Information propagation in social network

Jianping Qiu[1*], Lichao Chen[1], Guifang Chen[2]

[1]*Institute of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China*

[2]*Computer Science and Technology Department, Taiyuan Normal University, Taiyuan 030012, China,*

Received 1 March 2014, www.cmnt.lv

Abstract

Information propagation network analysis provides a new way to investigate online activities. From the perspective of information propagation analysis we can understand it in a constant evolving way, that is, the content of the information is modified by the netizens with a certain probability during the whole propagation process. By analysing the online behaviour of netizens, we constructed an information propagation network on social networks. In this paper, we found that the original information can keep its influence on the netizens only when most of them are forwarders. Meanwhile this paper reveals influence propagation is aggregated, for example, netizens tend to give a low rating after a low rating, as well as a high rating following a high rating. Our findings are helpful in better understanding information propagation.

Keywords: information propagation, social network, scale-free network, power-law distribution, influence propagation

1 Introduction

With the rapid development of social networks, more and more people are using social networks to obtain information [1]. We no longer need to physically enter a library to obtain the latest news or to read a scholarly journal. A simple search through any computer or mobile device is enough to put at our disposal not only what we search for but also a trove of related findings that increase our curiosity and expand our horizons. Add to that the ubiquity of e-mail, instant messaging and Microblogs, and we find ourselves in a world of instant connectivity and potentially productive connections with social networks across the globe.

A social network is a social structure made up of a set of social actors (such as individuals or organizations) and a complex set of the dyadic ties between these actors, and is main tunnel of information propagation [2]. P. S. Dodds and D. J. Watts studied the accumulation effect of information propagation and E. Agliari et al took into consideration the degeneration of information on a spatial system. In social networks, e.g., scale-free networks [3], the information propagation process is much more complicated than the ordinary scenarios. The information changes constantly in its propagation process. The behaviour comes from the cumulative modifications during the propagation process however, whether the information can spread through the whole network depends not only on the existence of the connections among nodes, but also on their strategies. We divide the strategies into two types: one is to forward information directly; the other is to modify information before spreading it out. In this paper, we investigate information propagation in social networks, the netizens forwarding the information are named as forwarders. The netizens modifying the information are named as modifiers.

2 Related work

In 1967, Stanley Milgram is credited with introducing the notion of a small-world network to the social science community. Milgram's famous "six degrees of separation" experiment suggested that the distance between two people selected at random from the entire population of the United States is approximately six intermediaries. In 1999, Barabasi created another line of investigation with the invention of scale-free networks (non-random networks with hubs). In a number of studies of the structure of the Internet and WWW, Barabasi et al. discovered an emergent property of the decentralized Internet that it had emerged without central planning into a structure consisting of a small number of extremely popular sites called hubs, which have more influence, and a large number of "unpopular" sites with few links. Instead of being random, the Internet topology was very non-random. In fact, the probability that a site has k links obeys a power law, which drops off quickly for large k. Furthermore, they speculated that this was the result of a microrule called preferential attachment that the probability a site will obtain a new link is directly proportional to the number of links it already has. Thus, the more links a site has, the more it gets the so-called "rich get richer" phenomenon.

Figure 1 shows the flow web of the 1000 most-visited sites [4]. Solid circles represent websites and edges show information propagation flow. The size of circles is proportional to the logarithmic value of their flow. The red circles is web 2.0 sites, the blue ones is web 1.0 and the white ones is search engine. As we can see from Figure 1,

[*] ***Corresponding author*** e-mail: 920868329@qq.com

netizens prefer web 2.0 to web 1.0.In Figure 2, the blue circles is statistics of the 1000 most-visited sites flow, the red line is analogue value of Zipf law. The reasons why there appear the phenomenon of a significant cut-off are lack of lower-ranking websites' data and rich-get-richer paradigm [5].As shown in Figure 2, the Internet is the power-law distribution which $\gamma < 1$.

