Can Network Help Chinese Microblogs Diffuse? Analyzing 118 Networks of Reposts About Social Issues in China

King-wa Fu

Journalism and Media Studies Centre, University of Hong Kong
kwfu@hku.hk

Abstract. The use of microblog, i.e. Twitter, in politics and social issues has been broadly studied. But research into the role of microblog in China's unique socio-political context is still limited and empirical data remain inadequate. In this study, we conceptualize the process of diffusion of microblogs in China as a "social network" and deploy social network analysis to operationalize the network characteristics of such process. Using Sina Weibo's Open API, we collected 118 networks of reposts related to various social issues in China. While our findings describe the network characteristics of the samples, we reveal that decentralized network is a key contributing factor for diffusion of microblogs in China. The implication of the result is then discussed.

Keywords: China, microblog, social network analysis, information diffusion.

1 Background

Broadly speaking, social media is a newer digital communication technology designed for facilitating human interactions and social connectivity. While its formal definition remains contestable [1], social media can be generally characterized by their capacity to produce and exchange user-generated content, participate in collaborative projects, and connect with one another in a variety of social communities [2]. Real life examples of social media applications include the online encyclopedia Wikipedia, video-sharing website YouTube, content sharing sites Flickr, Instagram, and Pinterest, social networking sites Facebook and Google Plus, microblogs such as Twitter or Chinese Sina Weibo or Tencent Weibo. (Remark: Throughout the text, Weibo is referred to the company or the title of Chinese microblog service; weibo is referred to a unit of message or post of the microblog service in China)

Microblog service is a self-publishing online application that enables Internet users to communicate with each other by posting or sharing short messages, say up to 140 characters in length in Twitter. Functionally, it allows users to subscribe the posts (or "tweets" in Twitter parlance) made by other users ("friends"), to be followed by others ("followers"), to forward message to one's followers ("retweet" or "repost" throughout the text), and to create searchable "hashtags" preceded by the pound sign

G. Meiselwitz (Ed.): SCSM 2014, LNCS 8531, pp. 331–341, 2014.

(#). Microblogging is characterized by its rapid diffusion of short messages through social networks comprising clusters of interconnected friends/followers relationships, i.e., friends, friends of friends, and so on. Such human connectivity has demonstrated powerful mass communication capacity during a variety of social and political incidents [3, 4], natural disasters [5-8], general elections [9-11], and political uprisings [12, 13].

In China, mass communication is vigorously controlled by the governments. Chinese traditional news media are subject to heavy-handed government monitoring and regulation. Much essential information on the newspaper or the Internet, which are not permissible by the government, is censored [14]. As a result, the country ranks at the bottom (173 out of 179) in the international rankings of freedom of the press and speech [15]. However, even though Chinese Internet platforms remain to be regulated and monitored [16], Internet users in China seem to possess greater autonomy over what they read and say concerning public affairs than they enjoy in their engagement with the traditional media and the public sphere [17]. Chinese microbloggers have successfully made local issues become international agenda, for example the "My father is Li Gang," "Guo Meimei," and "Wenzhou high-speed train crash" incidents were widely reported in the Western media after appearing in Chinese microblogs.

According to the China Internet Network Information Center [18], the number of Chinese microbloggers reached 331 million in mid-2013, accounting for 56% of the total Internet population. The two leading Chinese microblog platforms are Sina Weibo and Tencent Weibo, each of which claims to have 500 million registered accounts [19]. Nevertheless, our study reveals an alternative picture [17]: about 60% of Sina Weibo registered accounts are inactive and never make a post; the weibo contents are unevenly distributed among users and about 5% of whom create over 80% of the overall original posts.

1.1 Diffusion of Microblogs in China

The propagation of emerging social topics or critical novel ideas on the Internet can be conceptualized as a process of diffusion of innovations [20], under which is defined as "the process by which an innovation is communicated through certain channels over time among the members of a social system" (p. 5), where an innovation is defined as "an idea, practice, or object that is perceived as new by an individual or other unit of adoption" (p. 11). The process of a successful diffusion of innovation is determined by the following factors [21]: characteristics of the innovations (e.g., whether they have public consequence), characteristics of the innovators (e.g., their personal characteristics and position in the social network), and environmental and contextual setting (e.g., political conditions and societal culture).

