Research Article

Power-Law Properties of Human View and Reply Behavior in Online Society

Ye Wu,[1, 2] Qihui Ye,[2] Lixiang Li,[3] and Jinghua Xiao[1, 2]

[1] *State Key Lab of Information Photonics and Optical Communications,*
  Beijing University of Posts and Telecommunications, Beijing 100876, China
[2] *School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[3] *Information Security Center, Beijing University of Posts and Telecommunications, Beijing 100876, China*

Correspondence should be addressed to Ye Wu, wuye@bupt.edu.cn

Statistical properties of the human comment behavior are studied using data from "Tianya" and "Tieba" which are very popular online social systems (or forums) in China. We find that both the reply number R and the view number V of a thread in a subforum obey the power-law distributions $P(R) = R^\alpha$ and $P(V) \propto V^\beta$, respectively, which indicates that there exists a kind of highly popular topics. These topics should be specially paid much attention, because they play an important role in the public opinion formation and the public opinion control. In addition, the relationship between R and V also obeys the power-law function $R \propto V^\gamma$. Based on the human comment habit, a model is introduced to explain the human view and reply behaviors in the forum. Numerical simulations of the model fit well with the empirical results. Our findings are helpful for discovering collective patterns of human behaviors and the evolution of public opinions on the virtual society as well as the real one.

1. Introduction

Statistical properties and models of human behaviors have received much attention in different scientific fields, such as sociology, psychology, and economics. However, most of the existing findings are only qualitative analyses for the lack of real data about the complexity of human behaviors. Usually, it is assumed that the human behavior is a Poisson progress [1, 2] which is a kind of the Markov progress. However, some researchers found that the interevent time distribution of some human behaviors is power-law which means that it is a non-Markov [3–14] one. More and more researchers are interested in it for its importance in the theory and potential applications.

   As an important part of modern life and human dynamics, the human behavior on the Internet also attracts more and more attention. Chmiel et al. investigated the flows of visitors

migrating between different portal subpages. A model of portal surfing was developed where a browsing process corresponds to a self-attracting walk on weighted networks with a short memory [15]. Grabowski found that the distribution of human activity has the form of a power-law [16] distribution. Based on the data from "Tianya", Wu et al. found that the dynamics of human comment in the online society is non-Markov. Further, they proposed a model to explain it [17]. All these researches indicated that some kinds of human behavior in on-line systems were non-Markov. They have some common statistic properties. More and more researchers considered the forum as a virtual society to study the property and the evolution of complex friendship networks [18, 19].

A forum is very important for the information and the spreading of public opinions. Many public opinions were also formatted and then spread in the forum. Analyzing the user behavior in the forum is not only helpful for understanding the human behavior and enhancing the information spreading, but also for designing a better website which is important for the information spreading. Recently in China, the news about controlling public opinions on purpose by news have attracted more and more attention. There was a report that at least a half of public opinions in the Internet were proposed by some companies on purpose. So it is very important to study the human comment behavior in the forum. Yu et al. analyzed the view and reply data in the forum which was the beginning of researches on the human comment behavior in the forums. They found that the view and reply numbers of a thread in the sub-forum were power-law. However, they mainly considered statistic properties of the behavior and did not present a model to explain the basic mechanism [20].

In this paper, we consider the data collected from "Tianya" and "Tieba" which are very popular on-line social system in China and different from those in [17]. We show that both the view number (V) and the reply number (R) of a thread in the sub-forum obey power-law distributions which confirmed Yu et al.'s finding [20]. The relationship between V and R is also power-law. These present that a lot of topics are important in the formation and evolution of public opinions. Furthermore, based on the human habit, a model is proposed to explain these phenomena. Numerical simulations are given to explain the human comment behavior in the forum. We hope it is useful for understanding complex human behaviors in the forums.

This paper is organized as follows: in Section 2, the origin of the data is introduced. The statistical results are presented in Section 3. The model and numerical simulations are presented in Section 4. Finally, our conclusion is given in Section 5.

