Agent-Based Modeling of Netizen Groups in Chinese Internet Events

Zhangwen Tan, Xiaochen Li, and Wenji Mao

Institute of Automation, Chinese Academy of Sciences, Beijing, China
{zhangwen.tan,xiaochen.li,wenji.mao}@ia.ac.cn

Abstract. Internet events are public events with the participation of netizens to express their opinions or comments. As an emerging phenomenon, Internet events often draw nationwide attention and eventually influence offline events. Netizen groups who participate in the Internet events play a central role in such events. In this paper, we focus on the study of netizen groups and propose an agent-based model to capture their dynamics and evolvement in Internet events. Our experiment is based on two case studies of Chinese Internet events. We test the proposed model by running simulations and comparing experimental results with real social media data to show the effectiveness of our model.

Keywords: agent-based modeling, netizen group, Internet event.

1 Introduction

Internet events refer to public events which draw the attention and participation of large numbers of netizens. In an Internet event, unorganized netizens autonomously express their sentiments and opinions online which often influence the offline behavior associated with the event. Netizen groups who participate in the Internet events play a central role in such events. In recent years, with the development of Internet and mobile technologies, Internet events happen frequently and have become an emerging social phenomenon that significantly impact and reshape public and social lives.

Internet events are often involved with the participation and interaction of netizen groups——the biggest difference between Internet events and normal events. In order to understand this new phenomenon, it is important to model and capture the mechanism of netizen-centered Internet events from a computational perspective. This is a challenging issue, though. Some research attempts to investigate the issue through statistical analysis and data mining, but their focus is on the specific aspects such as netizen groups' past topics, interests or opinions. Due to the complexity of the interactions in an Internet event and the involvement of multiple parties, it is difficult to model netizen groups and capture the mechanism of Internet events using traditional methods such as statistical analysis or mathematical methods.

To tackle this issue, we ground our work on agent-based modeling of netizen groups and Internet events. Agent-based modeling is a tractable method for capturing netizen group's opinion and actions [1-3]. It is capable of modeling the emergence of social phenomenon through the decomposition and specification of individually

M. Chau et al. (Eds.): PAISI 2011, LNCS 6749, pp. 43–53, 2011.
© Springer-Verlag Berlin Heidelberg 2011

involved parties (i.e., agents) and their interactions. Recent studies show that the causes of Chinese Internet events can be located in public concerns, sentiments, and demands [4]. Thus netizen groups' sentiments and opinions, which largely reflect public concerns, are the key factors in promoting the evolution of Internet events. Therefore, one particular focus of our agent-based model is on the sentiments and opinions of the netizen groups in Internet events. In addition, to capture the interactions among multiple parties, our model considers all the parties involved in Internet events. We evaluate our model using two case studies of Chinese Internet events [5, 6]. The preliminary modeling results verify the effectiveness of our proposed model.

The rest of this paper is structured as follows. Section 2 briefly reviews the related work. Section 3 presents our agent-based model in detail. Then in Section 4, we test our model based on social media data of two Chinese Internet events occurred in 2010, Synutra event [5] and "360 versus QQ" event [6]. Section 5 concludes this paper.

2 Related Work

Netizen groups have been studied in crisis management domain. Hughes *et al.* [7] outlined several types of online social convergence behavior during times of crisis: *help*, *be anxious*, *return*, *support*, *mourn*, *exploit*, and *be curious*. Sutton *et al.* [8] studied information sharing and dissemination practice by the public during the October 2007 Southern California Wildfires. The authors suggested that community information resources and other unofficial communicative activities enabled by social media are gaining significant momentum in practice, despite concern by officials about the legitimacy of information shared through such means. They argued that these emergent uses of social media are precursors of broader future changes to the institutional and organizational arrangements of disaster response. Vieweg *et al.* [9] analyzed a selected set of online interactions that occurred in the aftermath of the 2007 shooting rampage at Virginia Tech, which represented a new and highly distributed form of participation by the public. These findings suggest strong self-organization which includes the development and evolution of roles and norms wildly exists among netizen groups.

