An Empirical Study of Information Diffusion in Micro-blogging Systems during Emergency Events

Kainan Cui[1,2], Xiaolong Zheng[2,3], Daniel Dajun Zeng[2], Zhu Zhang[2], Chuan Luo[2], and Saike He[2]

[1] The School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China
[2] The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
[3] Dongguan Research Institute of CASIA, Cloud Computing Center, Chinese Academy of Sciences, Songshan Lake, Dongguan, 523808, China
kainan.cui@live.cn, xiaolong.zheng@ia.ac.cn

Abstract. Understanding the rapid information diffusion process in social media is critical for crisis management. Most of existing studies mainly focus on information diffusion patterns under the word-of-mouth spread mechanism. However, to date, the mass-media spread mechanism in social media is still not well studied. In this paper, we take the emergency event of Wenzhou train crash as a case and conduct an empirical analysis, utilizing geospatial correlation analysis and social network analysis, to explore the mass-meida spread mechanism in social media. By using the approach of agent-based modeling, we further make a quantativiely comparison with the information diffusion patterns under the word-of-mouth spread mechanism. Our exprimental results show that the mass-meida spread mechanism plays a more important role than that of the word-of-mouth in the information diffusion process during emergency events. The results of this paper can provide significant potential implications for crisis management.

Keywords: Information diffusion, opinion dynamic, emergency response, social media, micro-blogging systems.

1 Introduction

The rapid diffusion of information and opinions through social media is affecting the development of crisis situations [1-3]. Many researches show that social media has great potential for improving the situational awareness during emergency situation, such as mining the actionable information for earthquake [4], tracking the transmission rate and studying the response behavior for Influenza outbreak [5, 6]. However, despite those opportunities, the social media may also contribute to the effective broadcasting of rumors and extreme opinions, which will lead to huge social and economic damage [7].

Generally, there are two main sources, including the word-of-mouth and the mass-media[8], for us to obtain information. Word-of-mouth information spreading via

Y. Gao et al. (Eds.): WAIM 2013 Workshops, LNCS 7901, pp. 140–151, 2013.

online social network is considered as the main feature of the social media. Researches on the this mechanism have exploded in recent years [9]. The affecting factors of the word-of-mouth information spreading mechanism include social network structure[10, 11] , geographic distance [12], opinion leader [13] and so on. On the other hand, there are also evidences for the mass-media information spreading in social media. A recent study found that micro-blogging resembles a traditional news media more closely [14]. In another word, information diffusion in social media could also follow the mass-media information spreading mechanism. Understanding the information spreading mechanism has several implications for emergency management. From the perspective of countering rumors, it will not only provide us tools to predict the rumor dynamics, but also guide us on how to reduce the loss. To the best of our knowledge, few works have been done to empirically examine the information spreading mechanism in social media during emergency events.

To fill the research gaps, we conducted a case study to investigate how information was spread during the 2011 Wenzhou train crash through the Sina Weibo, a popular Chinese micro-blogging system. In particular, we performed cluster analysis on the diffusion outcomes at first. Then we applied an agent-based model to simulate the temporal trends of diffusion. Our results indicate that the mass-media spreding mechanism could better reproduce the temporal trends of information diffusion than the word-of-mouth spreding mechanism.

The rest of this paper is organized as follows. In Section 2, we begin with a brief introduction to the Wenzhou train crash dataset and intuitions obtained from the dataset. Motived by the intuitions from Section 2, we perform empirical analysis and report results in Section 3. Both evidences for word-of-mouth spreading mechanism and mass-media spreading mechanism were found in this section. In Section 4, we further applied an agent-based Bass model to simulate the temporal trends of our dataset. The simulation results suggest that the mass-media spreading mechanism played a more important role in social media during emergency events than word-of-mouth spreading mechanism. We conclude our paper by summarizing the findings and discussing several key issues in Section 5.

2 The Dataset

"Wenzhou train crash" accident happened at night of July 23, 2011 in Zhejiang Provience. This event caused 40 deaths and over 100 people injured. Micro-blogging systems played a crucial role in the information sharing and dissemination during the event. More than 5.3 million messages were posted to microblog about this event[15].