2. Description of the Original Data

Our data are obtained from "Tianya" (http://www.tianya.cn) and "Tieba" (http://tieba.bai-du.com), which are two most popular on-line social systems in China. Our data are collected from the sub-forums of "Tianya" and "Tieba." Each user is assigned a different identity name (ID) in the forums. A topic in the sub-forum is called a thread. A thread is a minimal unit, and it can be divided into a root thread and the reply threads. A root thread is a new topic, and the reply threads are related to a root one. The users discuss the public opinion in both the root and reply threads. Until 2010/02/11, there were 33,296,350 IDs in "Tianya," and about 200,000 IDs on average were on-line at the same time. The topics and the public opinion in "Tianya" and "Tieba" reflect part of the public opinions of the real society in China. Our data sets are collected from the threads in four sub-forums. The types of these topics are different from public news to personal stories which indicate that our results are general for different

Table 1: Detailed format of a subforum.

Topic	Last update ID	Reply/View	Last update time
A	ID 1	123/23124	2009/02/12, 12:11:05
B	ID 3	3243/323532	2009/02/12, 12:10:05
C	ID 32	323/42421	2009/02/12, 12:03:05
…	…		
Y	ID 31	43/232	2009/02/12, 12:01:05

Table 2: Detailed informations about four randomly selected sub-forums.

Subforum	A	B	C	D
Total threads	19,492	19,479	27,359	5,302
Total clicks	1.05×10^9	1.62×10^9	2.33×10^8	5.19×10^6
Total replies	1.11×10^8	1.43×10^8	7.89×10^5	1.77×10^5
Duration (day)	247	2569	3	15

contents. The format of the data is shown in Table 1, where the first column is the title of a thread, the second one gives the author's name of a root thread, the third one shows R and V, and the last one is the last update time of a thread.

3. Statistical Results

In the forum, the view and reply times of a thread reflect the influencing ability of a topic. Further, more reply times mean more discussions and more communications. These two parameters play an important role in the public opinion formation and the web design. Hence, we study statistical properties of V and R in the thread of each sub-forum. Four sub-forums are randomly selected as our data sets. The topics and some prosperities are listed in Table 2.

The distributions of V and R in each sub-forum are shown in Figure 1 from which we can clearly see that all the distributions are power-law, although the threads differ in their contents. Their exponents vary with different sub-forums. These results show that the process of human comments is non-Markov which is the same as the human dynamics of the letter and e-mail communications, the web browsing, online movie watching, and broker trades. The heavy tail of the distribution allows for much more numbers of threads which have larger amounts of V and R than the Poisson progress. The thread which has more V and R has much more influences on the public opinion. The number of these kind of threads is so large that they cannot be ignored. A large population will read the thread by which their opinions may be influenced. So we must pay much attention to them.

As is known to all, the more the view, the more the reply. However, the quantity relationship between V and R is not very easy to know and it is the basic property of a thread. Hence, next we mainly focus on the relationship between the human's view and reply behaviors in Figure 2. We found that it can be illustrated as a straight line in a log-log plot, which means $R \propto V^\gamma$. It is easy to understand that the more the view, the more the reply. Moreover, the nonlinear relationship here also means that the reply number increases slower than the view one when the view number is large enough. It also indicates that human's interest in reply decreases as the increment of V.
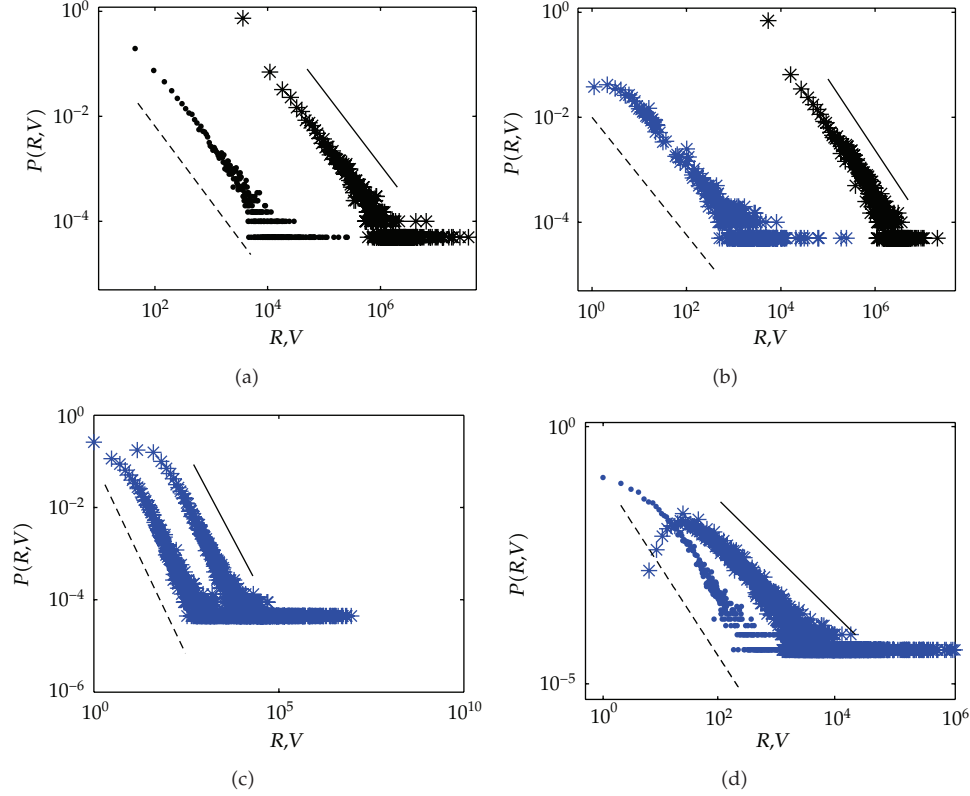
(a)

