

Research Article

Information Propagation in Online Social Network Based on Human Dynamics

Qiang Yan,¹ Lianren Wu,^{1,2} Chao Liu,^{2,3} and Xiye Li¹

¹ School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China

² Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287, USA

³ Department of Applied Mathematics, Shanghai University of Finance and Economics, Shanghai 200433, China

Correspondence should be addressed to Lianren Wu; lianrenw@asu.edu

Received 16 January 2013; Revised 10 April 2013; Accepted 18 April 2013

Academic Editor: Bo Shen

Copyright © 2013 Qiang Yan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We investigate the impact of human dynamics on the information propagation in online social networks. First, statistical properties of the human behavior are studied using the data from “Sina Microblog,” which is one of the most popular online social networks in China. We find that human activity patterns are heterogeneous and bursty and are often described by a power-law interevent time distribution $P(\tau) \sim \tau^{-\alpha}$. Second, we proposed an extended Susceptible-Infected (SI) propagation model to incorporate bursty and limited attention. We unveil how bursty human behavior and limited attention affect the information propagation in online social networks. The result in this paper can be useful for optimizing or controlling information propagation in online social networks.

1. Introduction

Rapid development of information and communication technology has increased the wide adoption of online social network in our life. Indeed, online social network such as Sina Microblog, Twitter, and Facebook had become an indispensable part of our life. Every day we sign into our homepages more than once to view and share information. These online social networks have common characteristics: instantaneity, simplicity, and universality. Taking Sina Microblog, for example, unlike the traditional blog, it allows the use of mobile devices to disseminate information by a length of 140 characters text at anytime and anywhere. Investigating the online social network is crucial in a broad range of settings from information propagation and viral marketing to political purposes.

Recent years, online social network as a platform for the empirical study of information has been widespread concern [1–4]. Despite the progresses that have been made, the empirical study of information propagation is still in its infancy. Studies in this direction have been mostly hindered by the shortcoming of available large-scale data. However, the availability of large-scale data from online social network has

recently created unprecedented opportunities to explore the impact of human behaviors on the information propagation.

Firstly, information propagation in online social network is determined by rhythms and activity patterns of human [5, 6]. An increasing number of recent measurements indicate that human activity patterns are heterogeneous and bursty [7–11]. If only considering the time interval between events, these human activity patterns are often described by a power-law interevent time distribution $P(\tau) \sim \tau^{-\alpha}$, where τ is the time interval between two consecutive activities [12]. Recently, the researchers began to realize that the bursty human behavior has an important impact on the dissemination of information [13, 14].

Secondly, the wide adoption of online social network has increased the competition among information for our limited attention. Every day we receive a lot of information from various online social networks. However, we do not have enough time and attention to disseminate each message which we received. It is an interesting question that whether such a competition may affect the velocity of information propagation. The issue of limited attention has been studied through messages posted and forwarded in online social

networks [15, 16]. However, how limited attention affects velocity of information propagation is still unclear.

In this paper, we propose an extended Susceptible-Infected (SI) propagation model, incorporating bursty human activity patterns and limited attention for the first time. Then, we obtain a large number of real data to test the model. Adopting the methods of theoretical research and empirical analysis, we study the information spreading process in social networking qualitatively and quantitatively. The key contributions of this study are summarized as follows.

- (1) From the empirical statistical results we find that at the group level, the interactive time (time interval between two consecutive login microblog homepage) follows power-law distribution with the slope ≈ 2.5 . And the distribution of newly infected individual (calculate as the number of new forwarding per day) follows power-law with the slope ≈ 1.5 . Two slope values satisfy the relationship $2.5 - 1.5 \approx 1.0$.
- (2) Through both the theoretical research and simulation, we prove that (a) if the generation time distribution follows power-law with exponent β , then the decay of propagation velocity will be characterized by the same power-law distribution; (b) if bursty human behavior follows a power-law distribution with exponent α , the decay of propagation velocity also follows a power-law with exponent $\beta \approx \alpha - 1$.