Poll system on the Sina Weibo gives us an excellent way to study the information spreading mechanisms during emergency events. The features of poll system make each poll in the platform a perfect diffuison process. At first, user can only paticipate specific poll through accessing certain URL. Second, each user is allowed to post once. There are two ways to spread a poll URL in the Micro-blogging platfrom. User sharing is the first one, which could be treated as word-of-mouth spreading mechanism. Users could also access certain poll from the new polls ranking and the hot polls ranking in the front page. This process is similar to the mass-media spreading mechanism. Our dataset is a poll titled "Will you still support China High-speed Train?" The poll started at 15:05, July 26, 2011 and closed by 15:04, August 2, 2011. There are 2071 voters totally. By applying a customized crawler, we collected all vote records with corresponding user profiles and following relationships.

The original poll has five options which could be easily classified into positive opinions, negative opinions and neutral opinions. After the division, we have 645 users who support CHT (China High-speed Train), 773 users who are against the CHT, 653 users hold neutral views. The exsiting of competive opinions in dataset provide us opportunity to study the difference among the diffusion outcomes of varisous opinions.

2.1 Social Network of Voters

The word-of-mouth spreading mechanism relies on the social network structure. In Micro-blogging platform, the social network is constructed by the following relationships. After removing the edges linking to users who did not vote, we have 3888 edges. We found that the social network is not fully-connected, which means the poll URL could not reach all voters via person-to-person spreading. Therefore, the word-of-mouth spreading mechanism could not explain the diffusion outcomes alone. As shown in Fig. 1, we visualized the giant component of voters' social network by representing voters as nodes and relationships among them as links (The figure was plotted by NodeXL http://nodexl.codeplex.com/). This network is the biggest weakly connected sub-graph of the whole social network. It contains 863 nodes and 3505 edges. The red nodes represent voters who support CHT. The blue nodes represent voters who oppose CHT. The green nodes represent voters who have neutral views.

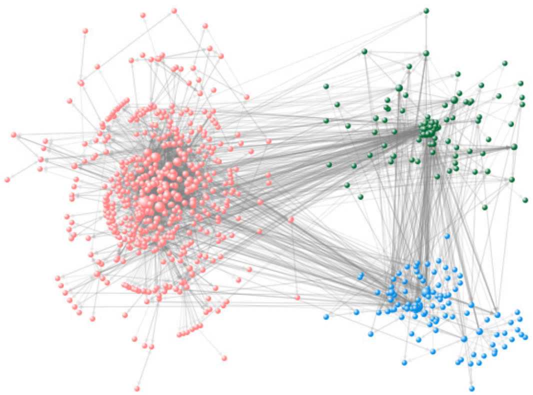

Fig. 1. The **structure of the giant component of voters' social network**, red nodes represent positive voters; blue nodes represent negative voters; green nodes represent neutral voters. The node size is in proportion to the out degree.

The exisiting of social network association is the essential condition for the word-of-mouth spreading mechanism. Therefore, if there is no social network association, then we can exclude the word-of-mouth spreading mechanism. We can observe an obvious cluster among positive voters. This intuition leads the social network association analysis in Section 3.1.

2.2 Geospatial Distribution of Voters

Geospatial distribution is another perspective of our dataset. The user profile contains the location information of most users. Under the assumption that people are more concerned about local events and have more willingness to post opinions on Micro-blogging, the areas near the accident spot will have more voters. After removing records outside mainland China or missing location information, we have 1882 records. Fig. 2 shows the geospatial variation of voters within province-level. This map intuitively presents that the voters are more concentrated in the accident spot and surrounding areas. We also found Beijing and Guangdong have more voters. Motivated by this result, we plan to quantify the geospatial correlation and identify the hot spot of voters in Section 3.2.

