The influence of opinion leaders on community evolution based on Signed Network

Linlin Cao, Mingchun Zheng *

College of Management science and engineering, Shandong Normal University, CHINA

Abstract

This paper presents a new perspective on the evolution of online social communities, and finds that the evolution of opinion leaders in the evolution process, which is different from the static time slice, which may miss the dynamic network evolution. In this paper, we analyze the dynamic process of the formation of the community network in the signed network, and analyze the evolution of the structure under the opinion leaders.

Keywords

Signed network opinion leader community evolution .

1. Introduction

With the rapid development of the Internet, human behavior has a great change, most of the behavior occurs in the network, in particular, the development of social networks almost occupy most of the space. Social networks have become an indispensable tool in life. Communication between users in a social network, share their own views and what one sees and hears. Social networks generally use graph to express, each node represents a user's individual, the edge represents connection between the user or the interaction. Individuals on the basis of mutual understanding, interest in the same or personal worship and other factors in the social network to form a complex structure.

Most social networks focus on the positive relationships in the network (friends, trust, etc.).however, if we analyze the relationship between users in online social networks, we will find positive links is only a simple assumption, in fact users not only show a positive attitude through labels between friends, but also appear to attitude toward opposition. In recent years, a great deal of research has focused on the negative relationship between social relations.

Signed network [1] is a network with a plus or minus edge attributes, in which the edges represent a positive relationship (friends, support, trust etc.), represented by the "+", minus represent the negative relation (enemy, do not entrust, opposition), represented with "-". There exists cooperation and opposition in social, biological, information fields and many complex systems .such as: in the international relations including cooperative and hostile relations, in the biological domain, in which information is transmitted by Inhibition or promotion. In the information field, in the realistic social network there exists supports and oppose between users, a user can label others as friends or enemy. In addition it is also has support or oppose opinions.in addition to the above mentioned, users may have some implicit or negative relations. You can show that there is support or between these two pages.

These need to be opposed by a complex, specific analysis to determine the support or opposition. These complex systems can be said for the abstraction of the signed network is described, the plus and minus can well reflect the reality of social relations.

The key to the research of the signed network lies in the comprehensive utilization of positive and negative information. The existing research ideas and methods for the two value network of no sign simply do not apply to the signed network. In any case, the symbolic network can be transformed into a simple two value network, which will lost valuable information or not worth being studied anymore .The formation of minus symbol of network has an important influence to the network structure, structure evolution and the dynamics. On the other hand, in the process of the formation of

the symbol network, how to define the negative edge and how to deal with the relationship between the positive and negative effects is the key and difficult point in the research of the symbolic network.

Opinion leaders[2] can influence other users' views and behaviors, which is an important factor that affects the evolution of the community. It plays an important role in the process of network evolution. Community evolution is to observe the changes in the community structure implied process as the goal, it mainly focus on the differences of community structure in the evaluation of different time segments, to observe the formation, development, change and the disappearance of the life cycle .

In order to study the evolution of this region, we should find the community. But in many scholars' studies, the two are not distinguished. Because in the research process of evolution, we must identify community first .so called dynamic community research .From the view of technology route, the research steps of network evolution are: time slice, computing the structure of each time slice, analyzing the structure of adjacent time slice, and finding the change of community

The main idea of this paper is to use the positive and negative edges of the signed network to study the impact of opinion leaders on the community structure evolution under the specific topic of the micro blog social network. The first section introduces the basic theory of community evolution, opinion leaders and the symbolic network, which includes the theory of equilibrium theory and position theory. In this paper, we only used the signed network as a tool to measure the evolution of the community. The second section describes the research status of community evolution. The third section describes the evolution of the community network analysis, from the change of the point and edge of the community. In the fourth section, the influence of opinion leaders is verified, and the evolution of the community is different from the previous trend. The fifth section is the summary and prospect.