In summary, although tremendous efforts have been made regarding the research about information propagation, further study based on human dynamics is still needed to unveil the role of human behaviors for the information propagation in online social network. In future studies, on the other hand, we can use other more mature theories to research the spreading dynamics, such as in the references [17, 18].

The rest of this paper is organized as follows. Section 2 gives the data description. In Section 3, we propose the extended SI model. In Section 4, we present simulation results and observations. Section 5 introduces theoretical analysis. Finally, in Section 6, we conclude the work.

2. Data Description

The dataset of this paper was collected from Sina Microblog (<http://www.weibo.com/>), one of the most popular microblog platforms in China at present. The dataset includes 345,095 messages from 41667 individuals during 2009/8/16 to 2011/6/4, collected by snowball sampling. These messages have been forwarded 203,997,094 times and triggered 58,617,139 comments. For each message, message ID, releasing time, times of forwarding, and number of comments were recorded. For each individual, the individual ID and the timing of individual sign in his/her microblog homepage were recorded.

The basic statistical results show that at the group level, the interactive time (time interval between two consecutive login microblog homepage) follows power-law distribution with the slope ≈ 2.5 (Figure 1(a)). And the distribution of newly infected individual (calculate as the number of new

forwarding per day) follows power-law with the slope ≈ 1.5 (Figure 1(b)). If set the slope of interactive time distribution is α and the slope of newly infected individual distribution is β , we find that there is the relationship $\beta \approx \alpha - 1$ between two slopes.

3. Model

3.1. Model Description. In this paper, we use the branching processes [19, 20] in conjunction with power-law human behaviors to describe the process of information propagation. We adopt the Susceptible-Infected (SI) propagation model for the simulation of information propagation in online social networks. Similar to the classical SI model, the population is divided into two states, either susceptible (S) or infected (I). In the information propagation model, however, the susceptible individual is defined as the one who has not yet known a piece of message, and the infected individual is defined as the one who knows the message and shares the message with his/her friends. After being infected, an individual will never return to susceptible state. At time t , there are $S(t)$ susceptible individuals and $I(t)$ infected individuals, and the population $N = S(t) + I(t)$.

Initially all individuals are susceptible except for a single infected individual. Different with the traditional model, at a given time step, an infected individual can be inactive; that is to say, infected individual will not infect connected susceptible individuals at that time step. The time interval between two consequent active steps of an infected individual is defined as the interactive time, which is often characterized by a power-law distribution $P(\tau) \sim \tau^{-\alpha}$ at the group level. Meanwhile, different individuals have different active time interval and each individual i acts with an unchanged interactive time τ_i .

On the other hand, the advent of online social network has greatly lowered the cost of information generation and propagation, boosting the potential reach of each message. However, the abundance of information to which we are exposed through online social networks is exceeding our capacity to consume it. Due to the limited time and attention, the individual cannot continuously check the update of information on his/her homepage. We assume that individuals interact on a directed online social network. Each individual is equipped with two lists. One is the screen where received messages are recorded and maintained a time-ordered list of messages. The other is memory where individual interested messages are recorded. Each individual can share some of the messages from the list with his/her friends. The friends in turn pay attention to a newly received message by placing it at the top of their lists. Because of the limited attention, we allow messages to survive in an individual's screen for a finite amount of time T . Meanwhile, we assume that each individual only forwards each message once, and then the individual loses interest in the message. In addition, if the individual no forwarding the message within T , the individual will no longer be concerned about the message and delete it from the screen. Each message may attract the individual's attention with probability λ ; that is to say, the individual will forward the message with probability λ .