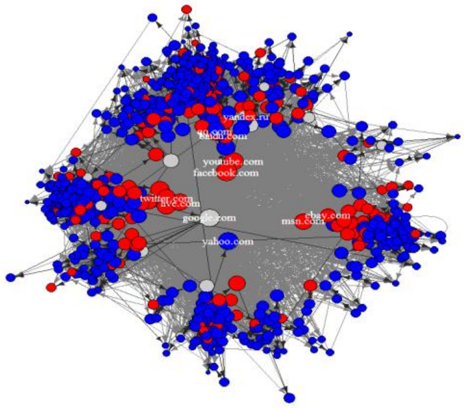

FIGURE 1 The network of top 1000 sites worldwide

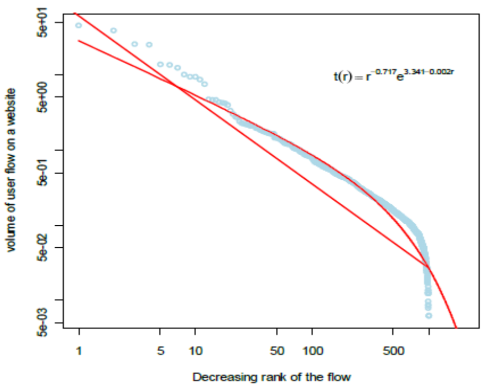

FIGURE 2 The flow distribution of top 1000 sites worldwide

3 Information propagation network

A information propagation network is a (directed or undirected) network, $G(t)=\{N(t), L(t), f\}$, where $N(t)$ is a set of actors (nodes) with time-varying state, $L(t)$ is a set of influence degree ϕ for information, and f is a static mapping of $N \times N$ that defines the topology of G. Nodes and links have a value property that defines influence degree (for links) and position (for nodes) associated with a proposition. Further, let $S(t)$ be the state vector representing an actor's position, where $-1 \le S(t) \le 1$ and $S(t) = [s_1, s_2, \cdots, s_n]^T$ = *state vector of* G. The rate of change in the state of a netizen is dictated by the difference between the states of adjacent netizens, State Equation is as follows:

$$s_i(t+1) = s_i(t) + \sum_j [s_j(t) - s_i(t)]\phi^T = s_i(t) + \sum_j \Delta s_i(t). \quad (1)$$

In this model, a netizen changes his or her position by an amount $\Delta S_i(t)$ after each interaction. A necessary (but not sufficient) condition for reaching a consensus in a communication network is $\sigma(L) < 0$, that is, a negative spectral gap, where σ is the largest nontrivial eigenvalue of L. At this moment, $\Delta s_i(t) = Ls(t)$, where L is the Laplacian of ϕ^T, as shown in Equation 2.

$$s_i(t+1) = s_i(t) + Ls_i(t) = [I+L]s_i(t) = [I+L]^{'} s_i(0), \quad (2)$$

Our network can model the consequences of information propagation. Let s be a measure of stubbornness over netizen. In Figure 3, $s = 0$ means the netizen who disseminates the information directly to their neighbours denoted by a circle; $s \ne 0$ means the netizen delivers the information after modifying denoted by a square. So the content of the information is modified by the netizens with a certain probability during the whole propagation process.

When receiving a certain version of the information, the netizen (the receiver) becomes a disseminator (denoted by white colour in Figure 3) of the version. When receiving two or more different versions of the information, the netizen accepts the latest version. If a netizen receives the original information or the revised versions, which he sent before, they will turn to be a terminator (denoted by black colour in Figure 3). This is simply because once the disseminator has disseminated similar information to their neighbours, this means there is no need to do that again. On the other hand, the neighbours who send the versions of the information to them would not be interested in the similar information as well.

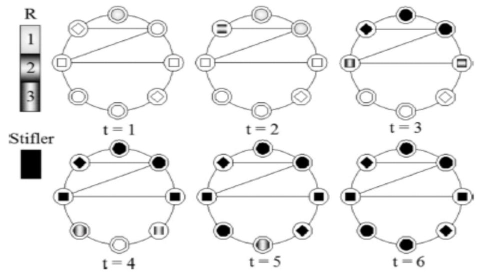

FIGURE 3 Information propagation network. *R* denotes the revised frequency of the information

4 Information propagation model

In our spreading model, netizens can play three roles: receivers, disseminators, and terminators, whose densities are denoted by $r(t)$, $d_x(t)$ and $e(t)$. Here, x denotes the version of information. The original version is one. We set the normalization condition $r(t) + d_{sum}(t) + e(t) = 1$ and $d_{sum} = \sum_x d_x$ .The role of a netizen starts with a receiver.

When information is injected into networks such as scale-free networks, the set of coupled properties can be written as the following. Consider a receiver forwarder i with degree K after t steps. When receiving the information x, a probability with which it becomes a spreader d_x is shown in Equation (3):

$$P^{rd}_{r\to d_x}(t,k) = FkP(k)r_k(t)\sum_{k'}\frac{k'P(k')d_{k'}(t)}{\langle k\rangle}, \quad (3)$$

where $P(k)$ denotes the degree distribution of the networks and F denotes the forwarders' fraction. If i is a receiver modifier, the probability with which it becomes a modified information spreader d_{x+1} is shown in Equation (4):

$$P^{rd}_{r\to d_{x+1}}(t,k) = (1-F)kP(k)r_k(t)\sum_{k'}\frac{k'P(k')d_{k'}(t)}{\langle k\rangle}. \quad (4)$$

The probability with which a disseminator d_x becomes a terminator e is shown in Equation (5).