Chinese microblogs are known to be a powerful tool for amplifying local incidents and setting the public agenda [22]. In a controlled media environment such as China's, local journalists and foreign correspondents alike rely on popular microblogs and social media as sources of information on breaking news. The use of microblogs in China thus has clear public consequence. Microblog opinion leaders are identified

as information hubs when reposting significant messages [17], thereby serving as "innovators" in this context. The heavily regulated media environment in China is a unique political and social setting that contextualizes the process of diffusion of novel information.

1.2 A Network Perspective on Microblogs

Human communication via the Internet has long been thought as a form of social network [23]. According to Manuel Castells [24], the widespread use of information and communication technology has given rise to the formation of a basic mode of social organization, namely "network society", along with the state and the capitalist market. Yochai Benkler puts forward the notion of the "networked information economy," [25] suggesting that decentralized individual actions can be interconnected by the network technologies for cooperative and coordinated work facilitated by advances in digital communication and computing technology.

Technological advances and the growing popularity of social media applications enable online users to be situated within a set of interconnected social networks that constitute the "dominant form of social organization (p. 11)" [26]. The networked individuals within such an organization (who are known as the nodes of the network) are able to take advantage of their connections with close friends, acquaintances, and friends of friends (represented by an edge between two nodes) to extend their ability to reach beyond a densely knit group, communicate effectively with the society at large, and become ubiquitously accessible [26]. Under such conception, social networks can be multi-dimensional in essence: behavioral (tweeting, retweeting, and following), semantic (sharing topics of interest, hashtags, or media content), cognitive (value or attitude), affective (emotion or happiness), or societal (reflection of real-life relationships) and various dimensions are distinct but are interlinked. As Rainie and Wellman put it, "the lines between information, communication, and action have blurred: Networked individuals use the Internet, mobile phones, and social networks to get information at their fingertips and act on it, empowering their claims to expertise (p.14)" [27].

With such backdrop of network perspective, we extend the idea to conceptualize the interconnections among microbloggers. The notion of community network has been used to describe the formation of a network of interlinked personal Twitter accounts [28]. Studies analyzing data collected from Twitter have consistently identified the characteristics of human social network such as scale-free and power-law distribution, homophily, and small world phenomenon [29, 30]. These previous works provide empirical support for our attempt to theorize Chinese microblogs as a social network.

Conventional social networks are mainly connected by weak ties [31], where the strength of a tie is "a combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie (p.1361)." Weak ties are thus those with casual acquaintances, not very close friends, neighbors, and co-workers. Granovetter [31] shows that weak ties can actually be quite strong in the sense that a weak-tie interpersonal network is more helpful than a

closed group in diffusion of information, e.g., information on job opportunities. Weak ties can bring new and diverse information from other social circles into a strongly connected relationship cluster. A recent study demonstrates that removing weak ties from a social network reduces the effectiveness of information diffusion to a considerable degree [32]. However, another study argues that Twitter networks do not fall perfectly within the scope of weak ties since the emotional intensity and reciprocal relationships among Twitter users are weak [33].

1.3 Research Question

The use of microblog, i.e. Twitter, in politics and social issues has been broadly researched. But study on the role of microblog as well as the formation of network in the Chinese context is still limited. While empirical evidence remains inadequate, the use of the notion "social network" in the Chinese context is still debatable [2]. Another challenge in research is the ubiquity of the mechanisms by which the Chinese authorities control the flow of public information [17].

Previous studies have found the rate of information diffusion via online network is influenced by the network structure [34, 35]. In this study, we seek 1) to describe the network characteristics of the flow of microblogs in China; 2) to examine the relationships between various network characteristics of repost networks in China.

2 Method

2.1 Data Collection

We make use of Sina Weibo's Open API (http://open.weibo.com/) to access raw microblog data in China. To gather the complete set of reposts of a list of original microblogs, we obtained an exhaustive set of reposts using Sina Weibo's repost timeline API (http://open.weibo.com/wiki/Statuses/repost_timeline/en). When the identity code of an original post is given as a parameter, the repost timeline API returns the list of reposts originated from that post. Since Sina imposes a limit of 200 reposts per API call, the whole list of reposts can be obtained by multiple calls of the repost timeline API.