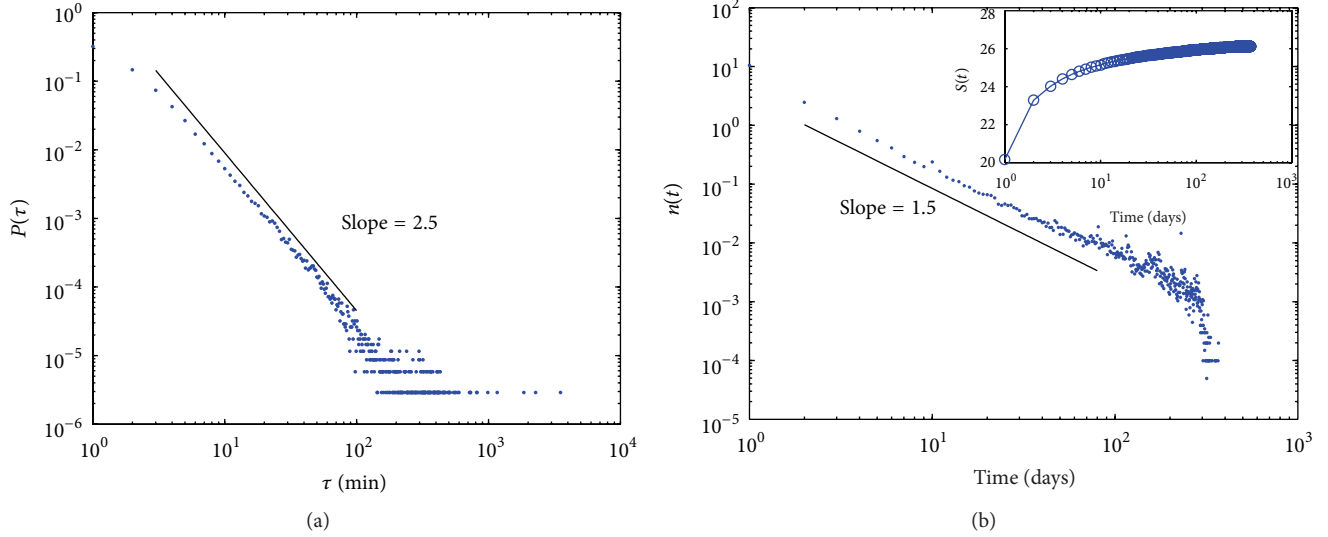


FIGURE 1: Empirical data. (a) The distribution of interactive time at the group level. (b) The distribution of newly infected individuals, inset: the cumulative distribution of newly infected individuals, namely, the distribution of all infected individuals. The results are the average of all messages.

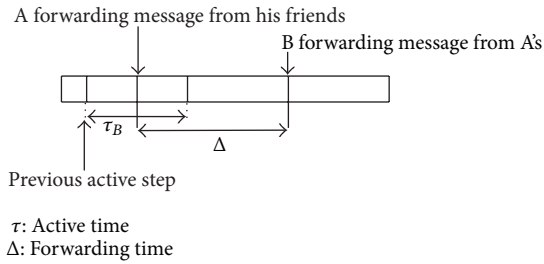


FIGURE 2: Schematic of individual interaction.

3.2. SI Model Based on Bursty and Limited Attention. According to the previous description, the SI model incorporating bursty and limited attention is illustrated in Figure 2. We characterize the timing of information propagation by the generation time Δ , which is defined as the time interval between the forwarding of an individual and the forwarding of his/her followers.

To sum up, the extended SI model is defined as follows.

Step 1. At time step $t = t_i$, an individual i posts a message. Meanwhile, individual j receives the message, where $j \in \delta_i$ and δ_i is the set of individual i 's neighbors.

Step 2. For each individual j , the first active time step is t_{j0} , $t_{j0} \in (t_i, t_i + \tau_j)$, and individual j will be active at the time steps $t = t_{j0} + k\tau_j$, $k = 1, 2, 3, \dots$, where τ_j is the active time interval of individual j .

Step 3. At each active time step, individual j will forward the message with the probability λ . If individual j forwards the message at the time step t_j , we obtain the generation time $\Delta = t_j - t_i$ and generation time must satisfy the condition $\Delta < T$.

