Journal of Computational Science & Engineering 5 (2013) 208-214

Journal of Computational Science & Engineering

Available at http://www.asocse.org ISSN 1710-4068

Microblog Social Network Analysis Based on Network Group Behavior

Lin Wang [a], Yang Zhao [a], Shehong Liang [b] and Kan Shi [b]*

[a] *Northeastern University at Qinghuangdao, Qinghuangdao, Hebei, China*
[b] *Graduate School, Chinese Academy of Sciences, Beijing, China*

Article Information

Article history:
Received 4 February 2013
Revised 10 February 2013
Accepted 20 February 2013
Available online 10 March 2013

Keywords:
Campus network
Network group incidents
Microblog
Social network
Data acquisition
Power

Abstract

This paper has searched the six key influential network group incidents in 2011 in China through SINA Microblog platform, screen the most active Microblog users who pay close attention to the incidents, their Microblog followers are above 50,000 , finally collects 30 Microblog user information by the way of snowballing, form 30*30 "follow – be followed " Microblog netizens relationship matrix. Specific to this asymmetric matrix, the authors use UCINET social network analysis software and visual software NETDRAW, analyze Microblog network structure, network centrality based on network group incidents, and comprehensively describe the laws to release and acquire group incidents under the mixed network environment of Microblog, combined with the massive Microblog exchange data, to analyze the non-rational behavior of the netizens, so as to provide a scientific basis for setting out the appropriate guiding strategy and interventions.

1. Introduction

Microblog (Microblogging) is an informal mini-blog. The user can release the text information and picture, video and other multimedia contents within 140 characters on their individual microblog

* Corresponding author: shik@gucas.ac.cn

2013 © The American Computational Science Society. All rights reserved.

through cellphone, instant messaging tool (including QQ, MSN, SKYPE, Fetion etc.), Microblog, light blogging and other platforms, show the latest information of the individual. Microblog is widely accepted due to convenient information release, take less time and energy.

Evan Williams was the first to put forward the idea of Microblog and founded the famous Microblog website-Twitter. Twitter's users are widely distributed, including the current President of the US Obama, the White House, Google, Forbes, GM and many international celebrities and organizations. The booming of Microblog in overseas countries has aroused the concern of the domestic Internet industry. In the first half year of 2007, JIWAI, FANFOU have become the China's first microblog platform. In August of 2009, the closed beta version of SINA Microblog has become the first portal website to provide microblog service. Microblog officially entered the view of the mainstream crowds who surfing the Internet in Chinese. In 2010, the people Microblog (t.people.com.cn), SOHU Microblog (t.sohu.com), Netease microblog (blog.163.com), Tencent microblog (t.qq.com) and others begin the close beta test. The microblog users cover individual netizens, enterprises, organizations and governments, such as Alliance (Zhaopin.com), New Oriental, China Writers' Association, CCTV's Today Observation, Information Office of Yunnan Provincial Government have "settle down" on SINA Microblog, getting up close interaction with the netizens.

Domestic and international scholars make full research on microblog. Through analysis, the author found that the researchers have made a comprehensive analysis on the characteristics of Microblog, compared to traditional blogging, Microblog allow users to exchange small text, with faster transmission speed, and the advantages of instant, interaction and freedom, more convenient to the peer-to-peer interpersonal communication, with the character of We Media. Meanwhile, the microblog has defects in information authenticity and deep transmission of ideology and other aspects. For the prospects of microblog, most of them are optimistic and think that it can blends with the traditional media to develop.

2. Internal relation between Microblog and network group incidents

With the rapid development of network society, the western scholars probe into and study the network group incidents and its political and social impact from different perspectives. They mainly discuss the following issues: (1) how to define the concept of network conflict and what types. The Professor Athina Karatzogianni from the University of Nottingham believed that, network conflict refers the concentrated expression of the conflicts existed in physical space on network. From the perspective of political sociology, network conflict is the network contests that affect or challenge the public opinion and realize some interest appeal. (2) What political and social functions the network conflict has. Peter Levine believes that the disputes and conflicts of cyberspace contributes to active civil society and wakeup public spirit, and is a powerful democratic resource (Peter Levine, 2001); Neil Weinstock Netanel warned that , the network conflict in a certain scope contributes to the generation of network corporatism and social progress, but over-expanded network conflict has the risk of "majority dictatorship", even form "network anarchism" (Neil Weinstock Netanel, 2000) . However, some scholars believe, whether the network is the new growth point of democracy or a new instrument of oppression, totally depends on different political and social context (Michael L. Best and Keegan W. wade, 2007 ; Norris, 2000 ; Lawrence lessig, 1999 ; Tyler Boas, 2003). (3) What challenges does the network