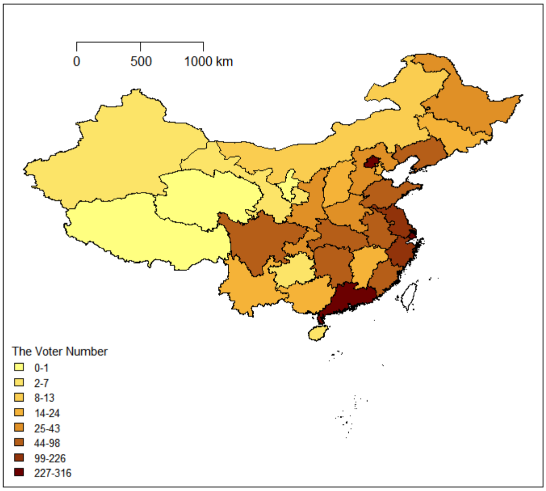

Fig. 2. Geospatial variation of diffusion outcomes in mainland China

3 Empirical Analysis

3.1 Social Network Association Analysis

Social Network Association of all Voters: We choose global cluster coefficient to analyze the whole voter network. Global cluster coefficient [16] is a measure to indicate the association degree of a graph and is defined as follows:

$$C = \frac{1}{n}\sum_{u \in G} \frac{T(u)}{deg(u)(deg(u)-1)} \quad (1)$$

Where G is the graph, n is the number of nodes in G, $T(u)$ is the number of triangles through node u, and $deg(u)$ is the number of neighbors of u.

As shown in Table 1, the global cluster coefficient of our voter network is a little bigger than the general network of Sina microblog. This result shows that the association level of our voter network is similar with the general network.

Table 1. A comparison of cluster degree between the general network of Sina microbog and the voter netwrok of our dataset

Network	APL	Diameter	Cluster Coefficient
General network[17]	4.0	12	0.21
Voter network	4.43	13	0.22

APL is the average path length; APL and Diameter is computed for the giant component of network

Social Network Association of Voters with Different Opinions: A commonly used method for social network association analysis among different categories is to construct a random mixing network and then compare the edges number in different categories[18]. In our dataset with 3 opinions, there are 9 types of edges (e.g., positive-positive edges, positive-negative, positive-neutral, etc.). Chi-square tests were used on the categorical data in order to assess whether there is a significant difference between the real network and the random mixing network.

Table 2 shows the comparison result between real network and random mixing network. From this table, we obtain the following observations: 1) the numbers of edges with same opinion are always greater than the numbers of edges with opposite opinion. For example, positive-positive edges (2591) are more than positive-negative edges (107), negative-negative edges (246) are more than negative-positive (97) edges, and so on; 2) the number of positive-positive edges from real network is much higher than the expected number; 3) the number of negative-negative edges and the number of neutral-neutral edges form real network are lower than the expected number. The results indicate strong social network association exists among positive voters, while the association level among negative voters and neutral voters are lower than random mixing network.

Table 2. A comparison of edges numbers between real network and random mixing network

followee / follower	Positive	Negative	Neutral
Positive	**2591**	107	182
Expected	376.7224	452.1838	381.9871
Negative	97	246	181
Expected	452.1838	541.2185	457.7923
Neutral	142	157	185
Expected	381.9871	457.7923	386.1327

(Chi-squared = 1722.219, df = 4, p-value < 2.2e-16).

Expected is the number of edges in random mixing network.

3.2 Geospatial Association Analysis

Geospatial Association of all Voters: "Local Moran's I" [19] is selected to identify the geospatial hot spot for all voters. It is defined as follows:

$$I_i = \frac{n(x_i - \bar{x}) \sum_{j=1}^{n} w_{ij}(x_j - \bar{x})}{\sum_{j=1}^{n} (x_j - \bar{x})^2} \tag{2}$$

Where x_i and x_j are the diffusion outcomes at area i and j, $\bar{x}$ is the average value of all areas in the entire region, w_{ij} is the weight of the spatial neighborhood relationship, and n is the population size.