(b)

(c)

(d)

Figure 1: Distributions of V ($*$) and R ($\bullet$) in each sub-forum. The solid line and the dashed one show the slopes of fitting function for distributions of V and R, respectively, where (a) sub-forum A, the slopes $\alpha = 1.40 \pm 0.01$, $\beta = 1.44 \pm 0.02$, (b) sub-forum B, the slopes $\alpha = 1.35 \pm 0.02$, $\beta = 1.12 \pm 0.01$, (c) sub-forum C, the slopes $\alpha = 1.51 \pm 0.01$, $\beta = 1.68 \pm 0.01$, (d) sub-forum D, the slopes $\alpha = 1.12 \pm 0.02$, $\beta = 1.76 \pm 0.02$.

4. The Model and the Simulations

In order to get a better understanding of our empirical observations in Section 3, we propose a model based on our intuitive experience about the human comment habit. We see that the view number of each sub-forum increases more quickly as the time evolves. There are many threads on each sub-forum. Each thread will be viewed based on its content and its previous view time. Hence, our model is defined by the following scheme.

Step 1 (growing). At time $t = 0$, there are a few threads on the sub-forum, and each thread has a random small V and R. At each step, a new thread is created, and there are $c * t^{\theta}$ views on the old thread. All the old threads have the probabilities to be viewed.

Step 2 (view habit). The probability that an old thread is viewed at each step is based on its attraction $\Pi(i) = A_i(t)/\Sigma A_i(t)$, where $A_i(t)$ is the attraction of a thread i at time t and it is reflected by the previous view number $V_i(t)$, that is, $A_i(t) = A(0) + V_i(t)$. Here $A(0)$ represents the initial attraction which is different due to different topics.

Step 3 (reply habit). At each step, when the user views a thread, he has a probability $P(i) = L * (R(i)/V(i))^{\eta}$ to reply the thread.
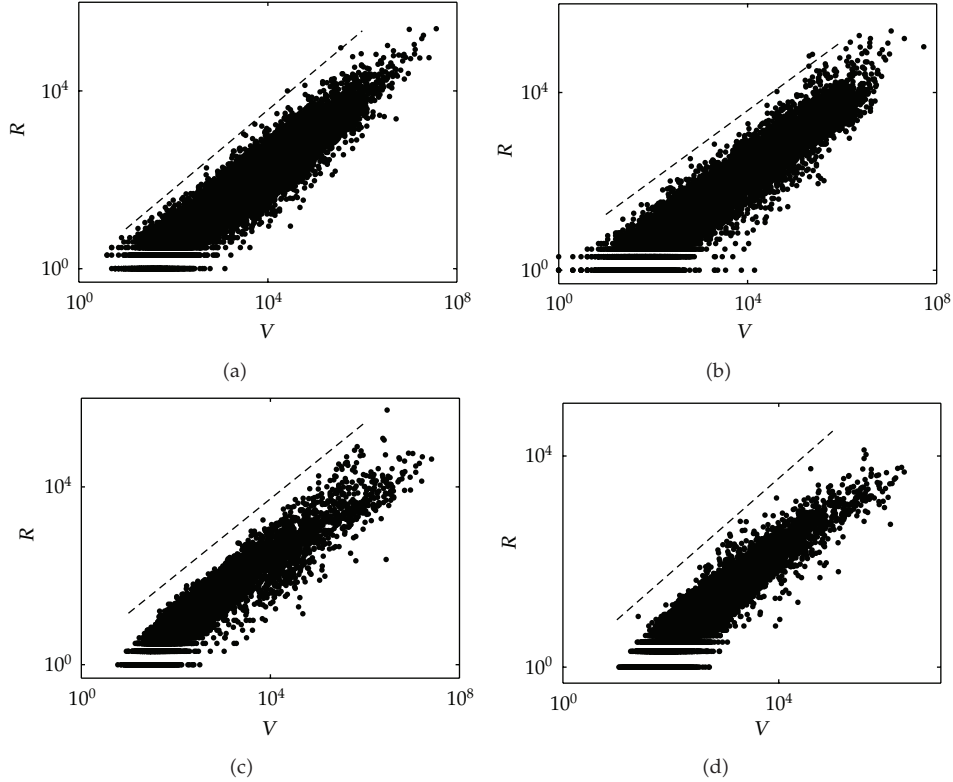
Figure 2: The power-law relationship between V and R, where (a) sub-forum A, the slope $\gamma = 0.77$, (b) sub-forum B, the slope $\gamma = 0.89$, (c) sub-forum C, the slope $\gamma = 0.85$, (d) sub-forum D, the slope $\gamma = 0.90$.