To study the pattern of repost network, we extracted the user names contained in the message content of the reposts. When copying a post, a reference in the form of '@XYZ', where XYZ is the displayed name of a user account, is offered by default by the system as a part of the reposting text and the entire form is preserved in the repost text if it is not deleted or modified by the user. We took advantage of this property to trace back the pattern of reposting.

We then collected 118 sets of reposts of Chinese microblogs in August 2012. The original posts were identified by keyword search in our Weiboscope database [17]. The samples covered a wide spectrum of social issues ranging from anti-corruption (keywords 贪污 or 腐败), housing demolition (拆迁 or 迁拆), misbehavior of city administrators (城管), and campaign calling for finding lost children ([失踪 or 寻人]

and 小孩). Because of political sensitivity of certain topics, some posts found in the Weiboscope database had been removed from the user timeline. However, the whole list of reposts was still accessible by using the repost timeline API even the original post no longer existed (remark: Since 2013, the updated version of Sina API does not allow access to censored reposts anymore). As the time stamp was attached with the returned data, cumulative distribution of each set of reposts was devised for the analysis. Their descriptive statistics were also calculated and were presented in table format.

2.2 Modelling Network Characteristics of Reposts

The network connectivity of microbloggers is represented by a directed social graph indicating the flow of repost messages between microbloggers, in which a link (an arrow) between two nodes signifies a microblogger forwarding another user's post. For example, If User B reposts a received message sent from User A, such relationship is represented by an arrow pointing from User A (a node) to User B (another node). In this way, microblog data are mapped into a network data structure in which a node represents a microblogger, the node attributes denote one's characteristics, e.g., sex, province, and follower count, and an edge between two nodes stands for a repost from one microblogger to another.

For instance, each individual node represents a microblogger who reposts a message, say microblogger M with a list of followers F(M)=[FM_1, FM_2,….. FM_f], where f is the total number of followers of M. The out-degree centrality of User M means the number of M's followers who eventually repost the message after receiving the message from M. Like in Figure 1, if M has 10 followers F(M)=[FM_1, FM_2,….. FM_{10}], and the followers FM_1, FM_3, FM_8, FM_9, and FM_{10} repost the message sent by M, the out-degree centrality of M is then 5. Only one follower of FM_1, namely FFM_1, reposts the message again and thus the out-degree of FM_1 is 1. Two followers of FM_9, FFM_{91} and FFM_{92}, reposts the message and thus the out-degree of FM_9 is 2.

The out-degree centrality serves as an indicator of the strength of the microblogger in propagating the message. The in-degree centrality of M represents the number of times that M reposts the same message received from other microbloggers and usually it is one time. When following a list of microbloggers, M does possibly receive the same message from multiple sources and thus M can still repost more than one time; The betweenness centrality of M represents the total count of pairs of nodes in the network whose shortest path between them consists of M, denoting the relative importance of the position where M is located as a bridging tie to link up different clusters within the network; Betweenness centrality is an indicator of the intermediary power of a node (microblogger) [36]. In this study, we deploy betweenness centralization to measure the extent to which the most central node in the network differs from all other nodes. Moreover, the proportion of out-degree of the maximum degree node to total number of edges, i.e. denoting as Degree (Max), is computed to indicate the role of the highest out-degree microblogger in the network. Therefore the higher value of the Degree (Max), the more important role of the highest out-degree microblogger one has (centralized network).

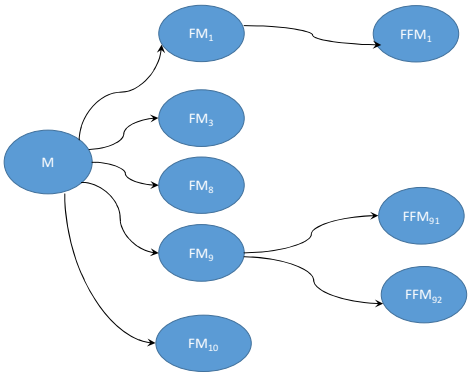

Fig. 1. Reposting Messages as a Network

Some additional network properties were measured [36]. The average path length is the average over all shortest paths between any pair of nodes in a network. The global cluster coefficient is an indicator of the extent to which a network's nodes cluster together, i.e. how many nodes form triangular connections with their adjacent nodes. The diameter of a network represents the longest path between any two nodes in the network. The entropy of degree sequence of a network is a measure of network randomness. For the sake of comparison across networks, the standardized entropy is computed by dividing the entropy by the maximum entropy, i.e. evenly distributed degree sequence. Therefore, the closer the standardized entropy to one, the higher randomness (even distribution) of the degree distribution. Lastly, the degree sequence of a network is fitted to power-law distribution and the power-law exponent is estimated.

