Approach to Communication Analysis Based on Topology and Viewpoint from Social Network

Hongming Shu

No.2 High School of East China Normal University, Shanghai, China

Hzhc_wjdi@163.com

Abstract: Today in technical communication network, rapid dissemination of a piece of news, in part, is because of the intrinsic value of the news itself, or in part, of the network node of influence or special place it reaches. Shared information, a culture characteristic, has been existing. However, dissemination of news never becomes so simple and wide as it does now. We attempt understanding the evolution of the methodology, purpose and functions of society information network. We mainly consider of three problems: 1. We define news to judge what news can be valued and disseminated in society. 2. A model is built for society information network, with which we analyze the influence of news characteristics on its dissemination. 3. Based on the model of the second problem, we try to explain how the news at different stage of human history is disseminated and the characteristics of the dissemination. 4. Based on the model of the second problem, we judge the effect of prejudice of people on the dissemination of news. Of them, the last three problems represent three elements that influence the dissemination of news: the news itself, viewpoint of people and the characteristics of the information network itself. Concentrating on these three problems, we find representative sets of data. Then we analyze the news source with the approaches of cellular automaton and data fitting, and the effect of the intrinsic value and strength of information network on the dissemination of news on Internet.

1. Introduction

In modern society, information of various types is disseminated rapidly on Internet, of which some are about the important changes in national policy, others are the rare natural wonders and still others are a picture posted by a celebrity. A piece of well-known information may because its content is attractive, or because it is released by an information source that arouses wide concern. Different people have different reaction to the same information. Process in communication technology makes the information, especially the category with excellent timeliness——mode of transmission of news undergoes a huge change. At present, a number of scholars have conducted study on the propagation model of information networks accordingly. Akshay and other scholars [1] pointed out in their articles the relations between Weibo network and geographic position in the characteristics of network structure. Generally speaking, users with adjacent geographic position in network are closely linked in Weibo network. Huberman, B.A [2] and others found in their research that in social network, users can only make information exchange and interaction with few people in their friend list under generation condition. So, they have important significance than other users in dissemination of information on social network. In the meantime, the author also defined the frequently-used concepts of social network, of which the relevant definition is applicable on Sina microblog and China Twitter. Teutle and other scholars conducted research into Twitter from perspective of the characteristics of dynamic evolution of the social network [3]. They described the evolution of Weibo network with the network parameter such as growth of user node degree, network density and betweenness on social network. Honey studied the information contents of the Twitter, especially the symbol of "@". The research findings proved that Weibo information has an

obvious effect on strengthening the exchange of information among users. Meeyoung Cha and others analyzed propagation characteristic of Twitter [4]. They compared in-degree of user node, reposted amount on Weibo and reference count on Twitter network from various perspectives to assess the influence of user node on the message propagation process. The findings show that the number of fans, message reposting and reference count does not present a directly proportional relationship. This conclusion indicates that the number of fans of the users does not present obvious relations with the users influence. However, users with great influence indeed can create hot topic in reality.

The paper aims to analyze the news dissemination status on the social information network and the factors that may influence the news transmission, which thus involves a simulation of the transmission. The author designed a model to describe the dissemination of news on social information network, which manifests intuitively mode of transmission and propagation path of the information.

2. News value and its changes with time

A great deal of news is disseminated on social information network each day. Today in explosive growth of the social network, a myriad of people shares what they see and hears on Internet. There are huge differences in the values for different news on the information network. The factors that affect the news value include:

1) News contents: news contents are the factors to differentiate different news, which is also one of the important influence factors in news value. News contents include the scarcity of the news, news report contents and level of detail. We believe that news with low probability of occurrence is easily widely spread.

2) Geographical and cultural factors of information network: the same news has different dissemination status on different information network, especially with undeveloped means of communication. For example, news in a region has high value in local news network while has low one out of the geographical factor.

3) Added value of news source: people make different reactions to the same news released by different sources. For example, news posted by a celebrity and politician is widely spread than an average person does. The same things happen in a celebrity will raise concern.

4) Timeliness of news: timeliness of news is an important basis for us to classify a piece of information into a news, an important factor to affect the dissemination of news. As time goes by, people's interest in a specific piece of news will come down rapidly, with a declining possibility of continuously propagating and reposting it.