2013 © The American Computational Science Society. All rights reserved.

conflict bring to the traditional political structure? Victoria Bemal thought the network has created a new public domain, challenging and changing the traditional political participation (Victoria Bemal, 2006); Michael Whine pointed that if not guide and control the network conflict, it will likely to be the new tool for political extremists ((Michael Whine, 1999). The scholars in China have just begun the research on network group incidents, mainly focus on transmission mechanisms, influencing factor and generation patterns.

3. Questions, Research contents and objectives

3.1 Question

1. Why the group incidents have very large attention and influence on microblog? Is there any clue features in the group behaviors on microblog?

2. Why the microblog users who have massive fans affect the wider netizens, how their social network structure affects the group behavior?

3.2 Research contents and objectives

Based on research results of predecessors and current problems of Microblog, the author makes the exploration and research on microblog social network analysis based on network group incidents.

The research is based on social network theory and technology acceptance model theory, combined with the enormous influence of Microblog and the frequency of network group incidents, explore the clues features and exchange mechanism, transmission characteristics of the Microblog posts be forwarded in a large number and leave comments and crowded aroused by typical network group incidents, aim at finding the clues to respond network group behavior, and work out the targeted strategy.

4. Research Process and Methods

4.1 Sample Selection

In this research, firstly, we have selected six well-informed network group incidents in the past year; secondly, search the Event 1-6 on SINA Microblog, the search result of Event 1: 10,462 Microbloggers follow this event on their microblogs; the search result of Event 2: 370,448 Microbloggers follow this event; the search result of Event 3: 133,577 follow; the search result of Event 4: 75,039 Microbloggers follow this event; the search result of Event 5: 688,469 Microbloggers follow this event; the search result of Event 6: 1,707,760 Microbloggers follow this event; Thirdly, select the microblogger with the most comments or forward in the search result, and the fans of the Microblogger is above 50,000, through the snowballing acquisition method, that is after collect the information of this Microblogger, collect the follow microblog users information of him, and still require that the fans of this Microblogger is above 50,000, the basic information including ID, name, sex, occupation, attention, fans, microblog posts, exchange content, communication channel etc.; finalize 30 Microbloggers as the research object, prepare for the information acquisition.

4.2 Data Acquisition

1. The data acquisition software in this research is the more mature LocoySPider in China, and select SINA Microblog platform for data acquisition. During the acquisition process, the author should read the source code of Microblog, compile acquisition rules, after repeated debugging without error, and then do formally acquisition. The acquisition data is from 08:00AM on Sep. 10 to 15:00PM on Sep. 11 of 2011, totally collects more than 1,500 Microblog exchange data, through data cleaning, removing the unrelated

2013 © The American Computational Science Society. All rights reserved.

acquisition content and empty acquisition content, and forms the initial original database.

2. On this basis, first duplicate the original database, and then carry data extraction, filtering out the basic information database of 30 Microbloggers.

3. According to the Microbloggers user information in acquisition database, the author adds these 30 Microbloggers on SINA Microblog, builds 30*30 matrix. It is an asymmetric matrix, because each follow of Microblogger is one-way, User A follows User B, but User B may not follow User A. Therefore, compare the Microblogger one by one; whether the Microblogger and I share the common follow Microblogger, through the above approach, to build this network relational matrix, to analyze the latter social network connection strength and network centrality analysis, it is the key matrix.

4.3 Descriptive Statistics

The research has sampled 30 Microbloggers in total, 18 males and 10 females, the occupation distributes senior executives, media reporter, anchor, actor, editor in chief, writer, Ph.D., commentator, and common staff etc., 67% Microbloggers are SINA's certified members, the sample has a certain influence as a whole.