$$P^{de}_{d_x\to e}(t,k) = kP(k)d_k(t)\sum_{k'}\frac{k'P(k')\left[d_{k'}(t)+e_{k'}(t)\right]}{\langle k\rangle}. \quad (5)$$

We define $\langle R_k\rangle = \frac{\sum_{r\in\{r|\deg ree(r)=k\}} r_R}{N_k}$, where r_R denotes the last version of the information at netizen i before netizen i turns out to be a terminator. $\langle R_k\rangle$ represents the frequency that information has been modified on average at a netizen with degree K. The rate equation for the average revised frequency $\langle R_k\rangle$ on degree K can be written as Equation (6):

$$\frac{d\langle R_k(t)\rangle}{dt} = (1-F)P(k)kr_k(t)\sum_{k'}\frac{k'P(k')d_{k'}\langle R_{k'}(t)\rangle}{\langle k\rangle}. \quad (6)$$

The evolution of the densities $d_k(t)$ and $e_k(t)$ satisfy the following set of coupled differential formula:

$$\frac{dr_k(t)}{dt} = -kP(k)r_k(t)\sum_{k'}\frac{k'P(k')d_{k'}(t)}{\langle k\rangle}, \quad (7)$$

$$\frac{dd_k(t)}{dt} = kP(k)r_k(t)\sum_{k'}\frac{k'P(k')d_{k'}(t)}{\langle k\rangle} - kP(k)d_k(t)\sum_{k'}\frac{k'P(k')\left[d_{k'}(t)+e_{k'}(t)\right]}{\langle k\rangle}, \quad (8)$$

$$\frac{dd_k(t)}{dt} = kP(k)r_k(t)\sum_{k'}\frac{k'P(k')d_{k'}(t)}{\langle k\rangle} - kP(k)d_k(t)\sum_{k'}\frac{k'P(k')\left[d_{k'}(t)+e_{k'}(t)\right]}{\langle k\rangle}. \quad (9)$$

To clarify the result of evolution, we run extensive simulations on scale-free networks for five different values of $F = 0.1, 0.3, 0.5, 0.7, 0.9$. We generate BA network which are generated by $m_0 = m = 3$. With the initial conditions $r(0) = \frac{N-1}{N}$, $d(0) = \frac{1}{N}$ and $e(0) = 0$, $k = 6$. We measure the distributions of R in Figure 4. We define $\Phi(R)$ as the number of netizens who were the disseminators of the information revised R times before the information vanishes. One can observe that the majority of the netizens are infected by the versions revised in Figure 4. This observation indicates that the original information can keep its influence on the netizens only when most of them are forwarders in the information propagation networks.

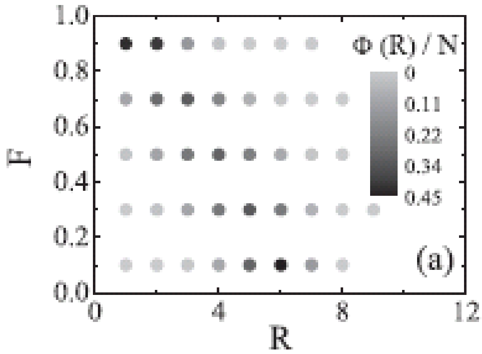

FIGURE 4 Simulation result

5 Influence propagation

There is tremendous interest in information propagation in social networks, fuelled by applications such as viral marketing, epidemiology, analysis of the spread of innovations, among many others. At the core of these applications there is a phenomenon called influence propagation, where actions performed by people propagate through a social network.

We collected data from MovieLens [6]. MovieLens is a movie rating system with five stars (i.e., ratings can be 1, 2, 3, 4 and 5). The dataset contains 6040 users, 3952 movies, and 1000292 ratings. From the dataset we construct a information propagation network, influence ϕ grows with the uploading index n. So we can obtain a simple formula $\phi = anY$, where a is a positive constant and Y is a multiplicative noise with mean one [7]. If a netizen's latest upload surpasses some "successful" threshold θ, she will continue to upload, otherwise she will stop [8]. Under these assumptions, a new upload with index n will fail to be successful with probability $P(anY < \theta) = P\left(\frac{aY}{\theta} < \frac{1}{n}\right)$. Let F be the Cumulative Distribution Function of the random variable $\frac{aY}{\theta}$. We see that a netizen who made n past uploads will stop at n with probability $F\left(\frac{1}{n}\right)=F(0)+\frac{F^{'}(0)}{n}+O\left(\frac{1}{n^2}\right)$, where $F^{'}(0)$ is a positive constant. If the number of uploads follows a distribution P_n, which means a fraction P_n of netizens stop at n uploads. We can be written as $\frac{P_n}{\sum_{i=0}^{\infty}P_{n+i}}=\frac{F^{'}(0)}{n}$.