Fig. 3 shows the cluster map of local Moran' I (The map was plotted by GeoDA software https://www.geoda.uiuc.edu/). A cluster is composed by locations which are more similar to its neighbors. The High-High locations refers to hot spot areas where the number of voters is higher than average, whereas the Low-Low locations refers to cool spot areas where the number of voters is lower than average. From Fig. 3, we can observe Shanghai, Zhejiang and Jiangsu as a hot spot, Xinjiang as a cool spot. These results indicate that the hot spot of voters is near the incident spot, while the cool spot of voters is far form the incident spot.

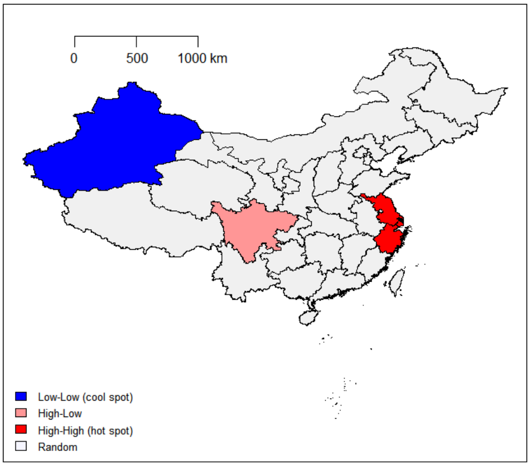

Fig. 3. Cluster map of diffusion outcomes based on Local Moran' I

Geospatial Association of Voters with Different Opinions: Global Moran's I [19] is selected to quantify the geospatial correlation for voters with different opinions. It is calculated as follows:

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})^2} \tag{3}$$

Where x_i, x_j, $\bar{x}$ and w_{ij} have the same meaning as in Equation (2). It is worthy to note that the global Moran's I is calculated by averaging the local Moran's I in each area.

The statistics are summarized in Table 3. From this table, we obtain the following observations 1) the geospatial distribution of all voters has positive geospatial association; 2) the geospatial distribution of positive voters is nearly random; 3) the geospatial distribution of negative voters and the geospatial distribution of neural voters have positive geospatial association.

Table 3. A comparison for the geospatial association of voters with different opinions

Opinions	Moran's I	999 MC simulation			
		E(i)	S_d	Z-score	P-value
Positive	0.092283	-0.0323	-0.0352	1.4447	0.093
Negative	0.271442	-0.0323	0.1075	2.7739	0.012
Neutral	0.228925	-0.323	0.1027	2.5496	0.019
All	0.20936	-0.0357	0.1008	2.4301	0.024

3.3 Summary of Empirical Analysis

In this Section, we performed empirical analysis to quantify the intuitions obtained in Section 2. In particular, social network association analysis and geospatial association analysis were conducted respectively. The results of association analysis are summarized in Table 4.

For all voters, we found a medium cluster degree in social network and a positive correlation in geospatial distribution; for positive voters, we found a high cluster degree in social network and a nearly random correlation in geospatial distribution; for negative and neutral voters, we found a low cluster degree and a positive correlation in geospatial distribution. Although the geospatial distribution of the voter population expresses a positive correlation as a whole, which could be better explained by the mass-media information spreading mechanism, we cannot rule out the word-of-mouth information spreading mechanism based on the results thus far. The high cluster degree of positive voters provides evidence for the word-of-mouth information spreading mechanism.

Table 4. A summary of results about social network association analysis and geospatial association analysis

Opinions	social network cluster degree	geospatial correlation
Positive voters	high	nearly random (0.09)
Negative voters	low	positive (0.27)
Neutral voters	low	positive (0.23)
All voters	medium	positive (0.21)

4 Experiments

In this section, we plan to further analyze the temporal dynamics of the diffusion process through simulation. We will first present the time series of the diffusion dynamics and discuss the design of experiments. In Section 4.1, we will describe an agent based model inspired by the Bass model. In Section 4.2, we will show the experiments results.

Because the voter's social network is not fully-connected, we simulate the information diffusion in the giant component of the voters' social network. Fig. 4 shows the time series for daily new voter numbers in the giant component of voters' social network. The plot shows a sharply increasing during the second day and a rapidly declining during the following period.