Step 4. Update the time step $t = t_j$ and repeat Step 1 to Step 3 until the preset time steps.

In addition, we also introduce two indicators to characterize the velocity of information propagation:

- (1) the first time step when the number of infected individuals exceeds half of the population, defined as half time T^* ;
- (2) the mean infection time of an individual after the outbreak, defined as mean time $T_m = \sum_{t=0}^{t_{\max}} (tn(t)/N)$, where t_{\max} is the maximum simulation step, such as in our simulation $t_{\max} = 10^4$.

4. Simulation Results and Observations

In our simulations, initially all individuals are susceptible except for a single infected individual. Each individual i has an unchanged interactive time τ_i , which follows power-law distribution $P(\tau) \sim \tau^{-\alpha}$ with $2 < \alpha < 3$. We set $T = 1440$ time steps. This is because messages will survive in an individual's list one day, namely, 1440 minutes [15]. Simulations were performed on a BA network with size $N = 10^4$ and $\langle k \rangle = 10$. We set the degree of attention $\lambda = 0.5$ and randomly select an initial infected node. For detailed comparison, we also performed the same SI dynamics with exponential interactive time distribution $P(\tau) \sim e^{-\alpha\tau}$. From the numerical simulation results (Figures 3 and 4), we have the following observations of the propagation process.

Observation 1. In power-law case, the average number of newly infected individuals $n(t)$ and the generation time $g(\Delta)$ follow power-law distributions with the exponent $\beta \approx \alpha - 1$ (Figure 3).