According to the above exposition, we can list the empirical formula for the news value as below:

$$V = \alpha I_o I_e e^{-\beta t}$$

Of the formula, α is the correction term for geographical factor, I_o is the added value of news source, I_e is the intrinsic value of news, t is the time for its dissemination and β is its timeliness coefficient.

3. Dissemination of the analog information of cellular automaton on social network

3.1 Cellular automaton model

Cellular automaton is a local dynamical model of discrete space-time [3]. It is made up of a number of table cells, each of which is at one of the limited possible status. So, we should simulate a model which has independence and connection among individuals. In our problems, information network is made up of a large number of nodes, each of which represent a man, with its own characteristic, including hobby and political attitude. In the meantime, there is specific link among nodes, namely, the social information network. During the process of the dissemination of news, the node on information network is merely affected by its adjacent node state, according with the characteristics of the localization of cellular automaton. Moreover, because news has strong

timeliness, the simulation process can be perceived as closely approximating the propagation process. During the process, information network presents no change, according with the characteristics of the cellular automaton. So cellular automation can be used to simulate the dissemination of news on the information network.

When simulating the dissemination of news on information network, we perceive the square adjacent to the node as its adjacent node in the picture. The node state includes not understanding the news, understanding it without reposting and understanding and reposting it. The process to simulate the cellular automaton with computer is divided into several stages: in the initialization phase, original state is allocated for each table cell. In social information network, each time when we think about the dissemination of a piece of news, its news source at initial phase has received and reposted it, while all of the rest nodes do not understand it. In the meantime, the unique characteristic of each node will affect the dissemination of information. In each round of simulation below, the state of each table cell is changed according to the status of the table cell adjacent to it in the last round with a set of rules. In dissemination of news involves two states, namely, "not understanding the news-understanding but not reposting it" and "understanding but not reposting it-understanding and reposting it". In each round of simulation below, for the two nodes that understand the news state in the last round, all of their adjacent nodes that do not understand the news is changed with the reposted probability p, that is, understanding but not reposting it.

3.2 Reposted probability

Another key factor to affect the dissemination, besides type of network, is reposted probability. There are many factors to influence a piece of news being reposted. When a piece of news is probably reposted from reposted node f to receiving node t, we should consider the effect from the following factors. Suppose each node of the network post message at the same rate, the reposed probability p is in inverse proportion to number of degrees d_t of the receiver on the network. According to the conclusion of model one, we can write the timeliness term for news $e^{-\beta_f t_f} e^{-\beta_s t_s}$, of which t_s is the time a news source posts the news, t_f is the time the reposting node post the news, and β_f and β_s are respectively the corresponding timeliness coefficient.

In the meantime, the authority of the news and its intrinsic value will influence its dissemination. So, we can add two coefficients to the probability p, namely, number of degrees of the news source d_s and the intrinsic value of news v. As a result, we can write the reposted probability of the news as follow:

$$p = vk \frac{d_s^{k_s}}{d_t} e^{-\beta_f t_f} e^{-\beta_s t_s}$$

In the formula, k is a constant.