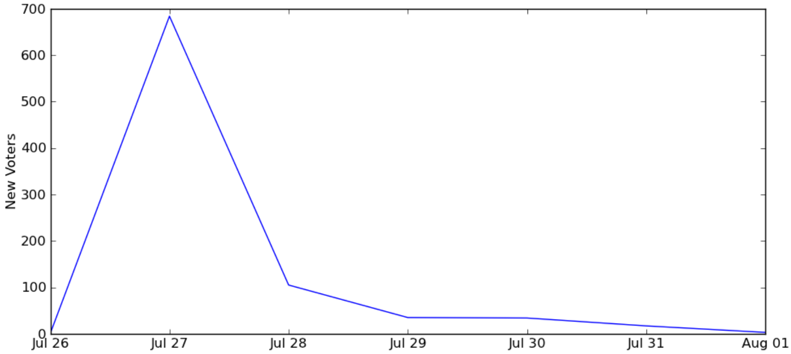

Fig. 4. Time series plots of new voters in the giant component of voter's social network

4.1 Models

The Bass model [8], which was originally developed to model the diffusion of new products in marketing, is defined as follows:

$$\frac{f(t)}{1-F(t)} = p + qF(t) \tag{4}$$

Where $f(t)$ is the rate of change of the adopted fraction, $F(t)$ is the adopted fraction, p is the coefficient for mass media effect, q is the coefficient for word-of-mouth effect.

In order to leverage the information of real-world network structure, we implement the Bass model in the framework of agent based modeling [20]. In this model, each agent is affected only by its neighbors, instead of the entire population. We introduce the model as follows: 1) the states of agents are classified into unaware state and aware state; 2) mass-media spreading mechanism: at each time step, every unaware agent has a possibility α to turn into aware state; 3) word-of-mouth spreading mechanism: at each time period, every aware agent attempts to affect their neighbors with transmission rate β; 4) At time step t=0, only 1 agent is aware; 5) the network structure is initialized according to the real-world network.

There are totally 8 experiments scenarios. Half of them are designed for the word-of-mouth spreading mechanism and half of them are designed for the mass-media spreading mechanism. In the scenarios for word-of-mouth spreading mechanism, the parameter α was set to 0, and the parameter β was set to 0.6, 0.7, 0.8 and 0.9 respectively. In the scenarios for mass-media spreading mechanism, parameter β was set to 0, and the parameter α was set to 0.6, 0.7, 0.8 and 0.9 respectively. We will run 50 times for each scenario. The results will be averaged and displayed in Section 4.2.

4.2 Results

We will first present the simulation results for the word-of-mouth spreading mechanism. Then will present the simulation results for the mass-media spreading mechanism. As Fig. 5 shown, word-of-mouth spreading mechanism could not reproduce the temporal trends of the diffusion process alone.

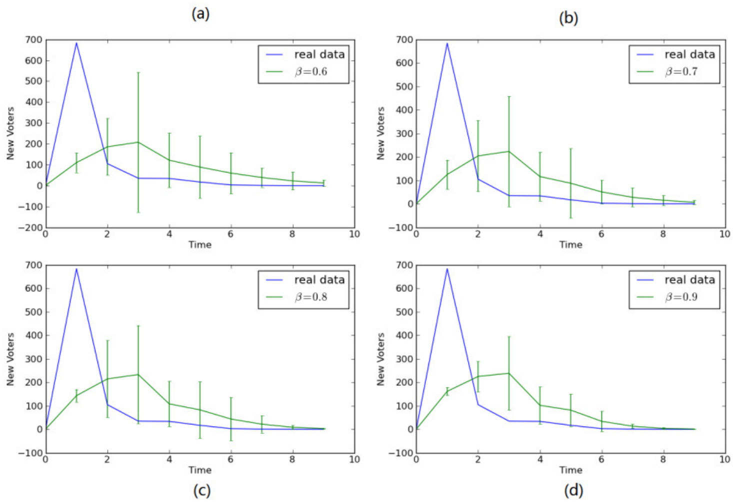

Fig. 5. Simulation result of word-of-mouth spreading mechanism. Four scenarios with different value of beta are displayed, each run 50 times and averaged. (a) beta=0.6. (b) beta=0.7. (c) beta=0.8. (d) beta=0.9.