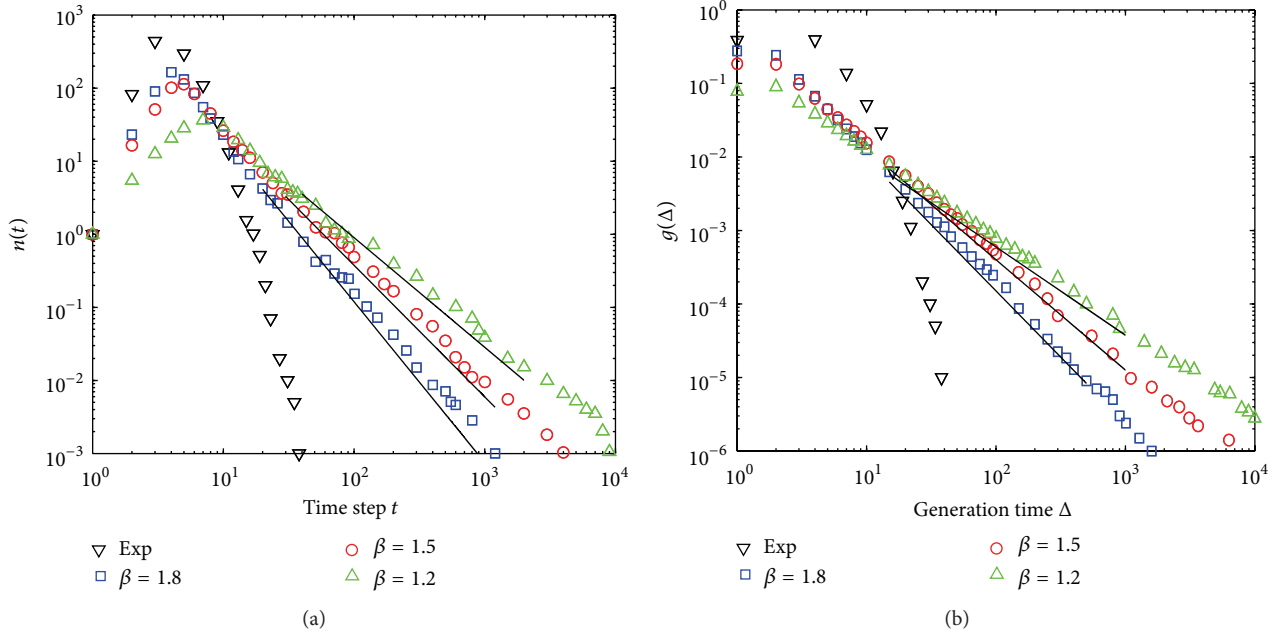


FIGURE 3: (a) The average number of newly infected individuals $n(t)$. Power-law distributions $p(\tau)$ with exponent $\alpha = 2.8$ (Squares), $\alpha = 2.5$ (circles), $\alpha = 2.2$ (up triangles), and the exponential $p(\tau)$ (down triangles). All $p(\tau)$ have the same mean interactive time $\langle \tau \rangle = 1.96$. (b) The generation time distribution $g(\Delta)$ for all $p(\tau)$. In both panels, the black lines have slopes -1.8 , -1.5 , and -1.2 . The results show that $n(t)$ and $g(\Delta)$ decay as a power law with the exponent $\beta \approx \alpha - 1$. In the exponential case, $n(t)$ decays fast, in stark contrast to the power-law case. The results are the average over 2×10^3 independent runs.

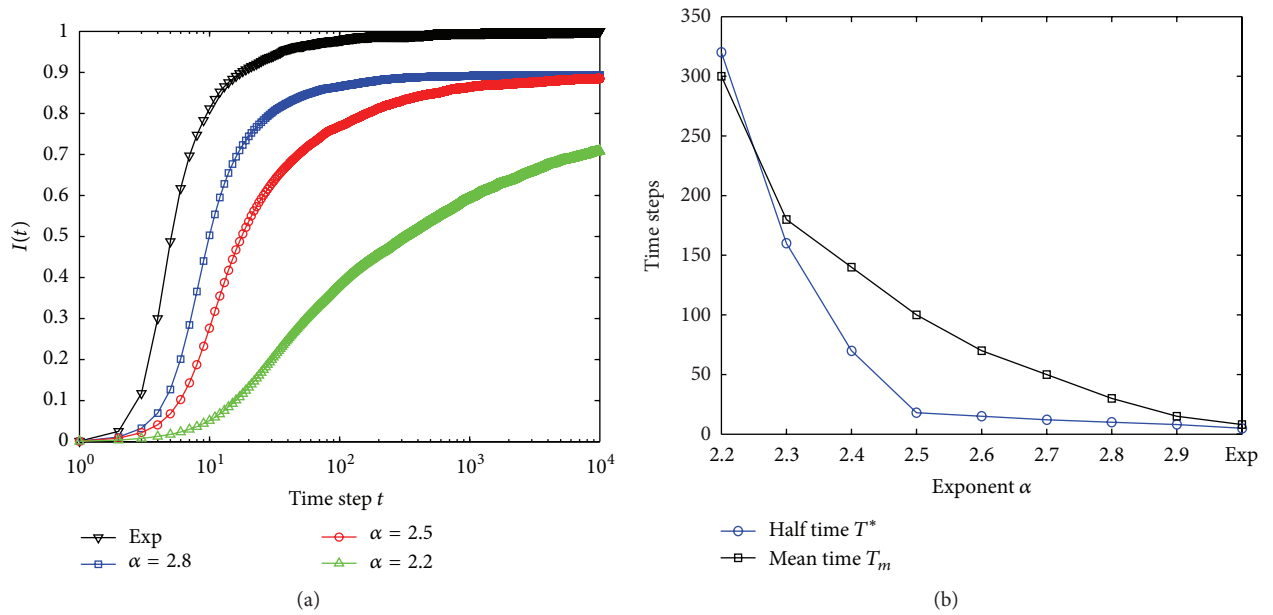


FIGURE 4: (a) The fraction of infected nodes I with different exponent α . (b) The half time T^* and the mean time T_m as the functions of exponent α .

Observation 2. The smaller the exponent α of interactive time distributions, namely, the larger heterogeneity of interactive time, resulting in the slower velocity. The half time T^* and mean time T_m monotonic decrease with the increase of exponent α (Figure 4).