3.3 Analysis into the influence factors

First, we analyzed the influence of news source on dissemination of news. So, we controlled other variables by only changing the start node of the news. We paid close attention to curve where the number of people who know the news and those who repost it is changed with the time. Because the simulating algorithm we used has randomness, we averaged through repeatedly simulation to reduce the error. As shown in the picture below, we compared three start nodes where the number of people who know and repost the news change with the time. It clearly shows that the news source more people pay attention to, the faster the dissemination of news will be (the bigger the slope will be), and more people will be to know and repost the news (the bigger the value of the trend will be). In a word, news from important news source is easily disseminated, with greater range of influence.
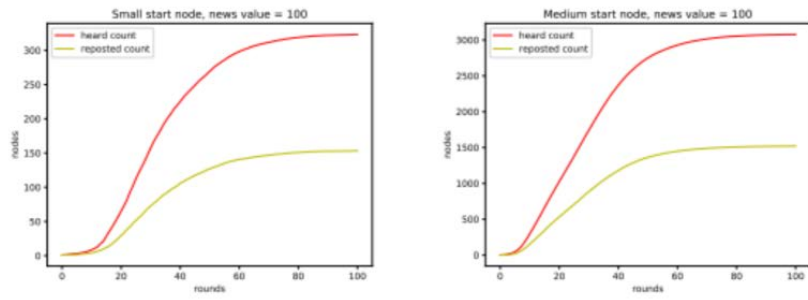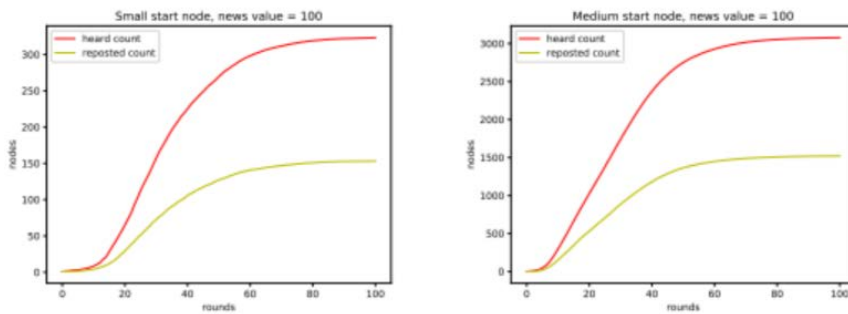
(a)Few concern about the news source (b)Moderate concern about the news source
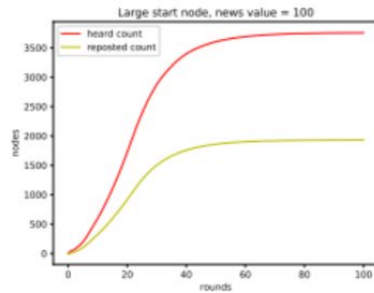
(c)Considerable concern about the news

Fig.1 Heat changes of Weibo with the time

Similar to the last analysis, the intrinsic value of the news was changed with controlling the variables. The influence of intrinsic value of the news is similar to that of news source. However, the former has greater effect than that of the latter.

(a)Few concern about the news source (b)Moderate concern about the news source
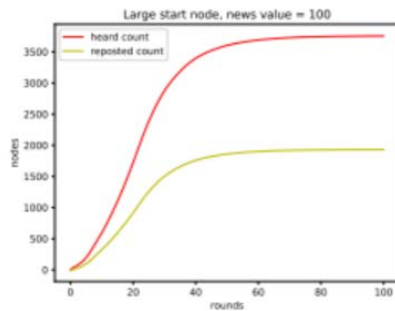
(c)Considerable concern about the news

Fig.2 Diagram for changes of news spreading scope of different intrinsic value with time

4. Influence of simulation viewpoint on dissemination of news

4.1 Viewpoint vector

When the Internet come into being, the influence of geographical factor on dissemination of information become increasingly small. The territory on Internet is manifested in another form. Viewpoint, hobby and prejudice exert great influence on the dissemination of news. According to the uses and gratification theory [14], the users would have an initiative selection of their consumption information. The emergence of Internet unprecedentedly increases the amount of information people can come into contact with, which makes the selection easier. Therefore, the territory on Internet is formed based on interest and viewpoint. Such a territory makes the viewpoint concentrated, with significant effect of initial viewpoint on dissemination of information.
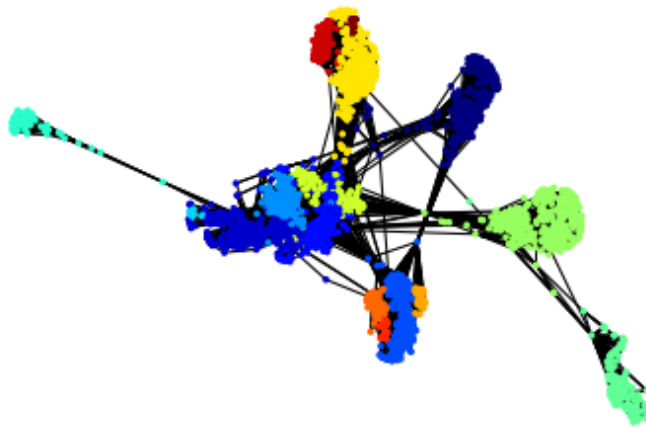