Netizen groups are a special type of online community. Research on general online communities can offer insights and research methods that could benefit investigating netizen-centered Internet events. Computational studies of online communities mainly fall into two categories: communities mining [10, 11] and community characteristics computing [12-15]. Community mining aims to identify online communities that are implicit, inconspicuous, or even hidden. A range of computing techniques such as Web crawling, social network analysis, content analysis, and link analysis, have been applied in community mining research. Community characteristics computing involves analyzing community structure [14], identifying leaders and experts from communities [15], searching for information, and finding friends with similar hobbies, among others.

Previous related work mainly focused on mining or modeling specific aspects of netizen groups based on online social media information. We aim at modeling the

dynamics and evolvement of netizen-centered Internet events. In next section, we shall propose an agent-based model of netizen group and other parties involved in Internet event.

3 Proposed Model

The growth of social media websites in China enables more and more netizens to engage in online discussions about hot topics and thus gives rise to the increasing occurrence of Internet events. Through investigating a number of Chinese Internet events occurred in recent years, we find there are five main parties involved in an Internet event, i.e., *main party*, *opposite party*, *netizen group*, *media*, and *government* (See Figure 1). Main party refers to people or group who initiates a hot event and the opposite party is the one having conflict interest against the main party. Both main party and opposite party are directly involved in the event. They are main parts of the event and their actions significantly influence the evolution of Internet events. Netizen group is the collection of netizens who are associated with the Internet event via online participation. Media includes traditional media (e.g., newspaper and TV) and online media (e.g., news websites and social network sites). Media reports latest news about Internet events. It is the medium of information diffusion. Government plays intermediate roles in some events. The response and policies made by government always impact netizens' opinion toward the event and government.

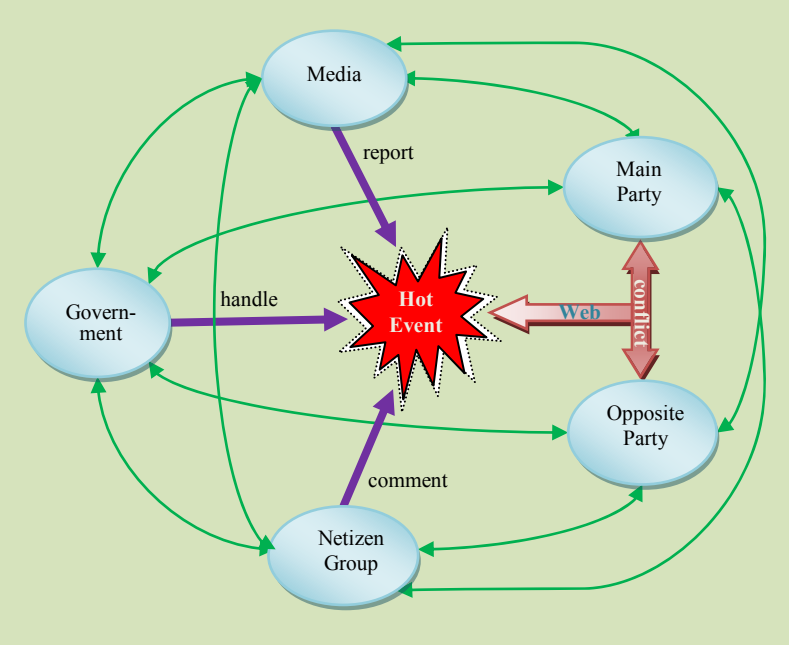


Fig. 1. Agent-based model for netizen-centered Internet events

Netizen groups play a key role in Internet events. In our model, we define the states and actions of the five parties (i.e., agents) as well as their interaction rules. Below we shall introduce the agent design.