In order to investigate the impact of attention on the propagation process, we fixed interactive time following power-law distribution with the exponent $\alpha = 2.5$ and randomly select an initial infected node. From other parameters $T = 1440$, simulations were also performed on a BA network with

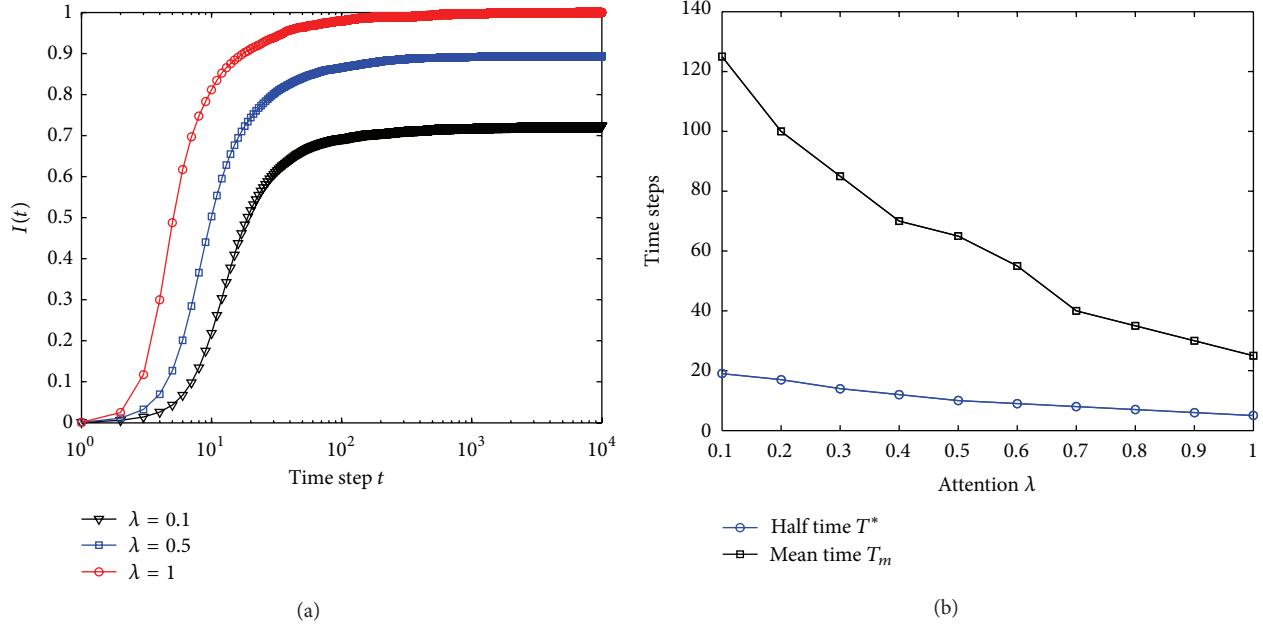


FIGURE 5: (a) The fraction of infected nodes I with different attention λ . (b) The half time T^* and mean time T_m as the functions of attention λ .

size $N = 10^4$ and $\langle k \rangle = 10$. The results are averaged over 2×10^3 independent runs. From the numerical simulation results (Figure 5), we have the following observation of the propagation process.

Observation 3. The higher the degree of attention, the faster the velocity. The half time T^* and mean time T_m monotonic decrease with the increase of attention λ (Figure 5).

5. Theoretical Analysis

In this section, the properties of propagation dynamics are analyzed. We prove that the decay exponent of propagation velocity equals that in the generation time distribution. Furthermore, we also proved that the exponent α characterizing the bursty is related to that in the decay of propagation velocity β by the relation $\beta = \alpha - 1$.