4.4 Social Network Analysis

In SINA Microblog, the transmission and spread of network group incidents are more influenced by those users who have massive fans/followers and more active in comments or forward, they received the attention of more people, so they play a more important role in the transmission of information and its network group incident. The Microblog users listed in this article have above 50,000 followers. Because the user group of SINA Microblog is very large(millions of users), the following group of users is only a sub-network of the whole Microblog social network, nested in the whole network, but because of the sampling is the most active people who follow the network group incidents among the users , they have a considerable extent of radiation. This article describes this small social network, and discusses the transmission characteristics of network group incidents on the whole Microblog social network.

	OutDegree	InDegree	NrmOutDeg	NrmInDeg
11 U011	12.000	6.000	41.379	20.690
14 U014	11.000	4.000	37.931	13.793
27 U027	10.000	6.000	34.483	20.690
15 U015	9.000	5.000	31.034	17.241
21 U021	8.000	6.000	27.586	20.690
25 U025	7.000	8.000	24.138	27.586
20 U020	7.000	0.000	24.138	0.000
29 U029	6.000	1.000	20.690	3.448
22 U022	6.000	4.000	20.690	13.793
3 U003	6.000	11.000	20.690	37.931
8 U008	5.000	1.000	17.241	3.448
13 U013	5.000	1.000	17.241	3.448
7 U007	5.000	14.000	17.241	48.276
9 U009	4.000	4.000	13.793	13.793
10 U010	4.000	5.000	13.793	17.241
23 U023	4.000	12.000	13.793	41.379
30 U030	4.000	3.000	13.793	10.345
24 U024	3.000	5.000	10.345	17.241
4 U004	3.000	2.000	10.345	6.897
2 U002	3.000	16.000	10.345	55.172
28 U028	3.000	2.000	10.345	6.897
1 U001	3.000	1.000	10.345	3.448
18 U018	2.000	4.000	6.897	13.793
17 U017	2.000	1.000	6.897	3.448
6 U006	2.000	4.000	6.897	13.793
5 U005	2.000	6.000	6.897	20.690
19 U019	1.000	1.000	3.448	3.448
12 U012	0.000	1.000	0.000	3.448
26 U026	0.000	3.000	0.000	10.345
16 U016	0.000	0.000	0.000	0.000

1) *Microblog social network "follow" matrix and its Inter-relationship diagraph*

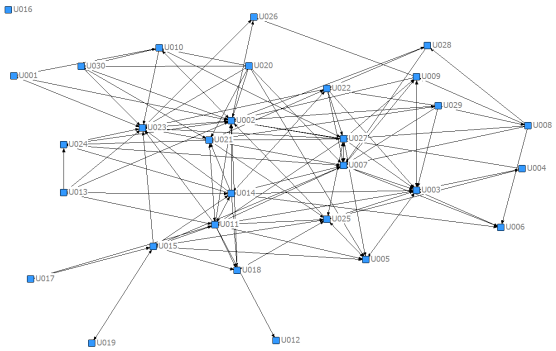
After a comparative analysis of the collected data records, use an adjacency matrix to show the relation between "follow" and "be followed" among the users. In this matrix, the row is the microblogger to "follow", and the list is the Microbloggers be followed. "1" means the relation of "follow" exists, "0" means that it does not exist.

Through the drawing function of UCINET and NETDRAW, the network relationship graph can be expressed intuitively, see Graph 1.

According to "Follow matrix", we can judge that the social network relationship graph of this group of users is digraph (directed graph). In this digraph, each node represents a user, the connection among them shows the relation of "follow "and "be

2013 © The American Computational Science Society. All rights reserved.

followed". The arrow from Node A direct to Node B, shows A "follow" B. the number of arrows from one node to other nodes is outdegree, and the number of arrows of one node directed by other nodes is indegree. Since these nodes exists the directed relation, that is, if A "follows" B, the information released by B will show on the homepage of A quickly. Among the Microblog social network, the transmission direction of information is opposite to the direction of "follow".