2. Related theory

2.1 community structure

Social networks connect users in different geographical ,age domain and others in virtual space .it has become an extension of the relationship in realistic world ,which has a distinct cluster structure and it is a kind of strong relation network . In addition to the same interest among users, the interaction relationship (such as the concern in the micro blog, etc.) constitutes a one-way weak link. Presents the salient features of the "birds of a feather flock together", in the real world friends, celebrity effect can become the network structure and browse the information, the same topic of interest and will encourage users to change the relationship between each other, and structural relationships in turn promote information dissemination within certain limits range, and then form the evolution of the system. Group structure[3] is an important structure in online social networks, We use graph to express. Based on sub-graph is defined locally is between individuals within the various more closely linked, and other graph with relatively sparse .Newman [4] called this graph structure which satisfies the characteristic community structure. Related terns are community group or group structure etc. In this paper, the network structure of all the users involved in a controversial topic in the micro blog is called the community which we want to study. Community structure [5] is changing with time, including the changes of network topology and the relationship between nodes in the network. At present, many researchers are dedicated to the discovery of the community, and then to study the evolution of the community.

In the research of complex networks, we use the knowledge of graph to represent the topological structure of the network, so we still use the graph to express in the signed network. The nodes in the graph represent the individual in the network, and the edges in the graph represent the interaction between the users in the network.

2.2 signed network

The constructed signed network represents a typical two value network .that is the topic of our research is a simple choice of supporting or opposing view. For example, a user initiated the topic: should retirement age be postponed a few years, the user's view is only considered should or should

not, the statement of the identity of their own identity, and does not agree with the reasons for the non-recognition, so that constitutes a network of only two relative views. Support for the statement of their identity .Reasons for the publication of the objection. So that constitutes a only two relative views of the network. We define the interaction between user nodes as the attribute values of the nodes in the network. There is an interactive communication between two nodes and they agree with each other, so they support each other, represented with "+", there is a mutual exchange but the view is not the same as the opposition, represented with "-", If there is no interaction between the two nodes, there is no edge.

In previous research, signed network is just an extension of complex network, the main theoretical results including Heider's[6] structure balance theory and the status theory, the main basis of balance theory is "friend's friend is a friend, the enemy's enemy is a friend, enemy's friend is an enemy, friend's enemy is an enemy", Such a relationship to define positive or negative relationships between nodes. In the equilibrium theory, there are two kinds of equilibrium states in figure (1): figure (a) (c), the other two kinds of situations are considered to have the trend of the transformation of the other two kinds of equilibrium. The structural balance of a triangle can be determined by the product of the sign of the 3 edges of a triangle. If the triangle is positive, the triangle is a structural equilibrium; otherwise, the structure is not balanced.
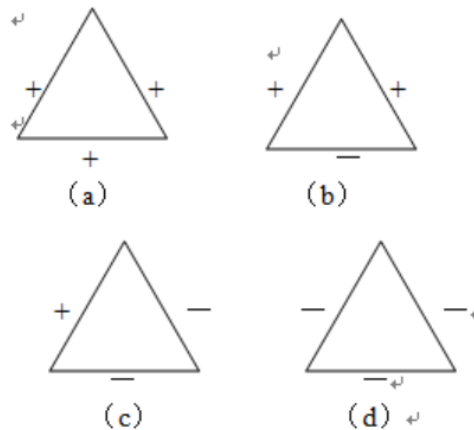

Figure 1: a combination of triangles in a signed network

The status theory: the structure equilibrium network provides a theoretical basis for the analysis of the undirected signed network, but there are some deviations in the characterization of the directional sign network. Therefore, Leskovec[7] and Kleinberg et through a large number of analysis of online social network users in the network to establish the intent of the network and the relationship between the triangle,proposed a new theory for the network, the status theory.

2.3 opinion leader

Our research is set to the micro blog environment, so we need to tap the opinion leaders in the micro blog, opinion leaders are defined as a class of people [8]: actively participate in the topic discussion, to express their views, opinions, actively respond to other people's comments and have a higher authority and influence. The relationship in the micro blog is divided into two kinds: the strong relationship based on the knowledge of the relationship and the weak relationship based on the topic. Such as friends, colleagues, relatives from the reality of the world extend to the network is a strong relationship social network structure. The relational structure, which is composed of a one-way relationship, is a weak relation structure. The factors that affect the evolution of the community structure in the strong relationship structure are some relations in the reality, but in the weak relationship, these factors are not affected, Mainly is whether the information source is reliable, for example, CCTV micro blogging released a news is more likely to accept than a certain user in the micro blogging released information for the user in the micro blog. But it also related to opinion leaders and the user's own personality, values, beliefs, and so on.