Proposition 1. *If the distribution of generation time follows power-law $g(\Delta) \sim \Delta^{-\beta}$ with $1 < \beta < 2$, the decay of propagation velocity also follows power-law $n(t) \sim t^{-\beta}$ and with the same exponent β .*

Proof. We consider a general theory of propagation process in online social networks. We assume that the propagation process outbreaks starting from a single infected individual at time $t = 0$. In this case, the average number of new infected individuals at time t is [19]

$$n(t) = \sum_{d=1}^D z_d (g^{(0)} * g^{(1)} * \dots * g^{(d)}(t)), \quad (1)$$

where z_d is the average number of individuals at generation d away from the first infected individual, where $*$ denotes the convolution operation; for example,

$$g^{(0)} * g^{(1)}(t) = \int_0^t d\tau g^{(0)}(\tau) * g^{(1)}(1 - \tau). \quad (2)$$

For the limited $1 \ll d$, we can obtain

$$g^{*d}(t) = g^{(0)} * g^{(1)} * \dots * g^{(d)}(t) \sim L_{\beta-1}\left(\frac{t}{t_d}\right) t_d, \quad (3)$$

where $t_d = \Delta_0 d^{1/(\beta-1)}$, Δ_0 is some characteristic time scale, and $L_\mu(x)$ represents the Levy distribution with exponent μ .

For $1 \ll x$, the Levy distribution $L_\mu(x)$ can be expressed as [21]

$$L_\mu(x) \sim x^{-(1+\mu)}. \quad (4)$$

To sum up, when $t \rightarrow \infty$, we obtain $g^{*d}(t) \sim t^{-\beta}$; namely, $n(t) \sim t^{-\beta}$. Thus, the proposition has been proved. \square

This preposition means that if the generation time distribution follows a power-law with the exponent β , then the decay of propagation velocity will be characterized by the same power-law distribution.

Proposition 2. *If the distribution of interactive time follows a power-law $p(\tau) \sim \tau^{-\alpha}$ with $2 < \alpha < 3$, the decay of propagation velocity also follows a power-law distribution $n(t) \sim t^{-\beta}$ with $1 < \beta < 2$ and $\beta = \alpha - 1$.*

Proof. When the distribution of interactive time follows a power-law $p(\tau) \sim \tau^{-\alpha}$ with $2 < \alpha < 3$, the active time interval τ_i has a finite mean $\langle \tau \rangle$.

Since the generation time probability density function is related to the interactive time probability density function [21], therefore we have

$$\begin{aligned} g(\Delta) &= \frac{1}{\langle \tau \rangle} \int_{\Delta}^{\infty} P(\tau) d\tau = \frac{1}{\langle \tau \rangle} \int_{\Delta}^{\infty} \tau^{-\alpha} d\tau \\ &= \frac{1}{\langle \tau \rangle} \frac{1}{-(\alpha-1)} \Delta^{-(\alpha-1)} \\ &\sim \Delta^{-(\alpha-1)}. \end{aligned} \quad (5)$$

According to Proposition 1, we obtain

$$n(t) \sim t^{-(\alpha-1)}. \quad (6)$$

Namely,

$$n(t) \sim t^{-\beta}, \quad \beta = \alpha - 1. \quad (7)$$

Thus, the proposition is proved. \square

6. Conclusion

An extended SI model is proposed in this paper. Different from the analysis of the network topology, we study the information propagation in online social networks from the perspective of human dynamics. We found that human behavior affects the range and velocity of information propagation greatly.

In the future, with the development of online social systems, there may be other factors influencing information propagation in online social network. Therefore, we must improve the propagation model in order to better explain the propagation process.

Acknowledgments

The authors would like to thank Liang Huang and Byungjoon Min for helpful discussions. This work was supported by Program for New Century Excellent Talents in University (NCET-11-0597) and the Fundamental Research Funds for the Central Universities (2012RC1002).