States. States are inherent attributes of agents which are closely related to their actions. All the five parties have belief states denoting their belief about an Internet event. For netizen group, media and government, they are not directly involved in an Internet event. They have *concern* as a belief state. Concern denotes the attention of a party to the event. It is a numerical variable ranging from 0 to 1. The main party and the opposite party are directly involved in an Internet event and therefore have *benefit* as a belief state instead of concern. Benefit denotes the benefit the main party/opposite party gets from the event. Benefit is a nominal variable with value *positive*, *negative* or *null*. In addition, we also use *opinion* to capture the positive/negative attitude one agent holds about another agent. For example, the *netizen group* has *opinion toward the main party, the opposite party*, *media*, and *government*, respectively. Each opinion is a numerical variable which ranges from -1 to 1, and positive/negative value represents positive/negative opinion polarity.

Actions. While the states of the five parties are a bit similar, their actions are rather diverse. Among all the possible actions of one party, we are particularly interested in those actions which have impact on the attitude/opinion of one agent toward others because our model aims to capture the dynamics aspect of a netizen-centered Internet event. An action has corresponding *target* and *intensity*. Target is the object of the action and can be one of the five agents defined. If an action has no target, the event itself is the default target. Intensity degree could be low, medium or high.

Actions of the main party and opposite party are the same. They consist of a number of actions, which are classified into positive actions and negative actions. Positive actions include those actions like *donation*, *compensation*, *surrender*, *praise*, etc. Negative actions include actions such as *prosecution*, *accusation*, *criticism*, *blackmail*. Actions of netizen group are also divided into positive actions and negative actions. Positive actions can be online or offline actions such as *help*, *support* or *digging* target's positive news. Negative actions are actions like *threat*, *accusation*, and *digging negative news*. Each action has its target and intensity degree. Media's actions include *praise*, *criticism* and *report-in-view*. Praise/criticism means media writes columns or comment articles to praise/criticize target. Praise and criticism are media's subjective opinions on the event. Report-in-view denotes media reports positive or negative news about target. Report-in-view is media's objective opinions on the event.

Government has five actions: *get-involved*, *judge-winner*, *appeasement*, *award*, *punishment* and *no-response*. Each action has intensity and target. Get-involved indicates government starts to get involved in the event. The intensity can be low, medium or high. The intensity is determined by the type of government, for example, the intensity is low if local government is involved. Judge-winner means government judges the target as winner after investigation, and the target is null when no party wins. Appeasement represents that government gives compensation to the target. Appeasement and award are positive actions, and no-response and punishment are negative actions.

Interaction rules. Interaction rules specifies how one party's states will change in response to another party's actions. For the main party, for example, the opinion/attitude toward the opposite party would decrease if the benefit of the main party is negative or the opposite party takes negative actions. It is also affected by government's judgement. The opinion toward the media is affected by the media's action. If the media reports positive news of the main party or negative news of the opposite party, the main party's opinion toward the media will increase. The opinion toward the government is affected by the government's action. If government takes negative actions against the main party, the main party's opinion toward the government will decrease.

With the increase of the conflicts between main party and opposite party, the concerns of media and netizen group will increase. Actions from other parties will increase concerns as well, such as criticism from netizen group to government. Negative actions often have greater influence than positive actions. This is probably due to the negativity bias according to social psychology [16]. If no action occurs, the concern will decrease with time following the power law, as found in social computing research.

Opinion of a netizen group can be influenced by other parties' states, for example, opinion toward the main party/opposite party is affected by the party's benefit. Actions from other parties can also change the opinion of the netizen group. For example, the opinion of the netizen group toward the main party will increase if the government judges the main party as winner. The netizen's opinion toward the government would decrease if government does not respond to the event in certain time limit or the investigation results/the actions of the government are contrary to the netizen's expectation. During the event, the netizen group's opinion will decay over time [17, 18]. Rules for the opinion of the media to the other parties are similar to those of the netizen group.

Below we choose two typical Internet events as case studies to test our proposed model.