Graph 1: Active Microbloggers Social Network Relationship Graph based on network group incidents

2) *Measures of Centralities and Central Potential*

The article combines the relation between "follow" and "be followed" among the Microbloggers, analyze the centralities and central potential index, identify those users who are at the core in this social network, that is to find out which users has greater "power" in Microblog information transmission, and can affect the transmission of group incidents in the whole social network to a large extent.

a) Degree Centrality Analysis

Input the above "follow matrix" into UCINET software, to calculate the degree centrality and central potential, the result is as follows:

It can be seen from the results, the user with higher indegree, that is User 2 who is more to be "followed" (indegree is 16), and User7 (indegree is 14). It has stated that they have greater power on the transmission of group incidents on this network; the news released by them is followed by more people.

The standard indegree and outdegree central potential of the whole network is 40.785% and 26.516% respectively. The central potential is closer to 1, the network has more centralization. It can be seen that the central potential be followed is larger, and the user be followed has more centralization. Similarly, "follow" centrality only reached 26.516%, indicating that centralization of "follow" users is not evident.

		OutDegree	InDegree	NrmOutDeg	NrmInDeg
1	Mean	4.567	4.567	15.747	15.747
2	Std Dev	3.116	4.031	10.746	13.899
3	Sum	137.000	137.000	472.414	472.414
4	Variance	9.712	16.246	115.484	193.170
5	SSQ	917.000	1113.000	10903.686	13234.245
6	MCSSQ	291.367	487.367	3464.526	5795.085
7	Euc Norm	30.282	33.362	104.421	115.040
8	Minimum	0.000	0.000	0.000	0.000
9	Maximum	12.000	16.000	41.379	55.172

Network Centralization (Outdegree) = 26.516%
Network Centralization (Indegree) = 40.785%

b) Betweenness Centrality

Input the above "attention" matrix into UCINET software, to calculate betweenness centrality and its central potential. The result is as follows:

		1 Betweenness	2 nBetweenness
27	U027	175.083	21.562
21	U021	118.380	14.579
25	U025	112.005	13.794
11	U011	111.989	13.792
15	U015	101.620	12.515
7	U007	76.160	9.379
2	U002	68.536	8.440
3	U003	56.405	6.946
10	U010	47.810	5.888
22	U022	41.593	5.122
14	U014	41.298	5.086
23	U023	38.248	4.710
30	U030	26.033	3.206
29	U029	25.510	3.142
9	U009	8.136	1.002
24	U024	4.426	0.545
5	U005	3.233	0.398
28	U028	2.500	0.308
8	U008	1.500	0.185
18	U018	1.386	0.171
4	U004	1.376	0.169
6	U006	0.450	0.055
13	U013	0.325	0.040
20	U020	0.000	0.000
17	U017	0.000	0.000
26	U026	0.000	0.000
12	U012	0.000	0.000
19	U019	0.000	0.000
1	U001	0.000	0.000
16	U016	0.000	0.000

Network Centralization Index = 17.79%

2013 © The American Computational Science Society. All rights reserved.

The betweenness centrality is measure the role of Node C on the information flow or control of tranmission in the whole network, that is, if the information transmit from Node A to Node B, on which extent to depend on Node C. we can clearly see from the analysis results, the betweenness centrality of User 27, 21, 25, 11, 15, 7, 2 and 3 is relatively high. In other words, other users want to get the further information of network group incidents, they are highly depend on the above users, so these users has larger power in this network, also can control the flow of related information of group incidents to a greater extent. The betweenness central potential of the whole network is 0.1779, not very high. For the whole network, the betweenness central potential is not very large, most nodes in the whole network don't need the other nodes as bridging, can get the information.

c) Closeness Centrality Measures

Input the above "attention" matrix into UCINET software, to calculate closeness centrality and its central potential.

The result is as follows:

	1 inFarness	2 outFarness
26 U026	115.000	870.000
2 U002	128.000	144.000
7 U007	131.000	126.000
3 U003	135.000	132.000
23 U023	136.000	128.000
27 U027	138.000	108.000
25 U025	138.000	117.000
5 U005	141.000	138.000
12 U012	143.000	870.000
10 U010	144.000	127.000
24 U024	145.000	143.000
9 U009	145.000	136.000
21 U021	146.000	109.000
11 U011	146.000	105.000
6 U006	148.000	144.000
30 U030	150.000	126.000
15 U015	153.000	113.000
4 U004	154.000	133.000
18 U018	154.000	138.000
14 U014	158.000	105.000
22 U022	161.000	117.000
1 U001	169.000	148.000
13 U013	171.000	118.000
29 U029	171.000	117.000
17 U017	178.000	128.000
19 U019	178.000	139.000
28 U028	184.000	132.000
8 U008	194.000	122.000
20 U020	870.000	91.000
16 U016	870.000	870.000

