

Research on the Ecological Dissemination Characteristics and Control of Information in Microblog Network Based on the Complex Network Theory

Xiaolan Guan¹, Zhenji Zhang², Shugang Zhang² and Feng Cao²

¹Beijing Institute of Graphic Communication, Beijing 102600, China

²Beijing Jiaotong University, Beijing 100044, China
{08113101, zhjzhang, 10113132, 09113108}@bjtu.edu.cn

Abstract

On the basis of the information dissemination law in Microblog and its network topology, this paper analyzes the characteristics of the user nodes, structure characteristics of the groups, and the dissemination characteristics of the information of Microblog, and then proposes some related countermeasures to control and guide the healthy and rapid information dissemination of Microblog network. The research shows that the Microblog network renders mesh topology with good characteristics of cluster, centrality and similarity. The efficiency of information dissemination in Microblog network is relatively very high and the network is not fragile. Meanwhile, the node's role in the dissemination of information varies greatly, and it is easy to form opinion leader. It has important significance to strengthen the government's authoritative coverage and rapid response capability, make the opinion leaders play an active role in guiding correctly for the information dissemination management in Microblog network.

Keywords: *Microblog Network, Topology Structure, Ecological Group, Information Dissemination, Opinion Leader*

1. Introduction

As a form of online social networking applications, Microblog has gotten rapid development in recent years, such as Twitter and Sina Microblog are just two typical examples. Compared with the Social Networking Sites and Blog, Microblog has the features of both of the “social network” and “media platform”. Its quickly publish way of short text greatly facilitates the users to share and discuss information, and thus contributes to the formation of a new social network. In addition, the huge numbers of Microblog users, frequent information exchange and rapidly spread make it one of the major sources of network carriers that reflecting public opinion. In summary, Microblog plays an extremely important role in the event of public opinion and information dissemination. Therefore, it has very important theoretical and practical significance to research the information dissemination in Microblog network and its properties, laws, etc., and then research how to control and guide online public opinion.

At present, the researches of scholars on Microblog network both at home and abroad mainly focus on the following two aspects. One is about the evolution of the network topology, such as the degree distribution, cluster coefficient, vertex degree correlations and other network properties [1-8]. For example, literature [1] collected data from four different social networks, namely Flickr, YouTube, LiveJournal, and Orkut, and then comparatively analyzed the topology features of these four kinds of social networks, and verified the characteristics of power rate, small-world effect and scale-free property of social networks. Then on the basis of [1], literature [2] analyzed the network topology distribution, temporal evolution distribution, and the information dissemination process of the pictures on Flickr. Literature [3] used the technology of crawling to collect more accurate data of Twitter than

previous studies, and discussed the network topology, user sorting method, transmission mode of hot topics and so on to achieve the quantitative analysis on the Twitter network and its information distribution.

And the other one is about the user interaction rules, most of which are still confined to qualitative or semi-quantitative analysis on the statistical nature, and lack of analysis on the theory and underlying mechanism [9-15]. For example, literature [9] analyzed the potential friendships on Twitter, and found that Twitter network is composed of network of mutual following with a large density and network of real friends with a small density. Literature [10] studied the role of social network of Twitter in information dissemination from the perspective of user behaviors. Using the ideas of social network analysis, literature [11] took the "Myspace 9911" Microblog community as an empirical analysis example, and discussed the network structure of user communication in Microblog community through the core-edge analysis, the overall network analysis, centrality analysis and subgroup analysis. And literature [12] explored the different structural characteristics of social networks and their relationships during the information exchange process of users through the center analysis and cohesion subgroup analysis.

However, the information dissemination in real Microblog network is a typical evolution process of complex network systems, and we need to take both of the users' interactions and network structure into account at the same time in order to describe the information dissemination characteristics in Microblog network more accurately. So based on this situation, this paper intends to use the theory of ecology and complex network to analyze the topology of Microblog network, and then researches the user behavior and the dissemination characteristics of the information in Microblog network. It aims to put forward some related countermeasures to the government departments to control and guide the healthy and rapid information dissemination in Microblog network.