4 Case Studies

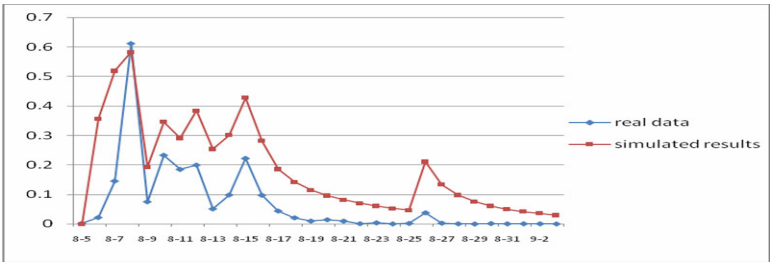
To illustrate the usefulness of our model, we test our model by simulating two typical Chinese Internet events occurred in 2010, i.e., *Synutra event* and "*360 vs. QQ*" event. We select Repast [19] as the agent-based modeling tool. In each scenario, we first initialize the model by setting event type and initial states of all the parties according to the related online news. The actions are directly transferred from real actions of all the parties in the event. The default action of every agent is no action and the default value of benefit is null. These parameters are extracted from text data by hand, and each time step corresponds to a day in the real world event. After running the model step by step, we get the netizen group's concern and opinions as the output of the simulation. To test the model's performance, the simulation results are then compared with results mined from real-time social media data using our sentiment analysis tool [20].

4.1 Synutra Event

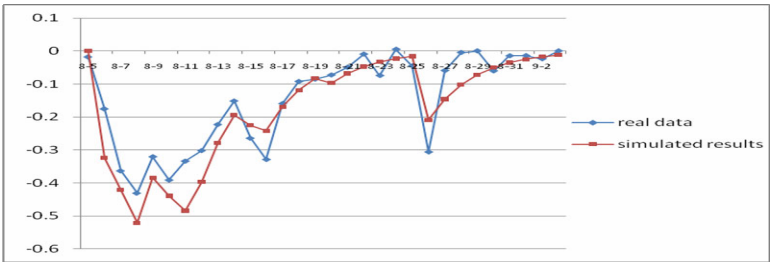
Synutra event is a food safety event happened in August 2010, in which Synutra baby milk powder was suspected to cause precocious puberty. Synutra event had drawn national attention and governmental investigation. At the beginning, three girls in Wuhan who had been fed with Synutra milk were suspected to be premature. Synutra denied the report later. Customers and large numbers of netizens do not believe the announcement given by Synutra and joined together to denounce Synutra. Finally Chinese Health Ministry was forced to initiate investigation. The investigation results showed that three girls were pseudo-precocious puberty. In this event, netizen group played an important role in the evolvement of the event.

We extract news reports and comments about Synutra event from three popular Chinese news sites, including QQ, 163, and Phoenix. Here we get 153, 104, and 580 reports, and 58307, 12146, and 13419 comments from QQ, 163, and Phoenix respectively.

In Synutra event, customers act as the main party, Synutra company is the opposite party and Health Ministry is the government. In the scenario we model the event from Aug 6 to Sep 3. Fig. 2 shows the comparison of simulation results and real data. In the figure x axis represents the date, and y axis denotes the positive/negative polarity of netizen group's opinion or the netizen group's volume of comments.