Mathematically, the model is similar to the growing networks in [21]. Based on the analysis of the growing network in this paper, we obtain that the distribution of V_i is a power-law one, that is, $P(V_i) \propto (V_i)^{-\alpha}$ at a large enough time t where the exponent α is $1 + 1/(1 + \theta)$.

To compare our model with empirical observation results, let us take the sub-forum C in our data sets as an example. Here we use the parameters $\theta = 0.9, L = 0.1, \eta = 0.5$ in the simulation. The results are shown in Figure 3. Figure 3(a) presents that the distribution of the view is indeed a power-law one with a similar exponent as that from the data. Figure 3(b) shows that the reply number also obeys a power-law distribution. The nonlinear relationship between the view and reply times is shown in Figure 3(c) which is the same as that from the data. In Figure 3(d), we further study the relationship between the parameter η and the slope γ. We see that γ decreases as η increases. From the analyses above, we can see that the proposed model can well-describe most important features in the human view and reply behaviors in online social systems.

5. Conclusion

In this paper, we analyze the statistical properties of the view and reply behaviors in on-line social systems. We find that they are different types of interactive human dynamics which are non-Markov. The view and the reply behaviors follow power-law distributions, and the relationship between them also follows a power-law one. A model based on the personal
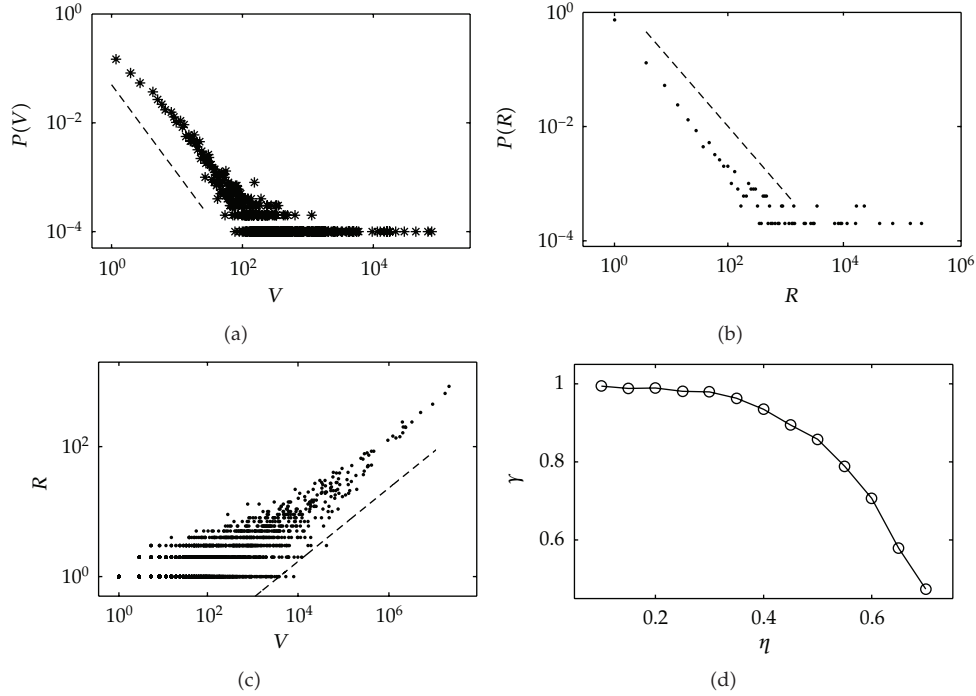
(a)