Opinion leaders are individuals who can affect others in the corresponding environment. Many times, they are also called influential people. According to the different data objects in the process of analysis, the method of opinion leaders tapping can be divided into the method based on network structure and the method of interactive information. The social network structure reflects the importance of the location of a node in the network to a large extent. If the position of the user in social networks can reflect his leader position in social activities, then based on the PAGERANK [9] algorithm can be measured. Some researchers [10,11,12,] have found that most of the opinion leaders have a higher status than ordinary users, and proposed a similar ranking of experts with PAGERANK algorithm. Song [13] proposed Influence Rank algorithm, it can be used to measure the influence of the user and publish information, so as to determine the user's ranking. According to the user's centrality in social networks, it is possible to infer that a person is a leader, and that a high degree of central user may be a local opinion leader, and the close central degree and Medium number Center can be regarded as the judgment index of global opinion leaders. The method based on network structure emphasizing the leader's leader characteristic and neglect the importance of the information released, Although the opinion leaders of social network have a greater degree of node, but the users with a large degree are not necessarily the opinion leaders, So the analysis of the influence of the user's information and the dissemination of information can be used to find the opinion leaders in the network , opinion leaders in the micro blog can be identified by the influential post. Each of these two methods has advantages and disadvantages, and the method based on network structure is simple, but the accuracy is relatively low, and there is the possibility of a miscarriage of justice in the real network, The results obtained from the analysis method based on mutual information are objective and accurate, but it is difficult to adapt to the large scale social networks because of the large amount of information.

We mining opinion leaders through the method based on network structure.

$$d_i = \sum_{j=1}^{n} a_{ij}$$

if node I is connected with j ,and i≠j , a_{ij}=1,orelse a_{ij}=o, N represents the number of nodes in the network. The degree centrality is the normalized processing of the nodes in the network.

$$C_D(i) = \frac{d_i}{N-1}$$

Where d_i represents the degree of node i, N represents the number of nodes in the network .close centrality measures the distance from one node to other nodes.it is defined as:

$$C_C(i) = \frac{N-1}{\sum_{i,\ j,\ i \neq j} d_{ij}}$$

Where N represents the number of nodes in the network, D_{ij} is the shortest path length of the node i to node j .Medium centrality is used to calculate the number of the shortest paths in the network, which directly reflects the importance of the nodes in the network communication:

$$C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

σ_{st} represents the shortest path number from s to t, σ_{st} (i) represents the shortest path that through the shortest path i .The greater the medium centrality is, the greater the probability that the information is transmitted through the node, then the information transfer to other nodes is more dependence on the node; the node is the core node.

We take these three metrics to take the Top-k strategy to select the user of large influence based on structure. Then make an intersection operation to obtain the alternative opinion leaders set :candidate= $DC_{top-k} \cap CC_{top-k} \cap BC_{top-k}$.

3. Preparatory knowledge

3.1 symbols and definitions

We use the map sequence express community structure in social networks G= {G$_1$, G$_2$... G$_n$}, in community structure G= {V, E, W}, the edge sets V and the edge set E changes with time constructed dynamic social network. W indicates the attribute values of the edges between nodes, defined:

$$w_{ij} \begin{cases} 1 & \text{There is an interaction between the node i and j, they support with each other} \\ -1 & \text{There is an interaction between the nodes but the view is not the same} \\ 0 & \text{There is no interaction between the nodes} \end{cases}$$

The network structure at a particular point in time i is defined as a snapshot at i, the social network in the snapshot I remembered as graph g$_i$= (v$_i$, e$_i$, w$_{ij}$).