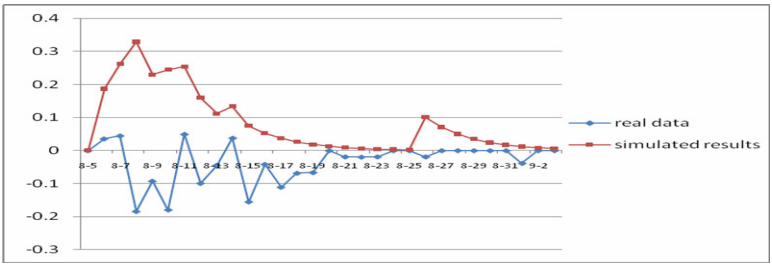


(a) Netizen group's concern

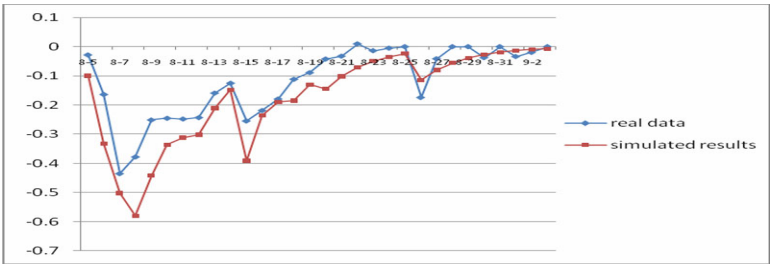


(b) Netizen group's opinion toward Synutra company

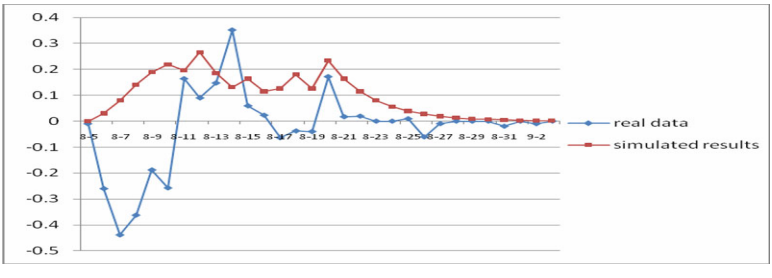
Fig. 2. Comparison between simulation results and real data



(c) Netizen group's opinion toward the customers



(d) Netizen group's opinion toward the government



(e) Netizen group's opinion toward media

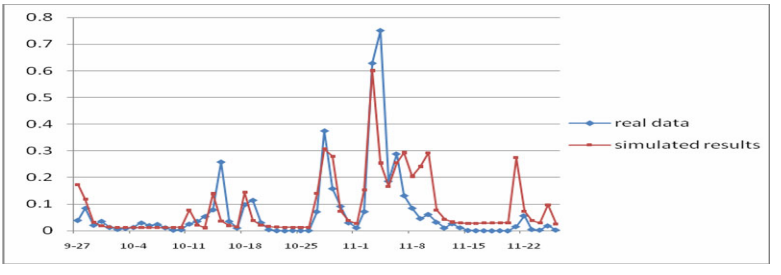
Fig. 2. (*continued*)

The curve of the simulated netizen group's concern matches well with the curve of the total amount of the comments, as shown in Fig. 2(a). From Fig. 2(b) we can see that the netizen group's negative opinion toward Synutra is rather strong on Aug 8 because a lot of victims of the Synutra milk powder are found. Fig. 2(d) shows the simulated opinion toward the government is similar to the real data. The opinion toward the government decreases at the beginning as the government refused to react to the event and thus raised the public rage. After the join of the Health Ministry, the opinion increases gradually. However, it drops suddenly on Aug 15 because the investigation results were not consistent with netizens' anticipation and they doubted the government's fairness. In general, the simulation results by our model fit the real social media data quite well. This basically verifies the effectiveness of the model.

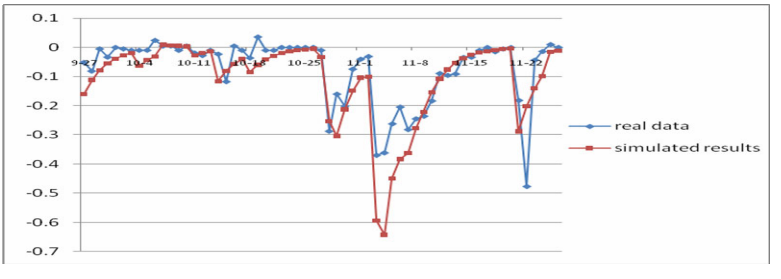
4.2 360 VS QQ

360 versus QQ is a business conflict between Qihoo and Tecent, two leading software companies in China. 360 published privacy protection software to monitor QQ software on September 27. Tecent announced QQ software would not compatible with 360 on personal computers. Netizens got rage soon and expressed their anger by posting threads online. The event ended after Industry and Information Technology Ministry and Public Security Ministry's engagement.