Fig.3 Classification of partitioning algorithm for community used in Facebook data set

To analyze the influence of viewpoint on dissemination of news, we used the discovery algorithm of Lauvain community to divide the community of information network, as shown in Fig.3:

We randomly generated the viewpoint vectors accordingly for each community, and guaranteed the length for each of them is 1. When the news is propagated to a node, the viewpoint in the node will help the dissemination of news no matter it is agree with or oppose to the news. So, we can add the viewpoint term p_c in the reposted probability, the formula is:

$$p_c = \overrightarrow{A_s} \cdot \overrightarrow{A_t}$$

Of the formula, $\overrightarrow{A_s}$ means the viewpoint vector of the news source while $\overrightarrow{A_t}$ is that of the receiver. Thus, the revised reposted probability p' is:

$$p' = p \cdot p_c = vk(\overrightarrow{A_s} \cdot \overrightarrow{A_t}) \frac{d_s^{k_s}}{d_t} \mathrm{e}^{-\beta_f t_f} \mathrm{e}^{-\beta_s t_s}$$

We used the revised reposted probability, and simulated the reposted news in the model in chapter two.

4.2 Analysis into the influence of viewpoint

As shown in the following picture, it is easy for the news agreeing with the mainstream opinion to be widely spread, which is disseminated faster and widely in the community with the same viewpoint.
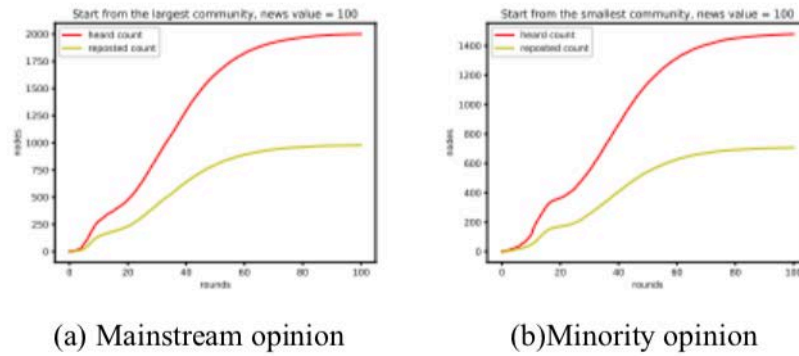
(a) Mainstream opinion (b)Minority opinion

Fig.4 Dissemination state under the influence by viewpoint

5. Rationality and validation experiment

5.1 Rationality of the validation formula

Because other factors in the formula have strong subjectivity, time t was selected as variable for validation. Because heat is an abstract concept, repost frequency (reposted times within unit time) can serve as the similar reposted heat. To be specific, reposting can be divided by equal interval according to the time quantum, where reposted amount was counted for each time quantum. And then we conducted exponential function fitting for the data and the above-mentioned data. Because of their strong subjectivity, It is difficult to differentiate α, I_o and I_e , so they are grouped into term α. In this way, we can simplify the fitting function to $V = \alpha' e^{-\beta t}$.

Here we selected a few reposted Weibo with largest reposted amount to be the analyzed object. They repost frequency and fitted function is shown as the following picture.

(a) Repost frequency of weibo I/II one changes with the time

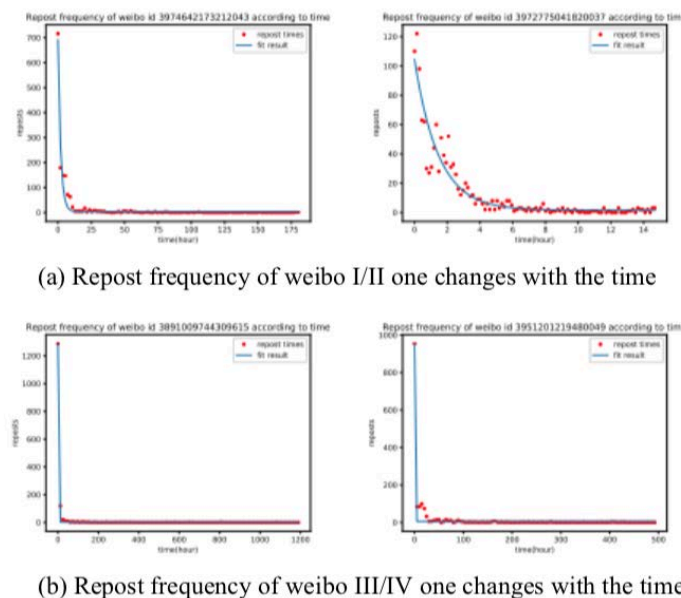(b) Repost frequency of weibo III/IV one changes with the time