Closeness Centrality Measures is the degree of an actor from other's control, on the contrary of the above two centralities, the value is smaller, indicating the node is in the core position. According to the meaning of closeness centrality, when the value is smaller, indicating the distance sum between this node and other nodes is smaller; the node is closer to the other nodes, and not easy to be controlled by other nodes when getting the information. From the analysis results, the node is more at the front, its centrality is higher. For example, User 26, the distance sum that the released information transmits to all other nodes is only 115. It is relatively difficult to get the information of other nodes, the distance sum is 870. As for the difficulty degree that release the information to other nodes , the order of top 10 in all users is 26, 2, 7, 3, 23, 27, 25, 5, 12, 10, the more front , less easily controlled by the others , and more independent. The difficulty degree to get information , the order is 20, 11(14), 27 , 21 , 15, the front is easy to access the information, not easily controlled by the others, and independent. In addition, the research found that although the fans of No.16 microblogger is over 50,000, the distance sum that he get and release information is very large, with worst independence. It can be proved in the above Graph 1.

5. Conclusion and Enlightenment

(1) When the outdegree centrality of a Microblogger is high, he can get more information of other Microbloggers, with relatively large "power" when getting the information of network group incidents. Similarly, when the indegree centrality of a Microblogger is high, he is be followed by many other Microbloggers, with large "power" to release the information of network group incidents; betweenness centrality refers to the "power" to control the flow of information, if it

2013 © The American Computational Science Society. All rights reserved.

is high , indicating he plays the role of bridging among many users, so he has large power when control other users to transmit network group incidents; closeness centrality is the "power" enables the information of network group incidents reach to each users rapidly and conveniently. When the closeness centrality of a user is high, it is stated that he is closer to any user in this social network, he can convenient or more directly (less to be controlled by the other users)to get the information of each user, so as to transmit the network group incidents to other users, less dependent on other users.

(2) The network structure of social network determines the "power" of network members, the centrality is an ideal method to describe the "power". The "power" affects the flow of network group incidents, the "power" here is not the property of social network members, but a model of interaction among the actors over the network. Through the observation of a group of Microblog users in reality, the article measures the centralities and gets the result, judge the "power" and sorting of each individual during the information transmission of network group incidents in Microblog social network. The "power" here is not the characteristics of individual or group, can be explained as the real or potential mode of interaction between the publisher and the recipient of Microblog information.

The samples in this research is smaller, can't describe a large-scale network structure, only trying to describe the transmission characteristics of Microblog network group incidents by the measures of social network centrality analysis, to analyze the relationship between "follow" and "be followed" of different users in the transmission process of network group incidents, and expected to provide a new perspective for the research on the transmission of network group incidents.

Acknowledgment

This research is supported by fund of China 973 Project No. 2010CB731406.

This research is supported by funding of Northeastern University at Qinghuangdao. Project No. XNB201313.

References

[1] P.L.K. Gross, E.M. Gross, Science, 1713 (1927) 385-389

[2] E . Garfield, JAMA, 1 (2006) 90- 93.

[3] Row lands I Journal diffusion fact or : A new approach to measuring research influence, J. A slib Proceedings, 2 (2002) 77-84.

[4] T.F. Frandsen, ASLIB Proceedings, 1 (2004) 5-11.

[5] T. Braun , W. Glanzel, A. Schubert, Scien tom etrics, 1 (2006) 169-173.

[6] J.D. Luo, M. Beijing: Social Sciences Academic Press, (2005) 156-159.

[7] S.F. Hu, Journal of Chongqing University of Posts and Telecommunications (Social Science Edition), 3 (2011) 33-36.

[8] Y.Q. Huang, Journal of Hubei University (Philosophy and Social Sciences), 5 (2010) 109-112.

2013 © The American Computational Science Society. All rights reserved.