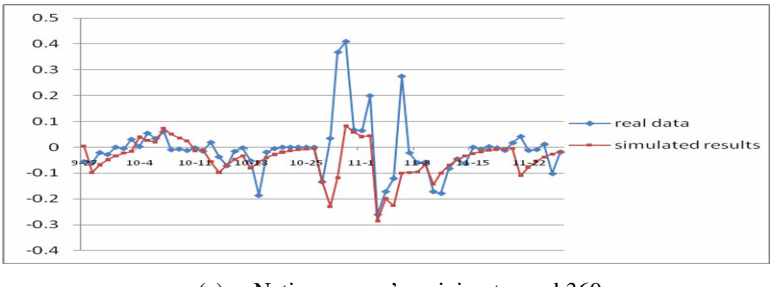
We downloaded 352 news reports and 169741 comments from 163. The y axis in Fig. 3(a) denotes the number of daily comments. The curves in the other figures show the netizens' opinions toward QQ, 360, the government and the media. We got these data according to the same procedure described in Synutra event case.



(a) Netizen group's concern

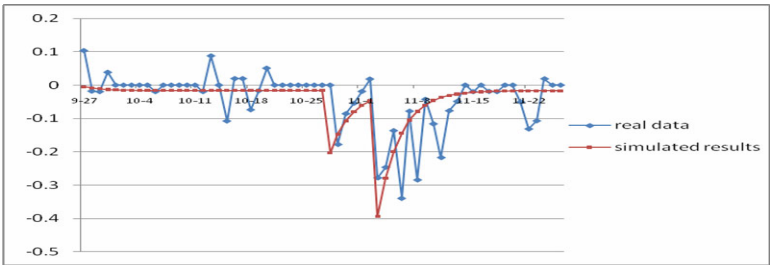


(b) Netizen group's opinion toward QQ

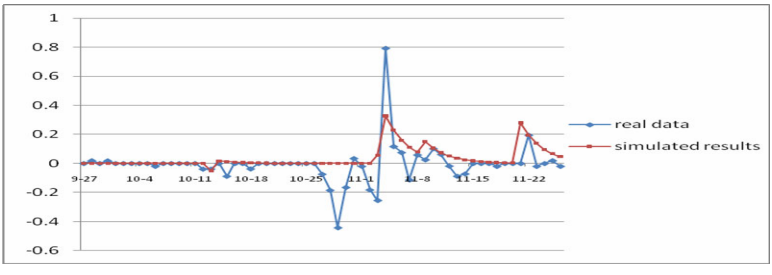


(c) Netizen group's opinion toward 360

Fig. 3. Comparison between simulation results and real data



(d) Netizen group's opinion toward the government



(e) Netizen group's opinion toward media

Fig. 3. (*continued*)

In this scenario, 360 is the main party, QQ is the opposite party, and Ministry of Industry and Information Technology and Ministry of Public Security are the government. Fig. 3 compares the simulated netizen group's concern and opinions with the real media data from Sep 27 to Nov 26. X axis and y axis denote the same meaning as those in Fig. 2.

Fig. 3(a) shows that the curve of the simulated concern matches well with the real data. In Fig. 3(b), it reaches its lowest point on Nov 4 because QQ exerted pressure on netizens and got criticism from media and netizens on that day. We can see from Fig. 3(c) that the simulated opinion toward 360 matches the real data well. In Fig. 3(d), the decreases on Oct 28 and Nov 3 are due to the direct conflicts between the two companies. In Fig. 3(e), opinion toward media in real data decreases on Oct 28 and Nov 3 for the same reason. The simulation results by our model generally fit the real data, which shows the model's effectiveness.

5 Conclusion

This paper proposes an agent-based model of netizen groups in Internet events. The proposed model considers all the main parties involved in an Internet event and captures how their interactions impact their opinions toward others. We focus on studying the evolution of netizen group's concern and opinion in the event and netizen group's interactions with the other parities. We select two typical Chinese Internet events as case studies and compare the experimental results with real social media data to verify the effectiveness of our model. This work contributes to the study of

social behavior in social computing [21]. Our future work shall explore the use of information extraction techniques to automatically extract media data and facilitate model construction.