Fig.5 Changes of Weibo heat with time

A special example is shown as Fig.5. The picture evidently can not be well described by prime formula. It is found after observation that the data is actually the results from the stack of repeated dissemination climax. There are many reasons for the phenomenon, such as reposting by an Internet celebrity and fermentation of news once again. For this state, we extend the formula:

$$V_{turns} = \sum_i (t - t_i)\epsilon(t - t_i)$$

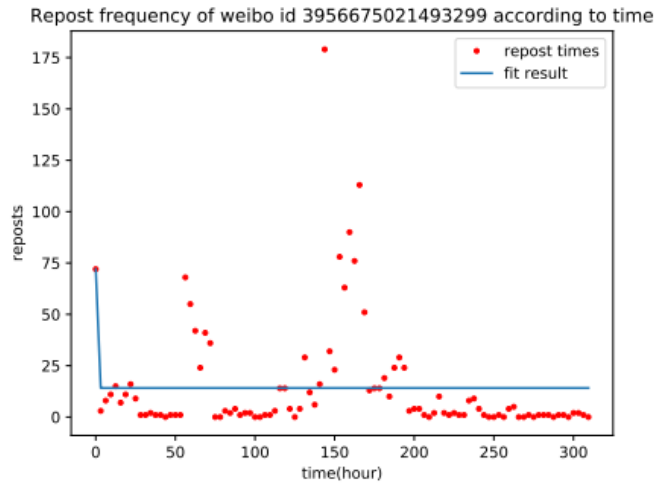In the formula, t_i is the time for the advent of a few climaxes.

Fig.6 Diagram for Weibo reposted heat not conforming with the prime formula

In conclusion, our empirical formulas accord with the physical truth in most instances, with universality.

5.2 Analysis into typical characteristic change

To simulate the truth as far as possible, we randomly deleted some sides according to specific proportion based on the previous information network of Facebook before operating simulation. The results are shown as the picture below. Obviously, the stronger the information network intensity it is, the faster and widely the news will be disseminated.

(a) weak network intensity

(b) moderate network

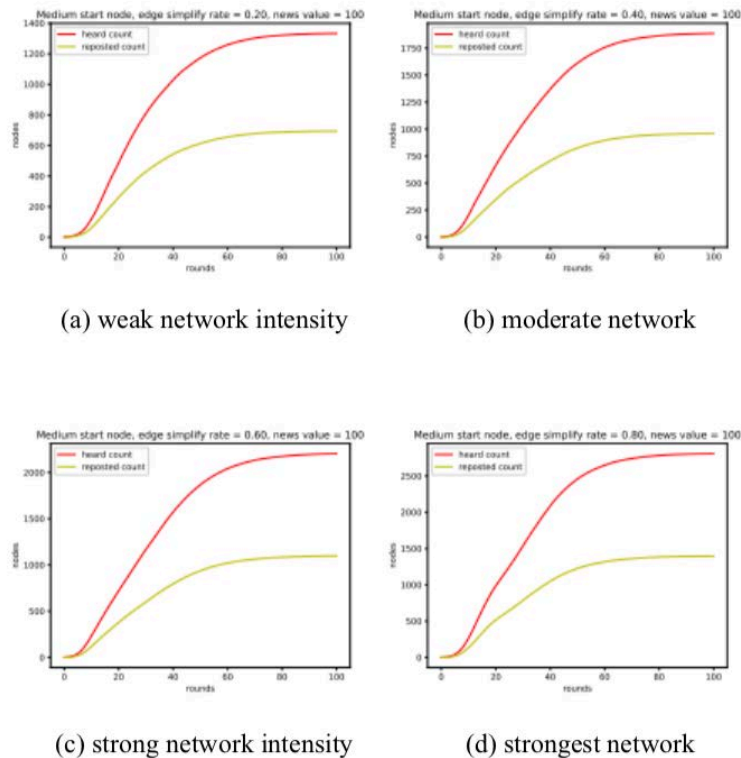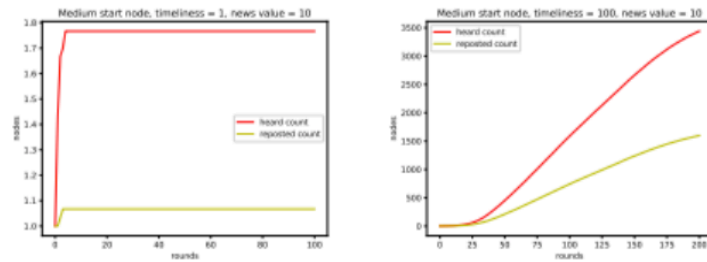(c) strong network intensity