The organization of this paper is as follows. In Section 2 we analyze the topology structure, ecological groups, and the inter-group relationships between them in Microblog network. Then, in Section 3, take the Sina Microblog network as an example, we model the characteristics of Microblog network from the perspective of user nodes, group structure and information dissemination. And finally in Section 4 we propose some related countermeasures to the government departments to control and guide the healthy and rapid information dissemination in Microblog network.

2. Structure Analysis of Microblog Network

2.1. Topology of Microblog Network

The information dissemination in Microblog network is not only different from the linear dissemination of traditional media (One to One), but also different from the mesh dissemination of network media (One to N), it is a new kind of dissemination with net nuclear shape and fission spread (One to N to N). This mode of dissemination can achieve geometric spreading of information in a very short time.

Define the users in Microblog network as the nodes, then the relationships between users can be abstractly represented by edges between nodes. Each node is not only a message sender, but also the recipient of information, and the information is only disseminated along the edges. In Microblog network, the information published by a node can be seen by the nodes that follow it, and commented, forwarded, shared, etc. with a certain probability, thus forming a node-based information network structure, as shown in Figure 1 [16].

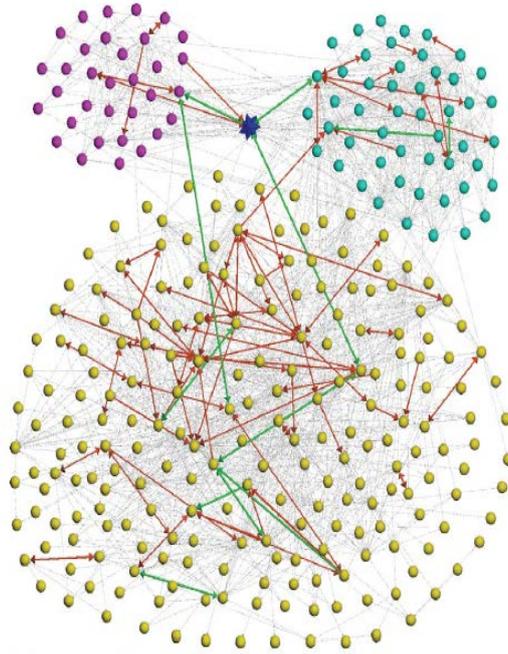


Figure 1. Topology of Microblog Network

2.2. Ecological Groups of Microblog Network

The information disseminated in Microblog network spreads in a wide range through forwarding and comments among different user nodes. When the topic released by the core nodes raises following from other users and also generates interactive behavior, then there will be interactions between groups on the platform of Microblog network, and thus form the group structure that takes the topic as the core. The only difference is the size of group and the interaction strength. Almost every topic in Microblog network can make users to form groups. The higher the density of groups is, the more topics that users are interested in. For example, there are three different sizes of groups in the topology of Microblog network in Figure 1, and the scale of the third group below is relatively larger, its density is higher and interaction is more intensive.

The groups of Microblog network can be divided into the following three levels from micro to macro, namely individual, population and community, as shown in Figure 2. The individuals refer to the subjects that are inseparable and can complete certain Microblog behavior independently in the activities of Microblog network including the natural individual Internet users and groups of individual namely virtual enterprises (or other organizations). And all the individual Internet users and the individual virtual enterprises that influence with each other and function as a whole to the outside world within a certain scope of Microblog network are called the population of Microblog network. They have same or similar ecological characteristics. The community of Microblog network refers to the collection with certain structure and functions composed by a variety of different populations of Microblog network in a certain Microblog network environment, such as all the users in the Sina Microblog network [17].