D$_i$ (V) indicates the degree number of node V at the time point i

E$_I$ (V) represents the number of neighbors of the node v at the time point i

βindicates the member fluctuations in the time period of (i, i+1)

σ(i, i+1) represents graph distance of time (i, i+1)

η indicates the active members of the community Definition 1: graph distancesσ(i, i+1) used to measure the differences between two graphs.

$$\sigma(i,i+1) = \sum_{v \in V} \left| \log \frac{d_{i+1}(v)}{d_i(v) + 1} \right| + \sum_{v \in V} \left| \log \frac{e_i(v) \cap e_{i+1}(v)}{e_i(v) \cup e_{i+1}(v)} \right|$$

By using the deformation form of the relative entropy, we can draw the graph distance between any two consecutive time slices. You can get a distance graph change sequence {σ(1,2), σ(2, 3)…σ(i,i+1)…σ(n-1,n)} .

Definition 2: the relative entropy   it is usually used to measure the distance between two distributions. In the statistics, it is defined as the logarithm of the likelihood ratio, which is defined as follows: the relative entropy of two probability density function p (x) and Q (x) is:

$$D(p\|q) = \sum_{x \in \chi} p(x) \frac{\log p(x)}{q(x)} = E_p \log \frac{p(\chi)}{q(\chi)}$$

Definition 3: Member volatility

For the topic community ,different users will join in the topic discussion with time going, to express their views, In the evolution of the community, Our edge represents the interaction of the two node users, In this paper, as long as the node to participate in the discussion of the topic, That is the affected node. A number of factors affect the user's membership. According to the research of Backstrom Lars [15], the cumulative effect of a user to join a virtual community is found.

$$\beta = \frac{\sum_{v \in V} d_{i+1}(v) - \sum_{v \in V} d_i(v)}{\sum_{v \in V} d_{i+1}(v) \cup \sum_{v \in V} d_i(v)}$$

Definition 4 :Active degree

In the evolution process of the topic, according to the general rule, a period of time after the beginning of the topic, the topic of discussion will reach a climax stage; it Is a relatively mature stage, Then began to decline, gradually quiet, no body participate in the topic of discussion any more, the topic of death. In our study, the evolution of the topic is influenced by the opinion leader, so the opinion leader will lead to a new round of discussion and reached another climax stage. In this process, the active degree of the members will have an obvious change.

$$\eta = \frac{\sum\limits_{v \in V} e_{i+1}(v) - \sum\limits_{v \in V} e_i(v)}{\sum\limits_{v \in V} e_{i+1}(v) \cup \sum\limits_{v \in V} e_i(v)}$$

Definition 5 :Time window setting

In the observation of community evolution, the research objectives and methods are different, but they are both based on the spatial dimension of the community, then extended time dimension, in order to find the community information and community changes at different time points or time window, so the time slice is a common computing features. First of all, we should divide the time slice of the social network data, and integrate the data in a certain time window. Then calculate the community structure of a time slice. Analysis and comparison of the community structure of the adjacent time slices Get the change of community network; finally find the influence of opinion leaders to the community.

In the process of data acquisition, we need to update the data continuously, access to the users 'statement of the latest network, But if monitoring the real-time data, the amount of computation is too large, the complexity of the algorithm is more complex, it is not suitable for us to conduct research.

So we can get a more appropriate time window according to the collected data set of micro blog, The length of the time window is set to WS (size window) =8 to observe the network changes on both sliding time sides.
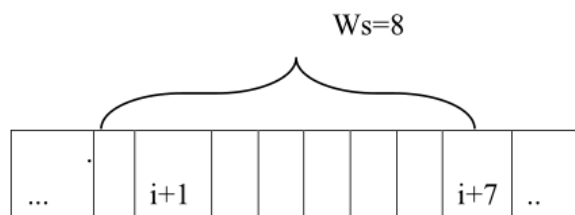

Figure 2: moving window with size=8

4. Generation and analysis of signed network

Our research is to extract the signed network from the text of the users, By analyzing the interaction between the users of the text to extract the user's attitude, so as to construct the signed network we want to study.