(d) strongest network

Fig.7 Dissemination of news with different network intensity

To analyze the influence of propagation speed of news on the dissemination of news, we made some modification based on the model two. The goal node will understand the news after it is reposted for the next round of simulation, which is changed to after n round of a specific round.

The news value is closely linked with time. So obviously, the slower the news is disseminated, the fewer rounds there will be to lose its value. As shown in the following picture, if the news is

disseminated slowly, its dissemination will be stopped for a few rounds, with limited propagation range.

(a) slow dissemination of the news   (b) rapid dissemination of the news

Fig.8 Dissemination of news with different network intensity

6. Conclusions

In the paper, we analyzed various influence factors for dissemination of news from three aspects, that is, the news itself, viewpoint of people and information network. Then we proposed some empirical formulas based on facts, and used real data to validate the formula. With cellular automaton, we simulated the influence from many conditions. Especially in terms of the effect of technical progress and viewpoint of people on dissemination of news, we draw a number of interesting conclusions:

1) News with important news source is easily disseminated, with greater scope of influence;

2) News agreeing with the mainstream opinion is easily widely spread. It is propagated faster and widely in a community with the same viewpoint;

3) The stronger the information network intensity, the faster and widely the news will be disseminated;

4) If the news is disseminated slowly, its dissemination will be stopped within few rounds, with limited propagation range.

A great deal of data from social network involves the privacy, which is difficult to access. So our conclusions are based on many hypotheses. Because of current restriction, the author of the paper wants to conduct an in-depth research into some interesting analyses in the future, such as the relations between intrinsic value of news and its contents, to further perfect the propagation model for information on Internet.

References

[1] Akshay J, Xiaodan S, Tim F, et al. Why We Twitter: Understanding Microblogging Usage and Communities [A]. In: Proceedings of the 9th Web KDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis [C]. San Jose, California, 2007: 56-65.

[2] Huberman B A, Romero D M, Wu F. Social Networks That Matter: Twitter under the Microscope [J]. FirstMondy. 2009, 14(1): 8-16.

[3] Teutle.Twitter: Network Properties Analysis [A]. In: Proceedings of 20th International Conference on Electronics, Communications and Computer (CONIELECOMP) [C], Cholula, 2010: 180-186.

[4] Cha M, Haddadi H, Benevenuto F, et al. Measuring User Influence in Twitter: The Million Follower Fallacy [A]. In: Proceedings of International AAAI Conference on weblogs and Social Media(ICWSM) [C], Washington, DC, 2010: 10-17.

[5] Gruhl D, Guha R, Liben-Nowell D, et al. Information Diffusion through Blogspace [A]. In Proceedings of the 13th International Conference on World Wide Web (WWW '04) [C]. New York, 2004: 491-501.

[6] Kempe D, Kleinberg J, Tardos É. Maximizing the Spread of Influence through a Social Network [A]. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '03) [C]. New York, 2003: 137-146.

[7] Goldenberg J, Libai B, Muller E. Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-mouth [J]. Marketing Letters, 2001, 12(3): 211-223.

[8] Lahiri M, Cebrin M. The Genetic Algorithm as a General Diffusion Model for Social Networks [A]. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence [C]. Atlanta, Georgia, USA, 2010:494-499.

[9] SNAP: Network datasets: Social circles, https://snap.stanford.edu/data/egonets-Facebook.html