Defining $G(n)=\sum_{i=0}^{\infty}P_{n+i}$, we get $-\frac{G^{'}(n)}{G(n)}=\frac{F^{'}(0)}{n}$. The solution to this equation is $G(n)\sim n^{-F^{'}(0)}$, which implies that $P_n=-G^{'}(n)\sim n^{-F^{'}(0)-1}$. Therefore the distribution of the information propagation network is a power-law distribution.

To show the existing of influence propagation, we look at a case: we will vote with influence propagation after voting on some very high-quality or very low-quality object. In the absence of influence propagation, the next votes should be more or less the same as usual votes; while if the influence propagation exists, a vote will become the anchor of the next vote, and thus in average we will give high rating after voting on a high-quality object and low rating after a low-quality object.

We use the average rating to estimate a movie's quality, and to reduce the possible errors caused by personalized tastes and unreasonable votes, we only consider the objects getting more than ten votes. Although ratings cannot perfectly reflect qualities, they are correlated with qualities and can be naturally treated as anchors by netizens. For MovieLens, a movie is distinguished as low-quality or high-quality object if its average rating is lower than 2.0 or higher than 4.5.

Figure 5a shows the rating series of a netizen in MovieLens. We divide ratings into two kinds: one is positive and the other is negative, and show them without explicit values in Figure 5b, where we could find that ratings in the same kind are aggregated. This kind of aggregation shows the influence propagation in voting behaviour, namely people are likely to give a high rating after a prior high rating while they are likely to give a low rating after a prior low rating. It is similar to the information propagation process that we are affected by neighbours in social network.

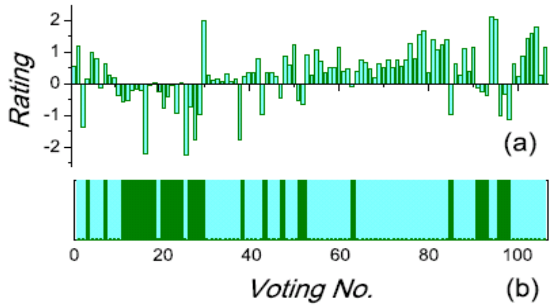

FIGURE 5 Rating series

6 Conclusions

Nowadays, Online social networking sites have become a popular way to share and disseminate content[9]. Their massive popularity has led to tremendous interest in information propagation. This paper presented information propagation network analysis. We showed the mechanisms of information propagation. As a result, we found that when the information spreads on the social networks the majority of netizens are influenced by the multirevised version. Meanwhile influence propagation is aggregated. Our result may provide a better understanding of information propagation.

Acknowledgments

This work is funded by Research Project Supported by Shanxi Scholarship Council of China (2013-097).

References

[1] Dyagilev K, Mannor S, Yom-Tom E 2013 *Social Network Analysis and Mining* **3**(3) 521-41

[2] Manuel G R, Jure L, Bernhard S 2013 Modelling information propagation with survival theory *Proceedings of the 30th International Conference on Machine Learning* Atlanta USA June 2013 JMLR:W&CP 9-16

[3] Lou S L Yang X H 2013 Journal of Information Technology and Computer Science **6**(5) 10-5

[4] Google 2013 The 1000 most-visited sites on the web http://www.google.com/adplanner/static/top1000 2013

[5] David E, Jon K 2013 Networks, Crowds and Markets Reasoning about a Highly Connected World *Cambridge University Press*

[6] Grouplens 2013 MovieLens database http://www.grouplens.org 2013

[7] Roberts J, Hann I H, Slaughter S 2006 Journal of Management Science **52**(7) 984-99

[8] Oreg S, Nov O 2008 *Journal of Computers in Human Behaviour* **24**(5) 2055-73

[9] Zhang C J, Zeng A 2012 Phys A **391** 1822-10

Authors

Jianping Qiu, born in June, 1973, Taiyuan, China

Current position, grades: teacher of Computer in Institute of Computer Science and technology, Taiyuan University of Science and Technology
University studies: Institute of Computer Science and technology in Xian University of Science and Technology
Scientific interests: computer and social network.
Publications: 10 papers
Experience: teacher at Taiyuan University of Science and Technology since 2005.

Lichao Chen, born in March, 1961, Taiyuan, China

Current position, grades: Doctor of computer, professor in Taiyuan University of Science and Technology.
University studies: Institute of Computer Science and technology, Beijing University of Science and Technology
Scientific interest: computer science.
Publications: 50 papers.

Guifang Chen, born in January, 1957, Taiyuan, China

Current position, grades: professor in Computer Science and Technology Department, Taiyuan Normal University.
University studies: computer science.
Scientific interest: computer science.
Publications: 30 papers.