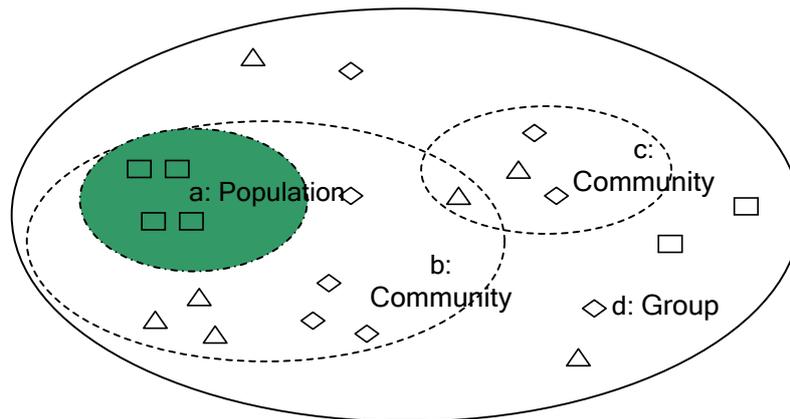


Figure 2. The Logical Structure of Ecological Groups in Microblog Network

2.3. Inter-group Relationships in Microblog Network

The groups of Microblog network are not closed. However, they have overlapping features, and are interconnected with intermediate nodes with each other. This kind of dissemination pattern can make the information disseminate and spread in a wider range on the Microblog platform, and it specific refers to that different groups have a certain proportion of common users. For example, the users who follow about the “Beijing rainstorm incident” also follow about the “Meimei Guo event”, so such users belong to two groups at the same time. These users are also links between different groups in addition to belonging to their respective groups, so the users of different groups can share information. In general, the links between Microblog groups can be divided into the following four types, namely internal links, between groups, intermediary and no-group links, as shown in Figure 3.

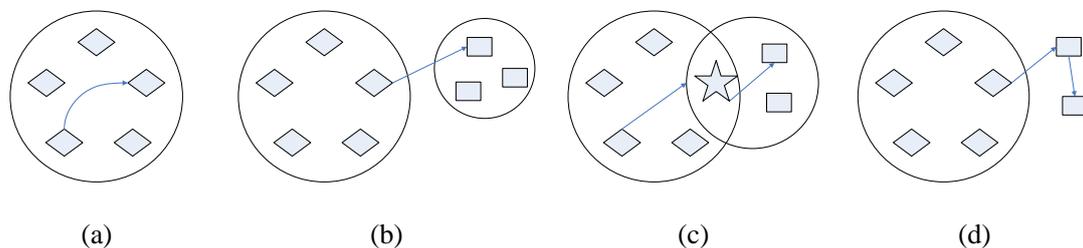


Figure 3. The Inter-group Relationships in Microblog Network

These inter-group relationships in Microblog network are clearly related to the activity in the network in terms of the information called comments and forwarding. Comments, which are supposed to be more personal information, tend to concentrate inside the groups or on links connecting close groups. This effect is stronger the larger the number of mentions exchanged and if they are reciprocated. Forwarding, which are associated to information dissemination events, appear with higher probability in links between groups, especially those that connect groups that do not show a high overlap, and more importantly on links connected to users who intermediate between groups. These intermediary users belong to multiple groups and play an important role in the spreading of information. They acquire information in one group and launch forwarding targeting the other groups of which they are members. At the same time, the access to new information can transform them into attractive targets to be forwarded by their followers.

3. Characteristics Analysis of Microblog Network-Take the Sina Microblog Network as an Example

3.1. Characteristics of the User Nodes

1. In and out degree distribution

The relationship of follow among Microblog users is directive, so the Microblog network is a directed graph, and it can be represented by in and out degree distribution, including the distribution of in degree ($p(k^{in})$) and the distribution of out degree ($p(k^{out})$). The nodes' in degree indicates how many people follow a certain user, or the equivalent, how many people received the user's information. On the other hand, the out degree means the number of people that a certain user follows, which indicates from how many people it has received information. The degree distribution is a distribution that can describe the importance of the network node. If the in degree is large, then it indicates that there are more users it follows, and it has more abundant sources of information, while if the out degree is large, then it indicates that there are more users following it, and it has greater influence [18].

In order to eliminate the impact of the network size, we use the cumulative distribution function to represent the degree distribution:

$$p_c(k) = \sum_{k' \leq k} p(k')$$

Take the Sina Microblog network as an example. The double logarithmic cumulative distribution plot of in and out degree of Sina Microblog network is shown in Figure 4.