2.3 Statistical Analysis

To analyze the set of social networks, we deployed the statistics software package R [37] and its network research package igraph [38]. Network-based commands, including the average shortest path length, standardized entropy, global cluster coefficient, diameter, power-law-fitted, out-degree, and betweenness centralization were used to measure the network characteristics of the samples [36].

Pearson correlation coefficients were used to test the associations between network characteristics. Because of multiple statistical tests between pairs of variables, p values were set as 0.001, 0.01, and 0.05 to indicate different levels of statistical significance.

3 Result

Totally 118 repost networks were analyzed. Their network size ranged from 102 nodes to 63,068 nodes (mean = 5,873, standard deviation = 8,745) and the number of

edges varied from 101 edges to 63,368 edges (mean = 5,898, standard deviation = 8,807). The time duration between the first repost and the last repost made (based on the time stamp of the data collected by the API) varied from 17 hours to 5,450 hours (mean = 1,820 hours, standard deviation = 1,463 hours).

Table 1 presents the descriptive statistics of the network characteristics of the samples. A few points are worth noting. First, the mean diameter and average path length were 8.85 and 1.98 respectively, indicating that on average each network had at least a microblogger who was about 9 steps away from the author of the original message. The central tendency of the networks' average shortest path between all microbloggers was about 2 steps. These results seem to suggest that the path of diffusion of reposts is highly efficient.

Second, on average 58% of the network connections were generated by the largest out-degree microblogger in the network and only a tiny proportion of the network, i.e. 3.04×10^{-5} as the mean cluster coefficient, was clustered.

Third, the mean and the standard deviation of the standardized entropy were 0.21 and 0.07 respectively, suggesting that the degree distribution of the sampled networks were mostly nonrandom. The mean betweenness centralization was found as 3.67×10^{-4}. These findings were consistent with the large value of maximum out-degree because of the central role of the key microblogger. Last, the average power-law exponent was 2.37, which is surprisingly close to the one estimated for Twitter connections [30]. This also suggests that a small group of high-out-degree microbloggers contributes the majority of network connections.

Table 1. Descriptive statistics of the network characteristics (n=118)

	Minimum	First quartile	Median	Mean	Third quartile	Maximum	Standard deviation
Diameter	2	5	8	8.85	12	34	5.10
Average Path Length	1.02	1.35	1.65	1.98	2.29	6.53	1.04
Cluster Coefficient	0.00	1.66E-7	3.31E-6	3.04E-5	1.32E-5	5.65E-4	9.30E-5
Degree (Max)	0.21	0.38	0.62	0.58	0.77	0.98	0.22
Betweenness	4.75E-7	1.66E-5	6.39E-5	3.67E-4	1.95E-4	8.33E-3	1.09E-3
Standardized Entropy	0.07	0.16	0.20	0.21	0.26	0.49	0.07
Power-law exponent	0.00	2.18	2.40	2.37	2.62	3.30	0.45

Table 2. shows the Pearson correlation coefficients between the network characteristics of the 118 repost networks.

Table 2. Pearson correlations between the network characteristic (n=118)

	No. of microbloggers	Diameter	Average Path Length	Cluster Coefficient	Degree (Max)	Betweenness	Standardized Entropy	Power-law exponent
No. of microbloggers	1	0.52***	0.33***	ns	-0.29**	ns	-0.19*	ns
Diameter		1	0.85***	ns	-0.66***	0.42***	0.25**	0.18*
Average Path Length			1	ns	-0.68***	0.59***	0.27**	ns
Cluster Coefficient				1	-0.19*	0.43***	0.4***	ns
Degree (Max)					1	-0.27**	-0.4***	ns
Betweenness						1	0.26**	ns
Standardized Entropy							1	0.2*
Power-law exponent								1

Remark: p<0.001***; p<0.01**; p<0.05*; "ns" stands for "non-significance"

The number of microbloggers in the network and the network diameter are the two measures for the extent to which the diffusion of microblogs reached out, i.e. effectiveness of the process of diffusion. As shown in the Table 2, the number of microbloggers in the network who reposted the original message was positively associated with the network diameter (r=0.52, p<0.001) and the average path length (r=0.33, p<0.001) but was negatively correlated with the maximum degree of network (r=-0.29, p<0.01) and the standardized entropy (r=-0.19, p<0.05).