Analysis of the attitude of the text, from another point of view is to extract the text of the emotion, the attitudes to one problem can be passed out from the feelings in the content of the text of the exchange between the users, so we can extract the user's feelings from the text, and then forecast the users' attitudes. Most of the methods for the evaluation of the scholars were extracted from the evaluation of words, to discriminate the feelings, The evaluation of the extraction and discrimination of words and phrases is often a work of integration, which is divided into two methods based on corpus and lexicon. The extraction and identification of words based on corpus is mainly the use of large corpus of statistical. Observe some phenomena to mine the evaluation words in the corpus and judge the part of speech. Some early scholars found: The two adjectives that are connected by a conjunction, such as: (but, and), tend to have a certain relevance. Such as and connection of the two adjectives (such as beautiful and lovely) has the same polarity, However, but connection of the two adjectives (such as lovely but unnatural) has the opposite polar. Based on this phenomenon, some scholars have developed a large number of adjectives evaluations of speech. Later Riloff [14] developed a mechanism for evaluating nouns, but there are some other words that have not been evaluated in general.

Methods for evaluating the extraction and discrimination of words based on lexicon mainly use the Semantic relation between words to mine the evaluation words. Dictionaries are generally used to use WordNet or HowNet [16] etc.In addition; some scholars identify the evaluated words' polarity by using the method based on graph.

Our goal is to tap the attitude of users in online communities, and we abstract the signed network representing the ideas of users from social networks. Note that our edge is the same as the user's view, and not only will the exchange be generated. The edge represents the point of view, which is different from previous studies. For example: A stated Lee was bitten by Ye's dog, but Ye said it is because Lee forced the dog to take pictures, so Lee sued Ye, Ye get out of the entertainment circle. B commented: the Ye is really too bad. C commented: unexpected events cannot be controlled, Ye is just tell the truth. As is shown in figure 3 ,so we consider that A is agree with B, there is a positive edge between the A and B . A has different opinions with C,so there is a negative edge between the two.
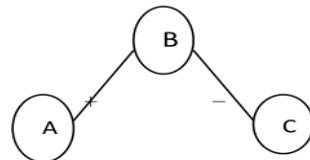

Figure3: signed between A, B, C

Method is divided into several steps, The first step is to identify the word's polarization, it is also known as polar words, emotional words, it particularly refers to words with emotional tendencies. Obviously, the word with polarization [17]is a good indicator of the subjectivity of the sentence, So their presence is highly correlated with the attitude of the user. Based on a large number of previous studies, the usual work is to establish approach based corpus or dictionary to extract and identify the polar of words. We recognize the polarity of words with a random walk method based on word association graph.

The second step is to turn to the sentence level, by detecting different sentences to discover the attitude from the sender to the recipient. We train a classifier[18] based on multiple information sources to make predictions. We use lexical items, polar tags, partial speech labels, and dependency parsing tree to train classifiers to recognize attitudes in a sentence.

Finally, we build a network based on user communication. We give each of the edges symbol based on the forecast by the words and sentences.

Our proposed method is applied to a large number of online exchange data sets. In order to evaluate the effectiveness of our automatic identification procedure, we turn to the human annotation to detect the positive or negative of the text of the communication, We compare the obtained results with the results obtained by the automatic recognition and the results are basically consistent.

After building the signed network we need, it's time to analyze the generated network; we start with the changes of the structure of the network and the changes of symbols of the signed network, and study the influence of opinion leaders on the community network evolution. In our study, according to the normal situation, users in the community network are influenced by the opinion leaders with the increase in the activity of opinion leaders. The positive and negative[19] numbers between the users will change. We first find a positive opinion leader, whose opinions are active and provide us guidance. From this view, the number of positive symbols in the network will increase, and the proportion θ will increase.

$$\theta = \frac{Nedges(+)}{Nedges}$$

5. Experimental simulation analysis

Sina Weibo provides a rich application programming interface for developers. We can get the part of the micro blog data through these interfaces. During the experiment, we used the method of controlling variables to carry out the experiment. Firstly, the opinion leader remained silent and we observed the development of the community network [20]. After a long time, the opinion leaders in the community network express their views and convey the positive information. Observe the structure [21]of the community network and the change of the positive and negative numbers between

the users. We select a topic which has 300 members participated in. using the method we generated a community network is generated based on the data we grabbed as figure 4.
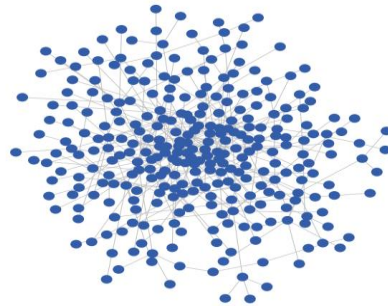