As shown in Fig. 6, mass-media spreading mechanism could well reproduce the temporal trends of the diffusion process. These results indicate that the mass-media spreading mechanism play a more important role in our dataset.

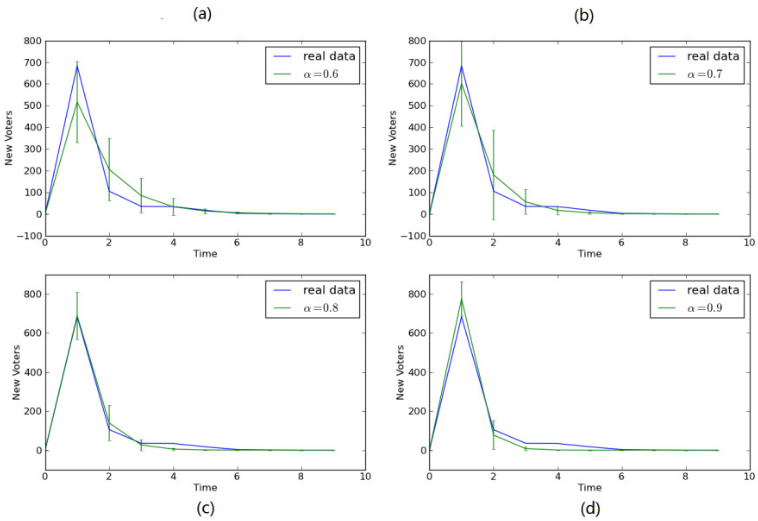

Fig. 6. Simulation result of mass-media spreading mechanism. Four scenarios with different value of alpha are displayed, each run 50 times and averaged. (a) alpha=0.6. (b) alpha =0.7. (c) alpha =0.8. (d) alpha =0.9.

To quantify the results in Fig.5 and Fig.6, the person correlation coefficient was calculated between the simulation output and the original data. As summarized in Table 5, the mass-media spreading mechanism has is obviously outperform word-of-mouth spreading mechanism.

Table 5. A comparison of correlation between the word-of-mouth spreading mechanism and the mass-media spreading mechanism

Mechanism	α	β	Person correlation	P-value
Mass-media	0.6	0	0.96761	P<0.001
	0.7	0	0.98735	P<0.001
	0.8	0	0.99726	P<0.001
	0.9	0	**0.99791**	P<0.001
Word-of-mouth	0	0.6	0.16003	0.510
	0	0.7	0.21718	0.372
	0	0.8	0.28764	0.238
	0	0.9	0.34966	0.159

5 Conclusion

In our research, we have investigated a real-world information diffusion case in Micro-blogging during the Wenzhou train crash event. Our study reveals that mass-media spreading mechanism can better explain the empirical data than the word-of-mouth spreading mechanism, which is considered as the main feature of information diffusion in social media. One possible explanation could be that the credibility of the information sources is more important during emergency events. According to the Bandwagon effect, people are likely to believe the choice of public. The hot topic ranking system as a mechanism to reflect the public choice will play the role like the mass media. If this assumption is true, we should focus on how to design ranking algorithms and borrow experience form mass media to improve the information credibility.

Another interesting phenomena are further found in this paper. We obtain that a high cluster degree in social network is related to a nearly random correlation in geospatial distribution and a low cluster degree in social network is related to a positive correlation in geospatial distribution. This phenomenon could be explained by the online communication, which break the physical boundary for information sharing and extend the association from offline world to the online world.

One of the limitations in this study is that we did not consider the two spreading mechanisms at the same time. However, this work still shed light on the information diffusion through Micro-blogging during emergency events. Further empirical work is needed to verify our findings.

Acknowledgments. The authors would like to thank Zhaonan Zhu and Jiaqi Liu for their helpful suggestions. This work was supported in part by the following grants: The National Natural Science Foundation of China, Grant No. 91124001, 71103180, 91024030 and by the Ministry of Health under Grant No. 2012ZX10004801.

