number of all user in the topic}} \quad (9)$$

Comparing the OLRA Algorithm with Interest-based PageRank algorithm, online time Algorithm, and Experience-based Algorithm to demonstrate the superiority of the algorithm. The results of Comparison are shown in Figure 7.



**Figure 7. Comparison of Four Algorithms about Correlation Coefficient**

As you can see in Figure 7, the in-degree of user has no effect on the online time Algorithm and Experience-based Algorithm. Therefore, it is not important to identify the opinion leader for the two algorithms. The in-degree value is closely related to Interest-based PageRank algorithm and the Opinion-Leader-Rank algorithm which is proposed in the paper and it is first increased then decreased. What's more, the in-degree value has stronger impact on OLRA algorithm than Interest-based PageRank algorithm, which demonstrates the correctness of considering the closeness degree to identify the opinion leader. Finally, in order to verify the efficiency of various algorithms in identifying the opinion leaders, we draw the Coverage Ratio to explore the efficiency as seen in Figure 8.

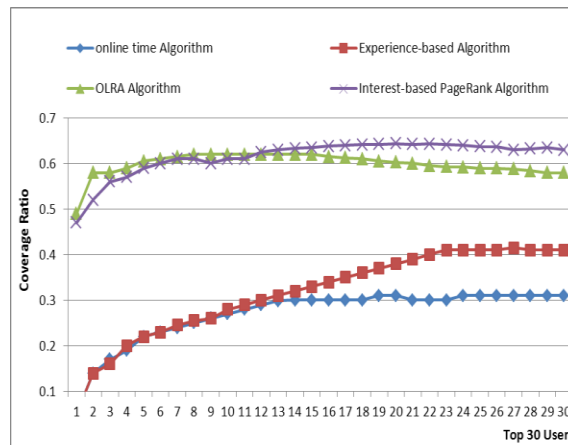


Figure 8. Coverage Ratio for Each Algorithm

Figure 8 indicates that the evaluation results of OLRA algorithm and Interest-based PageRank algorithm are relatively close, and they have better results than these simple sorting algorithm based on online time and experience, which means that the structure of the network itself has a stronger impact on identifying opinion leader. These two curves have a common characteristic that they level off after Top 10. In other words, opinion leader is often one of the top10 users. And OLRA algorithm calculating top 7 opinion leaders' score is more significantly than the efficiency of Interest-based PageRank algorithm. Therefore, it is very necessary to divide the posts into different topics according to the content of the posts. And algorithm based on online time and experience sorted is simple, and they do not take other factors that position online forum exchanges between users and their presence in the online community into account, leading to inaccurate results. In summary, this experiment based on real data demonstrates correctness and efficiency of the proposed algorithm.

## 5. Conclusion and Future Work

An OLRA algorithm is presented in to identify opinion leaders based on specific topic in online forums. The posts of BBS are clustered and classified into different topics according to their content. The closeness degree of author and the followers' emotional are taken into consider in the social networks. The experiments at last comparing OLRA and other algorithms verifies the accuracy and efficiency of OLRA.

Further research work should focus on improving the efficiency of OLRA. In order to improve the efficiency of the algorithm, how to deploy it to Hadoop distributed platforms is also an important point to study.

## Acknowledgments

The authors thank the editors and the anonymous reviewers for their helpful comments and suggestions. The work was supported by the National Natural Science Foundation (NSF) under grants (No.61472294, No.61171075), Key Natural Science Foundation of Hubei Province (No.2014CFA050), Applied Basic Research Project of WuHan (No. 2015010101010021), Program for the High-end Talents of Hubei Province.