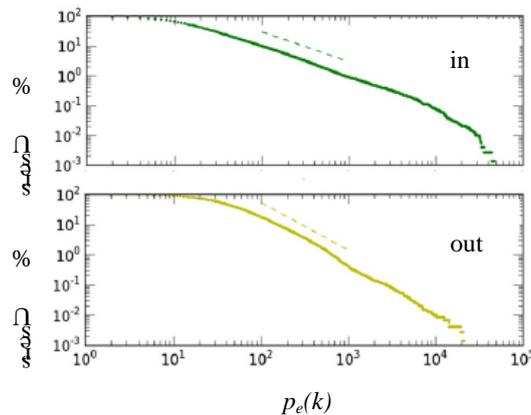


Figure 4. The in and out Degree Distribution of Sina Microblog Network

As can be seen from Figure 4, there is a big heterogeneity between the degree distributions of nodes in Sina Microblog network. Both of the distributions of in and out degree obey the exponential power-law distribution after removing the head, and they have a significant scale-free property. This indicates that they are very different in the information inflow and outflow. The data shows that the following number that the top 20% users with a high degree of following have accounts for 68% of the total, while the following number that the first 20% users who follow others more have accounts for 66% of the total. Although it not strictly complies the ‘‘Pareto’s Law’’, we can still see from their values that there are only a small part of important nodes in Sina Microblog network.

2. User Types

According to the different in and out degree distribution of the user, we can classify the users of Microblog network into the following three types: information producers, active information consumers and passive information consumers [19].

The information producers are the widely followed users who gain an enormous amount of retransmissions, whereas they have low activity. These users do not tend to follow a lot of people, nor retransmit many messages.

The active information consumers are users with high reciprocity in relations. They tend to gain as much audience and retransmission rate, as the amount of activity employed. They are very important in the information dissemination process, because they boost the content and serve as the propagators of the information producers.

The passive information consumers are the largest group of users who practically does not participate in the dissemination process. They consume more information than what they produce. They are characterized for having low activity rate, not retransmitting many messages and receiving messages from many more people than their audiences.

3.2. Structure Characteristics of the Groups

1. Cluster

Since the Microblog network has obvious group structure, so we can use cluster coefficient to describe the degree of clustering of Microblog network [20].

Define the cluster coefficient of undirected and unweighted Microblog network nodes is:

$$C_i = \frac{2l_i}{k_i(k_i-1)}$$

Here, l_i represents the actual number of edges between node i and the nodes that it connects, k_i represents the degree of node i . The cluster coefficient of the Microblog network is the average of cluster coefficient of all nodes.

Take the Sina Microblog network as an example. We can process the Sina Microblog network into an undirected graph when considering the cluster coefficients, so then we can describe the relationship between node groups. After calculation, we can get that the cluster coefficient of Sina Microblog network is 0.06826, which is larger than the cluster coefficient 4.026×10^{-5} of the same scale random network with several orders of magnitude, showing that the Sina Microblog network performance good characteristic of cluster.

2. Centrality

Centrality is an important concept that analyzes the group structure of Microblog network. It is about the central position measurement of a node in Microblog network, which reflects the position differences of various nodes in the group structure of Microblog network [21].

Define the centrality of Microblog network is:

$$C'_i = \frac{2 \sum_{i=1}^n (C_{iMax} - C'_i)}{(N-1)^2 (N-2)}$$

Here, C_{BMax} represents the maximum spacing of centrality.

Take the Sina Microblog network as an example. By calculating the spacing degree of centrality, it can reflect that there are differences in the position of various nodes in the group structure of Sina Microblog network, that is to say some users are in the heart of group while some users are in the edge of group. And this specific represents that the users of Sina Microblog network form groups that take some users as the core, and the users' audiences as the intermediary nodes and edge nodes around a particular event. Such groups can be described using the diagram form of satellite structure, namely there are a lot of sub-graphs with satellite structure in Sina Microblog network, and every sub-graph shows a partial group. The users within the group show a high degree of cohesion around the core users.

3. Similarity

The center tightness of the Microblog network reflects that the users are in the same group due to the similarities between them. During the dissemination of information, the impact of individuals within a group on the whole network is similar, so the nodes within a group are always given priority to pass this information on that the individuals disseminate, and then followed by the individuals in other group accessing the information that the nodes send. Thus, we can divide the individuals with more similarity into the same group by measuring the similarity of individuals in Microblog network, and otherwise.