For the network diameter, it was positively associated with the average path length (r=0.85, p<0.001), the standardized entropy (r=0.25, p<0.01), the betweenness centralization (r=0.42, p<0.001), and the power-law exponent (r=0.18, p<0.05) but was correlated negatively with the maximum degree of network (r=-0.66, p<0.001).

4 Discussion

In this study, we attempt to conceptualize the process of diffusion of reposting microblogs in China as a "social network" and deploy social network analysis to operationalize and describe the network characteristics of such interconnections between "networked" microbloggers who involve in reposting an original message.

Based on our findings, we identify some network characteristics that are conducive to the spread of messages about social issues via microblogs in China. While the presence of highly centralized microbloggers is essential [39], say higher betweenness centralization, our study finds that a lower ratio of maximum degree to total number of edges in the network, i.e. a less dominant position of the highest out-degree microblogger or a relative decentralized network, is a key contributing factor for information diffusion. It suggests that the effectiveness of microblog diffusion is not overly dependent on a dominant "online opinion leader", but rather a number of

players who have high power to generate reposts and subsequent reposts. Moreover, our study finds a conflicting result on the role of randomness in information diffusion, despite previous finding that randomness helps diffuse information [35]. This part of question requires further investigation.

The finding of this study is particularly timely under the circumstance that the Chinese government has enacted a new regulation to control the spread of "rumor" via reposting weibos (being reposted 500 times) and subsequently some well-known "online opinion leaders" were allegedly targeted [16]. Since the implementation of the real name registration system for microbloggers in 2012 [17], such measures of Internet regulation may be another political intervention of the Chinese government that could limit the free flow of information in China.

Our findings provide an empirical base to understand the role of network characteristics in diffusion of microblogs and we have demonstrated that the conception of "social network" can be applied to theorize and operationalize the flow of microblogs. Nevertheless, our analysis is preliminary and exploratory. Future research is warranted to investigate the information diffusion as a function of time and its association with a broader range of network characteristics.

References

1. Dijck, J.V.: The culture of connectivity: A critical history of social media. Oxford University Press, Oxford (2013)
2. White, J., Fu, K.-W., Benson, B.: Social Media: An Ill-Defined Phenomenon. In: Ozok, A.A., Zaphiris, P. (eds.) OCSC 2013. LNCS, vol. 8029, pp. 422–431. Springer, Heidelberg (2013)
3. Efron, M.: Information search and retrieval in microblogs. Journal of the American Society for Information Science and Technology 62, 996–1008 (2011)
4. Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment in Twitter events. Journal of the American Society for Information Science and Technology 62, 406–418 (2011)
5. Mendoza, M., Poblete, B., Castillo, C.: Twitter under crisis: Can we trust what we RT? In: Social Media Analytics, KDD 2010 Workshops. ACM, Washington, DC (2010)
6. Preis, T., Moat, H.S., Bishop, S.R., Treleaven, P., Stanley, H.E.: Quantifying the Digital Traces of Hurricane Sandy on Flickr. Sci. Rep. 3 (2013)
7. Qu, Y., Huang, C., Zhang, P., Zhang, J.: Microblogging after a major disaster in China: A case study of the 2010 Yushu earthquake. In: CSCW 2011, pp. 25–34. ACM (2011)
8. Vieweg, S., Hughes, A., Starbird, K., Palen, L.: Microblogging during two natural hazards events: What Twitter contribute to situational awareness. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 1079–1088. ACM (2010)
9. Adamic, L., Glance, N.: The political blogosphere and the 2004 US election: Divided they blog. In: Proceedings of the 3rd International Workshop on Link Discovery, pp. 36–43. ACM (2005)
10. Burns, A., Eltham, B.: Twitter free Iran: An evaluation of Twitter's role in public diplomacy and information operations in Iran's 2009 election crisis. Communications Policy & Research Forum 2009, University of Technology, Sydney (2009)