Figure 4: the signed network

According to the evolution of our research community , we first observe the edge and the point of the network, and get the trend of the change and the development situation of community evolution. The conclusion is obtained by the relative entropy of the edges and points of the community network: At the beginning of the establishment of the community, a large number of participants discussing the topic. Everyone express different views, Each airs his own view, then reached a climax in the discussion after a certain period of time, According to the general rule ,after the climax, the topic gradually to decline, Everyone's heat reduced, the topic gradually died away. But we add the influence factors of opinion leaders in our study, after the opinion leaders have joined the discussion; it is surprising that our community network has reached another climax. Although it is not as intense as the initial stage, it is more active than before. This is different from previous studies. Because of the special nature of the community network we selected, the conclusion of the research may be different from other scholars. Conclusions are shown in Figure5:
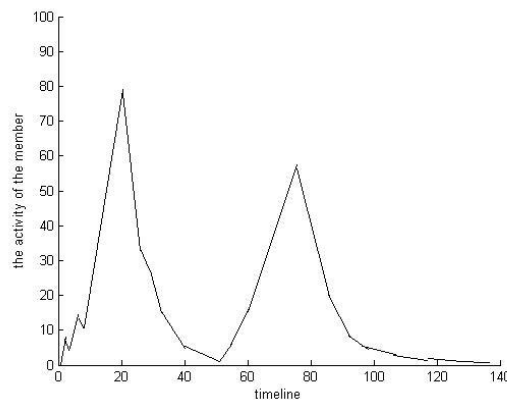

Figure 5: the activity of the members

We observe the development and evolution of the community based on the sliding time window of time window. In the process of experiment, we found that the proportion of positive and negative edges in the community is almost the same at first. Before we find out the opinion leaders both of the two relations are showing an increasing trend .After our opinion leaders began to move, the opinion leaders we chosen are positive opinion leaders, leading the public to the positive side. The percentage of negative symbols in the community begins to fall, and the final state of the stability is maintained. According to the calculation, the trend is shown in figure6:
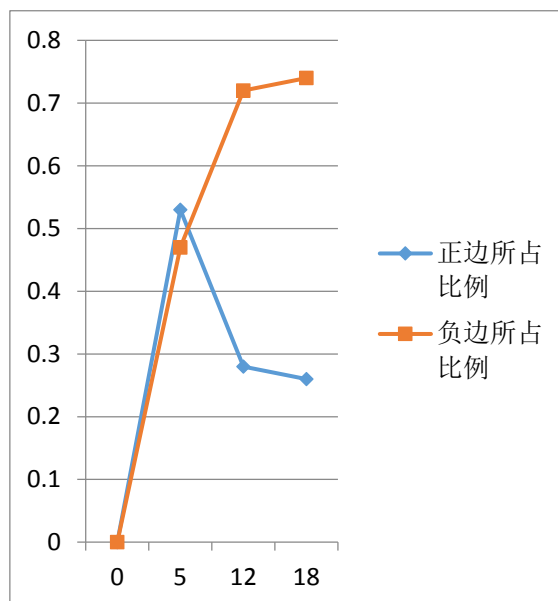
Figure 6: the proportion changes

6.　Conclusions and outlook

The rising of social network research provides an effective tool for people to study the communication behavior of real society. This paper starts from the evolution of the network, from a new perspective study of the influence of opinion leaders on community evolution. We have the same method as the traditional research network evolution. We both start from the division of time, the emphasis of the analysis of the trend based in the time division is placed on the evolution of the relationship between the network nodes within the community. Traditional research scholars mainly analyze the changes of the external characteristics of the network community, and then find the network's growth, reduce or the evolution of the demise of the trend. From the experiment we can see the impact of opinion leaders on the evolution of community networks, positive opinion leaders can guide the majority of users to establish a correct attitude towards things, So we can start from the opinion leaders in the network of public opinion control, so as to reduce the generation and spread of network public opinion, increase the security of the internet. At the same time, it provides a broad channel for the dissemination of information on the Internet.