Define the center tightness of the group structure of Microblog network is:

$$C_i^n = \frac{\sum_{i=1}^n (C_{iMax}^n - C_i)(2N - 3)}{(N - 1)^2 (N - 2)}$$

Here, C_{iMax}^n represents the maximum relative center tightness.

Take the Sina Microblog network as an example. The users in Sina Microblog network can be divided into 13 categories in accordance with the nature of industry, as shown in Table 1, and then we can subdivide the group into small populations again according to this category.

Table 1. Group Categories of Sina Microblog Network

1	2	3	4	5	6	7
Entertainment	Fashion	Finance and Economics	Sports	Science and Technology	Media	Art
8	9	10	11	12	13	
Education	Government	Literature	Car	Game	Estate	

3.3. Dissemination Characteristics of the Information

1. Evolution of the Number of Users

Figure 5 shows the cumulative curve of the number of users who followed about the incident in Sina Microblog network after the “Beijing rainstorm incident” occurred. We can see from the curve that the number of users steadily increased in the early event, and then maintain stable later. There is also a sharp growth of the number of users in Sina Microblog network during the emergence of “Beijing rainstorm incident”, indicating that there is a special event that causing great following of the users.

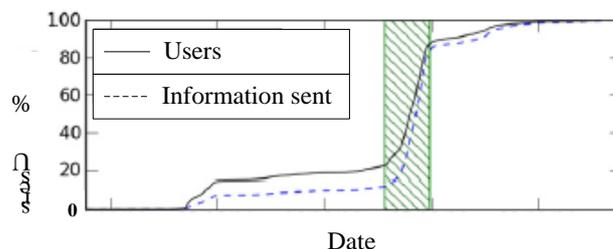


Figure 5. The Cumulative Curve of the Users who followed about the "Beijing Rainstorm Incident" in Sina Microblog Network

2. Distributions of the Information Comments and Forwarding

We can know the information dissemination characteristics by analyzing the distributions of the information comments and forwarding. Figure 6 shows the distributions of the information comments and forwarding in Sina Microblog network after the “Beijing

rainstorm incident” occurred, which has obvious characteristics of fat tail. The overall trend of the distribution meets the power-law distribution, and its power index is 1.5612. It means that only a minority of information can get a lot of comments and forwarding, and this conclusion of analysis is also consistent with the observation of people that the authoritative information will get a lot of forwarding by users, and some other public opinions will also get lot of comments from their audiences after the emergency occurs, while the information of most of the users rarely gets comments and forwarding since this type of information is the information of expressing emotion and opinion and it does not provide much value.

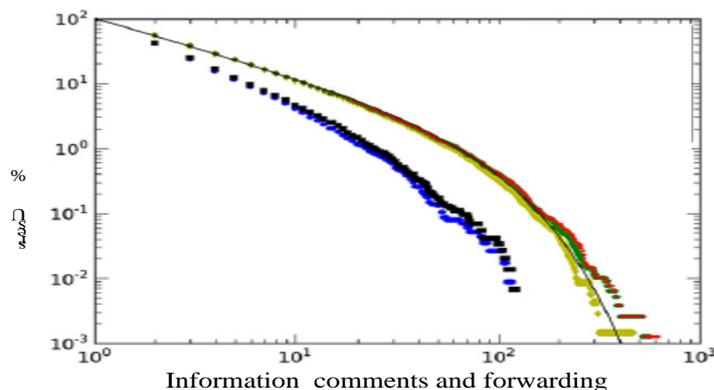


Figure 6. Distributions of Information Comments and Forwarding about the “Beijing Rainstorm Incident” in Sina Microblog Network

3. Efficiency and Vulnerability of Information Dissemination

The average path length and diameter of network are closely related with the characteristics of connectivity, reachability and dissemination delay, etc., so they can be used to measure the information dissemination efficiency of the Microblog network. The average path length of network refers to the average of the shortest path between all nodes in the network, while the diameter of network refers to the maximum value of the shortest path between any nodes in the network [22].