## References

- [1] W. Cheng and X. Gao, "Network Communication and Social Development", Broadcasting Institute Press, Beijing, (2001).
- [2] S. C. Chu and Y. Kim, "Determinants of consumer engagement in electronic word-of-mouth (eWOM) in social networking sites", *Journal of International Journal of Advertising*, vol. 30, no. 1, (2011).
- [3] H. Liu, X. Yu and J. Lu, "Identifying Top-N Opinion Leaders on Local Social Network", *Journal*, vol. 268, no. 271, (2013).
- [4] L. Page, S. Brin and R. Motwani, "The PageRank citation ranking", *Bringing order to the web*, *Journal*, (1999).
- [5] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment", *Journal of the ACM*, vol. 46, no. 5, (1999).
- [6] J. Weng, E. P. Lim and J. Jiang, "TwitterRank: Finding Topic-Sensitive Influential Twitterers", *Proceedings of the third ACM international conference on Web search and data mining*, (2010).
- [7] N. Ma and Y. Liu, "SuperedgeRank Algorithm and its Application in Identifying Opinion Leader of Online Public Opinion Supernetwork", *Journal of Expert Systems with Applications*, vol. 41, no. 4, (2014).
- [8] J. Wang, J. Zeng and B. Zhou, "Online Forum Opinion Leaders Discovering Method Based on Clustering Analysis", *Journal of Computer Engineering*, vol. 37, no. 5, (2011).
- [9] Z. Zhai, H. Xu and P. Jia, "Identifying Opinion Leaders in BBS. Web Intelligence and Intelligent Agent Technology. WI-IAT'08", *IEEE/WIC/ACM International Conference on IEEE*, (2008).
- [10] M. V. Mahoney and P. K. Chan, "Learning nonstationary models of normal network traffic for detecting novel attacks", *Proceedings of the eighth ACM SIGKDD International conference on Knowledge discovery and data mining ACM*, (2002) July 23-25, Calgary, Canada.
- [11] B. Huang, G. Yu and H. R. Karimi, "The Finding and Dynamic Detection of Opinion Leaders in Social Network", *Journal of Mathematical Problems in Engineering*, (2014).
- [12] X. Ding, Y. Hu, W. Zhao and R. Wu, "A Study on the Characters of the Public Opinion Leader in Web BBS", *Journal of Sichuan University*, vol. 42, no. 2, (2010).
- [13] Z. Liu and L. Liu, "Recognition and Analysis of Opinion Leaders in Microblog Public Opinions", *Journal of Systems Engineering*, vol. 6, (2011).
- [14] L. Jiang, B. Ge and W. Xiao, "BBS opinion leader mining based on an improved PageRank algorithm using MapReduce", *Chinese Automation Congress (CAC)*, IEEE, ChangSha, HuNan, (2013) November 7-8, pp. 392-396.
- [15] J. Gao and J. Yang, "Analysis of Opinion Leader in On-Line Communities", *Journal of University of Electronic Science and Technology of China*, vol. 36, no. 6, (2008).
- [16] X. Li and C. Zhang, "Research on enhancing the effectiveness of the Chinese text automatic categorization based on ICTCLAS segmentation method", *Software Engineering and Service Science (ICSESS)*, 2013 4th IEEE International Conference on IEEE, (2013), pp. 267-270.
- [17] <http://www.datatang.com/>, (2014).
- [18] A. Bavelas, "A Mathematical Model for Group Structures", *Human Organization*, vol. 4, (1948), pp. 16-30.

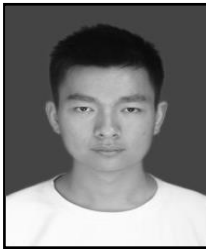
## Authors



**Chao Wu**, is a M.S. student in the Department of Computer Science at Wuhan University of Technology. He received his B.S. degree in Software Engineering from Huazhong University of Science and Technology Wuchang Branch in 2012. His research interests are in cloud computing.



**Chunlin Li**, is a Professor of Computer Science at Wuhan University of Technology. She received her M.S. in Computer Science from Wuhan Transportation University in 2000 and her Ph.D. in Computer Software and Theory from Huazhong University of Science and Technology in 2003. Her research interests include cloud computing and distributed computing.



**Wei Yan**, is a M.S. student in the Department of Computer Science at Wuhan University of Technology. He received his B.S. degree in Software Engineering from Wuhan Institute of Technology in 2014. His research interests are in cloud computing.



**Youlong Luo**, is a vice Professor of Management at Wuhan University of Technology. He received his M.S. in Telecommunication and System from Wuhan University of Technology in 2003 and his Ph.D. in Finance from Wuhan University of Technology in 2012. His research interests include cloud computing and electronic commerce.