11. Lassen, D.S., Brown, A.R.: Twitter: The electoral connection? Social Science Computer Review 29, 419–436 (2011)
12. Castells, M.: Networks of outrage and hope: social movements in the Internet age. Polity, Cambridge (2012)
13. Grossman, L.: Iran protests: Twitter, the medium of the movement. Time Magazine (2009)
14. OpenNet Initiative: China (2009)
15. Reporters Without Borders, http://en.rsf.org/press-freedom-index-2013,1054.html
16. Chin, J.: China Intensifies Social-Media Crackdown. The Wall Street Journal (2013)
17. Fu, K.W., Chan, C., Chau, M.: Assessing Censorship on Microblogs in China: Discriminatory Keyword Analysis and the Real-Name Registration Policy. IEEE Internet Computing 17, 42–50 (2013)
18. China Internet Network Information Center: 32nd statistical report on Internet development in China. China Internet Network Information Center (2013)
19. Mozur, P.: How many people really use Sina Weibo? Wall Street Journal China, Beijing (2013)
20. Rogers, E.M.: Diffusion of innovations. Free Press, New York (1995)
21. Wejnert, B.: Integrating models of diffusion of innovations: A conceptual framework. Annu. Rev. Sociol. 28, 297–326 (2002)
22. Fu, K.W., Chau, M.: Use of Microblogs in Grassroots Movements in China: Exploring the Role of Online Networking in Public Agenda-Setting. Journal of Information Technology & Politics (online first, 2014)
23. Wellman, B.: Computer Networks As Social Networks. Science 293, 2031–2034 (2001)
24. Castells, M.: The rise of the network society. Blackwell Publishers, Cambridge (1996)
25. Benkler, Y.: The wealth of networks how social production transforms markets and freedom. Yale University Press, New Haven (2006)
26. Wellman, B.: Little boxes, glocalization, and networked individualism. In: Tanabe, M., van den Besselaar, P., Ishida, T. (eds.) Digital Cities 2001. LNCS, vol. 2362, pp. 10–343. Springer, Heidelberg (2002)
27. Rainie, L., Wellman, B.: Networked: The new social operating system. The MIT Press, Cambridge (2012)
28. Gruzd, A., Wellman, B., Takhteyev, Y.: Imagining Twitter as an Imagined Community. Am. Behav. Sci. 55, 1294–1318 (2011)
29. Huberman, B., Romero, D., Wu, F.: Social networks that matter: Twitter under the microscope. Available at SSRN 1313405 (2008)
30. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: WWW 2010, pp. 591–600. ACM (2010)
31. Granovetter, M.S.: The strength of weak ties. Am. J. Sociol. 78, 1360–1380 (1973)
32. Zhao, J., Wu, J., Xu, K.: Weak ties: Subtle role of information diffusion in online social networks. Physical Review E 82, 016105 (2010)
33. Takhteyev, Y., Gruzd, A., Wellman, B.: Geography of Twitter networks. Social Networks 34, 73–81 (2012)
34. Opuszko, M., Ruhland, J.: Impact of the Network Structure on the SIR Model Spreading Phenomena in Online Networks. In: ICCGI 2013, The Eighth International Multi-Conference on Computing in the Global Information Technology, pp. 22–28 (2013)
35. Watts, D.J.: The "New" Science of Networks. Annu. Rev. Sociol. 30, 243–270 (2004)
36. Lewis, T.G.: Network science: Theory and practice. John Wiley & Sons, Hoboken (2009)

37. R Development Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing (2011)
38. Csardi, G., Nepusz, T.: The igraph software package for complex network research. Inter. Journal, Complex Systems, vol. 1695 (2006)
39. Guan, W., Gao, H., Yang, M., Li, Y., Ma, H., Qian, W., Cao, Z., Yang, X.: Analyzing user behavior of the micro-blogging website Sina Weibo during hot social events. Physica A: Statistical Mechanics and its Applications 395, 340–351 (2014)