Define the average path length of Microblog network is:

$$l = \frac{1}{n(n-1)} \sum_{i \neq j} d_{i,j} (i, j = 1, 2, \dots, n)$$

Here, $d_{i,j}$ represents the shortest path between node i and node j .

Take the Sina Microblog network as an example. Its network diameter is approximately 12.0 and average path length is approximately 4.0, which means that the information between each node can reach each other after an average of 4 steps or maximum 12 steps up. It is consistent with the same scale random network whose average path length is approximately 3.88, indicating that the connectivity and reachability of Sina Microblog network is very good, and the dissemination delay is short, so it is very conducive to the dissemination of information.

4. The Information Dissemination Control of Microblog Network

4.1. Strengthen the Government's Authoritative Coverage and Rapid Response Capability

Open, fast and smooth information channel is very necessary to guide the information of Microblog network. The government should release press for the first time as far as possible to win the right to speak, minimize or avoid public speculation and inaccurate news media reports, and thus grasp the initiative in public opinion. Once an unexpected event occur, it

usually takes 2 or 3 hours to appear on the Internet, 6 hours later may be reproduced on other websites, and after 24 hours the online discussion will reach climax. If the positive sound can't keep up, then the negative information will take advantage. In the management of crisis and issues, the government should respond and launch emergency plans within 12 hours after the incident occurs. The government should strengthen its emergency authoritative coverage and rapid response capability, strive to head start, release accurate authoritative information, and thus play a guiding role as a main channel of online media.

4.2. Make the Opinion Leaders Play an Active Role in Guiding Correctly

As early as the 1940s, Lazarsfeld *et al.*, have proposed the concept of "opinion leader" in "People's Choice" [23]. The so-called opinion leader refers to the "activist" who often provides information for others and also influences others in that interpersonal communication network, and plays an important role in brokering or filtration during the formation process of mass media effect. Generally speaking, the number of most of the audiences of Microblog users is typically tens to hundreds of magnitude, while the number of audiences of a small number of users is over thousands or even tens of thousands, called star user (opinion leader). Therefore, it has important significance to monitor and guide the remarks of opinion leaders in Microblog network for information dissemination management.

5. Conclusions

The area, speed and efficiency of information dissemination have been greatly improved in Microblog, so the research on the information dissemination law in Microblog has a very important theoretical and practical significance to control and guide public opinion. Using the theory and method of ecology and complex network, this paper analyzed the network topology, ecological groups and network characteristics of the Microblog network, and then proposed some related countermeasures to the government departments in order to control and guide the healthy and rapid dissemination of information in Microblog network. The research shows that the Microblog network renders mesh topology with good characteristics of cluster, centrality and similarity. The efficiency of information dissemination in Microblog network is relatively very high and the network is not fragile. Meanwhile, the node's role in the dissemination of information varies greatly, and it is easy to form opinion leader. It has important significance to strengthen the government's authoritative coverage and rapid response capability, make the opinion leaders play an active role in guiding correctly for the information dissemination management in the Microblog network. However, there are still many other uncertain factors during the information dissemination process in Microblog network, so there is still much room for improvement to this research.

Acknowledgements

This work is partially supported by the General Research Project (18190113002) and the Quality Improvement Project of Personnel Training (03150113016) of Beijing Municipal Education Commission, and also the Institute Level Project (E-b-2012-20) and the Course Construction Project (22150112088) of Beijing Institute of Graphic Communication.

References

- [1] A. Mislove, M. Marcon and K. P. Gummadi, "Measurement and analysis of online social networks", Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement. New York: ACM, (2007), pp. 29-42.
- [2] M. Cha, A. Mislove and K. P. Gummadi, "A measurement-driven analysis of information propagation in Flickr social network", Proceedings of the 18th International Conference on World Wide Web. New York: ACM, (2009), pp. 721-730.
- [3] H. Kwak, C. Lee and H. Park, "What is twitter, a social network or a news media", Proceedings of the 19th International Conference on World Wide Web. New York: ACM, (2010), pp. 591-600.
- [4] J. Goldenberg, B. Libai and S. Moldovan, "The NPV of Bad News", International Journal of Research in Marketing, vol. 24, no. 03, (2007), pp. 186-200.