References:

[1]. Cheng SQ, Shen HW, Zhang GQ, Cheng XQ. Survey of signed network research. Ruan Jian Xue Bao /Journal of Software, 2014, 25(1):1—15(in Chinese). http://www. jos. org. cn / 1000 · 9825/4503. Htm

[2].Zhiming Liu, Lu Liu, The identification and analysis of opinion leaders of network public opinion in micro blog./systems engineering,2011,06:8-16.

[2]. Lan MW, Li CP, Wang SQ, Survey of sign prediction algorithms in signed social network/Journal of computer research and development, 2015, 02:410-422

[3]. David Lo, Di di Surian, Philips Kokoh Prasetyo, Kuan Zhang, Ee-Peng Lim. Mining direct antagonistic communities in signed social networks [J]. Information Processing and Management, 2013, 494:.

[4]. Girvan M, Newman MEJ. Community structure in social and biological networks. Proc. of the National Academy of Sciences of the

United States of America, 2002, 99(12):7821−7826. [doi: 10.1073/pnas.122653799]

[5]. BIN Wu, Bai Wang, Qisheng Yang. Framework for Tracking the Event-Based Evolution in Social Networks./Journal of software.2011.07:1488-1502

[6]. Heider F.Attitudes and cognitive organization.Journal of Psychology, 1946,2l(1): 107—112. [doi:10.10803/00223980.1946. 9917275]

[7]. Leskovec J, Huttenlocher D, Kleinberg J. Signed networks in social media. In: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems. New York：ACM Press, 2010. 1361—1370. [doi：10. 1 145/

[8]. Research on micro blog opinion leader identification model. /Journalism and Communication Studies. 2011, 06:81-88+111.

[9]. Pu Bing-yuan,Huang Ting-zhu,Wen Chen,An improved PageRank algorithm:immune to span [C].IEEE 2010 Fourth International Conference on network and system Security,2010:425-429.

[10]. Junze Wang, Yalei Wang, Identification model of opinion leader [J]. Journalism and communication research.

[11] Xing Hua Fan, Zhao Jing, Fang Binxing, Li Yuxiao . Influence Diffusion Probability Model and Utilizing It to Identify Network Opinion Leader [J].Chinese journal of computers. 2013, 02:360-367.

[12]. Xuefeng Ding, Hu Yong, Wen Zhao .A study of the characters of the public opinion leader in web BBS [J].journal of Sichuan university (engineering science edition),2010,02:145-149

[13]. Yu He, Hongli Zhao, Overview of complex network evolution, /Journal of equipment command and Technology Institute,2011,01:120-125.

[14]. Larusso N, Bogdanov P, Singh A. Identifying communities with coherent and opposing views. In: Proc. of the 1 5th Annual Graduate Student Workshop in Computing. Santa Barbara：UCSB. 2010. 31-32. http://gswc.cs.uesb.edu/2010/proceedings.pdf

[15]. Pang B, Lee L. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2008, 2(1—2)：1-135. [doi: 10. 1561／1500000011]

[16]. Cartwright D,Harary F. Structural balance: A generalization of Heider'S theory. Psychological Review,1956, 63(5):277—293.[doi: 10. 1037／h0046049]

[17]. Patrick Doreian, Andrej Mrvar. Partitioning signed social networks [J]. Social Networks, 2008, 311:.

[18].. Ahmed Hassan and Dragomir Radev. 2010. Identifying text polarity using random walks. In ACL'10.

[19]. Asur S, Parthasarathy S, Ucar D. An event-based framework for characterizing the evolutionary behavior of interaction graphs. In: Proc. of the KDD 2007. 2007. 913−921. http://www.sigkdd.org/kdd2007/ [doi: 10.1145/1281192.1281290]

[20]. Girvan M, Newman MEJ. Community structure in social and biological networks . Proc. of the National Academy of Sciences of the United States of America, 2002, 99(12):7821−7826. [doi: 10.1073/pnas.122653799]

[21]. Hopcroft J, Khan O, Kulis B, Selman B. Tracking evolving communities in large linked networks. Proc. of the National Academy of Sciences, 2004, 101(Suppl. 1):5249−5253. [doi: 10.1073/pnas.0307750100]