References

1. Adam, N.R., Shafiq, B., Staffin, R.: Spatial Computing and Social Media in the Context of Disaster Management. IEEE Intelligent Systems 27, 90–96 (2012)
2. Daniel, Z., Hsinchun, C., Lusch, R., Shu-Hsing, L.: Social Media Analytics and Intelligence. IEEE Intelligent Systems 25, 13–16 (2010)
3. Zeng, D., Wang, F.-Y., Carley, K.M.: Guest Editors' Introduction: Social Computing. IEEE Intelligent Systems 22, 20–22 (2007)
4. Tyshchuk, Y., Wallace, W.A.: Actionable Information during Extreme Events – Case Study: Warnings and 2011 Tohoku Earthquake. In: Conference Actionable Information during Extreme Events – Case Study: Warnings and 2011 Tohoku Earthquake, pp. 338–347 (2012)
5. Culotta, A.: Towards detecting influenza epidemics by analyzing Twitter messages. In: Proceedings of the First Workshop on Social Media Analytics, pp. 115–122. ACM, Washington D.C (2010)
6. Cui, K., Cao, Z., Zheng, X., Zeng, D., Zeng, K., Zheng, M.: A Geospatial Analysis on the Potential Value of News Comments in Infectious Disease Surveillance. In: Chau, M., Wang, G.A., Zheng, X., Chen, H., Zeng, D., Mao, W. (eds.) PAISI 2011. LNCS, vol. 6749, pp. 85–93. Springer, Heidelberg (2011)
7. Il-Chul, M., Oh, A.H., Carley, K.M.: Analyzing social media in escalating crisis situations. In: Conference Analyzing Social Media in Escalating Crisis Situations, pp. 71–76 (2011)
8. Bass, F.M.: Comments on "A New Product Growth for Model Consumer Durables The Bass Model". Management Science 50, 1833–1840 (2004)

9. Zheng, X., Zhong, Y., Zeng, D., Wang, F.-Y.: Social influence and spread dynamics in social networks. Front. Comput. Sci. 6, 611–620 (2012)
10. Kim, M., Xie, L., Christen, P.: Event Diffusion Patterns in Social Media. In: Conference Event Diffusion Patterns in Social Media (2012)
11. Zheng, X., Zeng, D., Li, H., Wang, F.: Analyzing open-source software systems as complex networks. Physica A: Statistical Mechanics and its Applications 387, 6190–6200 (2008)
12. Toole, J.L., Cha, M., González, M.C.: Modeling the Adoption of Innovations in the Presence of Geographic and Media Influences. PLoS ONE 7,e29528 (2012)
13. Watts, D.J., Dodds, P.S.: Influentials, Networks, and Public Opinion Formation. Journal of Consumer Research 34, 441–458 (2007)
14. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World Wide Web, pp. 591–600. ACM, Raleigh (2010)
15. Xu, X.: Internet Facilitated Civic Engagement in China's Context: A Case Study of the Internet Event of Wenzhou High-speed Train Accident (2011)
16. Watts, D.J., Strogatz, S.H.: Collective dynamics of /'small-world/' networks. Nature 393, 440–442 (1998)
17. Pengyi, F., Pei, L., Zhihong, J., Wei, L., Hui, W.: Measurement and analysis of topology and information propagation on Sina-Microblog. In: Conference Measurement and Analysis of Topology and Information Propagation on Sina-Microblog, pp. 396–401 (2011)
18. Thelwall, M.: Homophily in MySpace. Journal of the American Society for Information Science and Technology 60, 219–231 (2009)
19. Moran, P.A.: The interpretation of statistical maps. Journal of the Royal Statistical Society. Series B (Methodological) 10, 243–251 (1948)
20. Li, X., Mao, W., Zeng, D., Wang, F.-Y.: Agent-Based Social Simulation and Modeling in Social Computing. In: Yang, C.C., et al. (eds.) ISI Workshops 2008. LNCS, vol. 5075, pp. 401–412. Springer, Heidelberg (2008)