Acknowledgement

This work is supported in part by the National Natural Science Foundation of China under Grant Nos. 60875028, 60875049, 60921061 and 90924302.

References

1. Grimm, V., Revilla, E., Berger, U., et al.: Pattern-Oriented Modeling of Agent-based Complex Systems: Lessons from Ecology. Science 310(5750), 987–991 (2005)
2. Kaiser, J.: BIODEFENSE: Senate Bill Would Alter Biosafety, Select Agent Rules. Science 320(5883), 1573–1573 (2008)
3. Zacharias, G., MacMillan, J., Van Hemel, S.B.: Behavioral Modeling and Simulation: From Individuals to Societies. The National Academies Press, Washington, DC (2008)
4. Jiang, M.: Chinese Internet Events. In: The Internet in China: Online Business, Information, Distribution and Social Connectivity. Berkshire Publishing, New York (2010)
5. Synutra event (2010), http://www.bbc.co.uk/news/world-asia-pacific-10978702
6. 360 v. QQ (2010), http://en.wikipedia.org/wiki/360_v._Tencent
7. Hughes, A.L., Palen, L., Sutton, J., et al.: Site-seeing in Disaster: An Examination of Online Social Convergence. In: Proceedings of the 5th International ISCRAM Conference, Washington, USA (2008)
8. Sutton, J., Palen, L., Shklovski, I.: Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires. In: Proceedings of the 5th International ISCRAM Conference, Washington, USA, pp. 624–632 (2008)
9. Vieweg, S., Palen, L., Liu, S.B., et al.: Collective Intelligence in Disaster: Examination of the Phenomenon in the Aftermath of the 2007 Virginia Tech Shooting. In: Proceedings of the 5th International ISCRAM Conference, Washington, USA, pp. 44–54 (2008)
10. Blood, R.: How Blogging Software Reshapes the Online Community. Communication of the ACM 47(12), 53–55 (2004)
11. Zhang, H., Giles, C.L., Foley, H.C., et al.: Probabilistic Community Discovery using Hierarchical Latent Gaussian Mixture Model. In: Proceedings of the 22nd National Conference on Artificial Intelligence, Vancouver, British Columbia, Canada (2007)
12. Golbeck, J., Parsia, B., Hendler, J.: Trust Networks on the Semantic Web. In: Proceedings of Cooperative Intelligent Agents, Helsinki, Finland (2003)
13. Lim, E.P., Vuong, B.Q., Lauw, H.W., et al.: Measuring Qualities of Articles Contributed by Online Communities. In: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, Hong Kong, pp. 81–87 (2006)
14. Wang, F.-Y., Zeng, D., Hendler, J.A., et al.: A Study of the Human Flesh Search Engine: Crowd-Powered Expansion of Online Knowledge. Computer 43(8), 45–53 (2010)
15. Nolker, R.D., Zhou, L.: Social Computing and Weighting to Identify Member Roles in Online Communities. In: Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence, Washington, DC, USA (2005)

16. Niu, X.: Psychological Effects Analysis of the Dissemination of Internet Hot Events. Chinese Journal of Voice and Screen World 2009(1), 36–38 (2009)
17. Chen, Y., Yu, J.: Mechanisms of Mass Emergency's Occurrence, Response, and Prevention, and Public Opinion. Journal of Social Science Review 21(7), 135–137 (2006)
18. Plutchik, R.: The Nature of Emotions. American Scientist 89(4), 344–350 (2001)
19. Collier, N.: Repast: An Extensible Framework for Agent Simulation. The University of Chicago's Social Science Research (2003)
20. Zhang, C., Zeng, D., Li, J., et al.: Sentiment Analysis of Chinese Documents: From Sentence to Document Level. Journal of the American Society for Information Science and Technology (JASIST) 60(12), 2474–2487 (2009)
21. Wang, F.-Y., Carley, K.M., Zeng, D., et al.: Social computing: from social informatics to Social Intelligence. IEEE Intelligent Systems 2(22), 79–83 (2007)