References

- [1] Q. Yan, L. Wu, and L. Yi, "Research on the human dynamics in mobile communities based on social identity," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 672756, 13 pages, 2012.
- [2] Y. P. Xiao, B. Wang, Y. B. Liu et al., "Analyzing, modeling, and simulation for human dynamics in social network," *Abstract and Applied Analysis*, vol. 2012, Article ID 208791, 16 pages, 2012.
- [3] Y. P. Xiao, B. Wang, B. Wu, Z. Yan, S. Jia, and Y. Liu, "A hybrid human dynamics model on analyzing hotspots in social networks," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 678286, 13 pages, 2012.
- [4] Q. Yan and L. Wu, "Impact of bursty human activity patterns on the popularity of online content," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 872908, 11 pages, 2012.
- [5] A. L. Barabási, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, no. 7039, pp. 207–211, 2005.
- [6] H. H. Jo, M. Karsai, J. Kertesz, and K. Kaski, "Circadian pattern and burstiness in mobile phone communication," *New Journal of Physics*, vol. 14, Article ID 013055, 2012.
- [7] A. Vazquez, J. G. Oliveira, Z. Dezso, K. I. Goh, I. Kondor, and A. L. Barabasi, "Modeling bursts and heavy tails in human dynamics," *Physical Review E*, vol. 73, Article ID 036127, 2006.
- [8] J. G. Oliveira and A. L. Barabási, "Human dynamics: Darwin and Einstein correspondence patterns," *Nature*, vol. 437, no. 7063, p. 1251, 2005.
- [9] Z. D. Zhao, H. Xia, M. S. Shang, and T. Zhou, "Empirical analysis on the human dynamics of a large-scale short message communication system," *Chinese Physics Letters*, vol. 28, no. 6, Article ID 068901, 2011.
- [10] W. Hong, X. P. Han, T. Zhou, and B. H. Wang, "Heavy-tailed statistics in short-message communication," *Chinese Physics Letters*, vol. 26, Article ID 028902, 2009.
- [11] Y. Wu, C. Zhou, J. Xiao, J. Kurths, and H. J. Schellnhuber, "Evidence for a bimodal distribution in human communication," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 44, pp. 18803–18808, 2010.
- [12] Y. Wu, Q. Ye, L. Li, and J. Xiao, "Power-law properties of human view and reply behavior in online society," *Mathematical Problems in Engineering*, vol. 2012, Article ID 969087, 7 pages, 2012.
- [13] B. Min, K. I. Goh, and A. Vazquez, "Spreading dynamics following bursty human activity patterns," *Physical Review E*, vol. 83, no. 3, Article ID 036102, 2011.
- [14] A. Vazquez, B. Racz, A. Lukacs, and A. L. Barabasi, "Impact of non-Poissonian activity patterns on spreading processes," *Physical Review Letters*, vol. 98, Article ID 158702, 2007.
- [15] L. Weng, A. Flammini, A. Vespignani, and F. Menczer, "Competition among memes in a world with limited attention," *Scientific Reports*, vol. 2, article 335, 2012.
- [16] Q. Yan, L. L. Yi, and L. R. Wu, "Human dynamic model co-driven by interest and social identity in the MicroBlog community," *Physica A*, vol. 391, pp. 1540–1545, 2012.
- [17] H. L. Dong, Z. D. Wang, and H. J. Gao, "Distributed filtering for a class of time-varying systems over sensor networks with quantization errors and successive packet dropouts," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 3164–3173, 2012.
- [18] H. L. Dong, Z. D. Wang, D. W. C. Ho, and H. J. Gao, "Distributed H-infinity state estimation with stochastic parameters and nonlinearities through sensor networks: the finite-horizon case," *Automatica*, vol. 48, no. 8, pp. 1575–1585, 2012.
- [19] A. Vazquez, "Polynomial growth in branching processes with diverging reproductive number," *Physical Review Letters*, vol. 96, no. 3, Article ID 038702, 2006.
- [20] T. E. Harris, *The Theory of Branching Processes*, Springer, Berlin, Germany, 2002.
- [21] W. Feller, *An Introduction to Probability Theory and Its Applications*, Wiley, New York, NY, USA, 1967.