- [5] A. Java, S. Xiaodan and T. Finin, "Why we twitter: Understanding microblogging usage and communication", Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis. New York: ACM, (2007), pp. 56-65.
- [6] L. Ping and L. Zong, "Research on Microblog Information Dissemination Based on SNA Centrality Analysis—A Case Study with Sina Microblog", Documentation, Information and Knowledge, vol. 6, (2010), pp. 92-97.
- [7] Y. Jaewon and J. Leskovec, "Modeling Information Diffusion in Implicit Networks", IEEE. 2010 IEEE 10th International Conference on Data Mining. Berlin, Germany: IEEE, (2010), pp. 599-608.
- [8] M. Lahiri and M. Cebrian, "The Genetic Algorithm as a General Diffusion Model for Social Networks", The 24th AAAI Conference on Artificial Intelligence. Atlanta, Georgia, USA: [s.n.], (2010), pp. 76-81.
- [9] B. A. Huberman, D. M. Romero and W. Fang, "Social networks that matter: twitter under microscope", First Monday, vol. 14, no. 1, (2009), pp. 1-5.
- [10] L. Kristina and G. Rumi, "Information Contagion: an Empirical Study of the Spread of News on Digg and Twitter Social Networks", Venue: in ICWSM, (2010).
- [11] X. Wang and S. Teng, "Empirical Analysis on Informal Communication in Microblog Community —Taking "Myspace 9911 microblog" for an Example", Library and Information Service, vol. 55, no. 4, (2011), pp. 39-43.
- [12] Y. Yuan and C. Yang, "Empirical Analysis of All Kinds of Social Networks and Their Relationships Formed by Information Communication among Microblog Users", Library and Information Service, vol. 55, no. 22, (2011), pp. 31-35.
- [13] K. Saito, M. Kimura and K. Ohara, "Learning Continuous Time Information Diffusion Model for Social Behavioral Data Analysis", Advances in Machine Learning, Springer, vol. 5828, (2009), pp. 322-337.
- [14] H. Haddadi, BENEVENUTO F, GUMMADI K P. Measuring User Influence in Twitter: The Million Follower Fallacy[C]. Proceedings of International AAAI Conference on Weblogs and Social. Washington, DC: ICWSM, (2010), pp. 97-105.
- [15] Y. Hui and J. Wu, "Complementary Structure of Network Modeling and Analysis of Statistical Characteristics", Communications Technology, vol. 44, no. 05, (2011), pp. 94-96.
- [16] P. A. Grabowicz, J. J. Ramasco, E. Moro, J. M. Pujol and V. M. Eguiluz, "Social Features of Online Networks: The Strength of Intermediary Ties in Online Social Media", PLoS ONE, vol. 7, no. 1, (2012) January, pp. 1-9.
- [17] Z. Zhang and R. Zhang, "Ecology of Cyber-society", Publishing House of Electronics Industry, (2008) July.
- [18] L. Zheng and S. Li, "A Novel Information Diffusion Model based on Microblog Network", Communications Technology, vol. 45, no. 02, (2012), pp. 39-41.
- [19] A. J. Morales, J. C. Losada and R. M. Benito, "Users structure and behavior on an online social network during a political protest", Physica A, vol. 391, (2012), pp. 5244-5253.
- [20] Z. Tian and Y. Sui, "The Empirical Analysis of Micro-blog Information Flow Based on Complex Network Theory", Library and Information Service, vol. 56, no. 8, pp. 42-46.
- [21] W. Lv, "Research on Group Structure of Micro-Blogging Network", Beijing Jiaotong University, (2011) December.
- [22] P. Fan, H. Wang, Z. Jiang and P. Li, "Measurement of Microblogging Network", Journal of Computer Research and Development, vol. 49, no. 4, (2012), pp. 691-699.
- [23] Z. Liu and L. Liu, "Recognition and Analysis of Opinion Leaders in Microblog Public Opinions", Systems Engineering, vol. 29, no. 6, (2011) June, pp. 8-16.