(b)

(c)

(d)

Figure 3: Simulation results of the model whose parameters are selected as $\theta = 0.9, L = 0.1, \eta = 0.5$, where the dashed line shows the slope of the fitting function. (a) The distribution of V of a thread. The slope of fitting function is $\alpha = 1.51 \pm 0.02$. (b) The distribution of R of a thread. The slope of fitting function is $\alpha = 1.41 \pm 0.04$. (c) The relationship between V and R. The slope of fitting function is $\gamma = 0.9 \pm 0.03$. (d) The relationship between the slope γ and the parameter η.

attraction is introduced to explain the human complex behavior. Numerical simulations of the model fit well with empirical results. Our work is useful to understand the human complex behavior in realistic society, for example, the human discussion behavior in a meeting or group communications in trunked mobile telephony [22]. We expect that quantitative understanding of human view and reply behaviors, when combined with additional content analyses, will open a new perspective on distinguishing fraud public opinions from realistic opinions.

Acknowledgment

References

[1] F. A. Haight, *Handbook of the Poisson Distribution*, John Wiley & Sons, New York, NY, USA, 1967.
[2] P. Reynolds, *Call Center Staffing*, The call Center School Press, Lebanon, Tenn, USA, 2003.

[3] A. L. Barabási, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, no. 7039, pp. 207–211, 2005.

[4] 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.

[5] A. Vázquez, J. G. Oliveira, Z. Dezsö, K. I. Goh, I. Kondor, and A. L. Barabási, "Modeling bursts and heavy tails in human dynamics," *Physical Review E*, vol. 73, no. 3, Article ID 036127, pp. 1–19, 2006.

[6] U. Harder and M. Paczuski, "Correlated dynamics in human printing behavior," *Physica A*, vol. 361, no. 1, pp. 329–336, 2006.

[7] T. Zhou, H. A. T. Kiet, B. J. Kim, B. H. Wang, and P. Holme, "Role of activity in human dynamics," *EuroPhysics Letters*, vol. 82, no. 2, Article ID 28002, 2008.

[8] M. Li and W. Zhao, "Visiting power laws in cyber-physical networking systems," *Mathematical Problems in Engineering*, vol. 2012, Article ID 302786, 13 pages, 2012.

[9] M. Li, C. Cattani, and S. Y. Chen, "Viewing sea level by a one-dimensional random function with long memory," *Mathematical Problems in Engineering*, vol. 2011, Article ID 654284, 13 pages, 2011.

[10] M. Li, "Fractal time series-a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.

[11] Z. Jiao, Y. Q. Chen, and I. Podlubny, *Distributed-Order Dynamic Systems*, Springer, New York, NY, USA, 2011.

[12] H. Sheng, H. Sun, Y. Chen, and T. Qiu, "Synthesis of multifractional Gaussian noises based on variable-order fractional operators," *Signal Processing*, vol. 91, no. 7, pp. 1636–1644, 2011.

[13] H. Sheng, Y. Q. Chen, and T. Qiu, "Heavy-tailed distribution and local long memory in time series of molecular motion on the cell membrane," *Fluctuation and Noise Letters*, vol. 10, no. 1, pp. 93–119, 2011.

[14] H. Sheng, Y. Q. Chen, and T. S. Qiu, *Fractional Processes and Fractional Order Signal Processing*, Springer, New York, NY, USA, 2012.

[15] A. Chmiel, K. Kowalska, and J. A. Holyst, "Scaling of human behavior during portal browsing," *Physical Review E*, vol. 80, no. 6, Article ID 066122, 7 pages, 2009.

[16] A. Grabowski, "Human behavior in online social systems," *European Physical Journal B*, vol. 69, no. 4, pp. 605–611, 2009.

[17] Y. Wu, C. Zhou, M. Chen, J. Xiao, and J. Kurths, "Human comment dynamics in on-line social systems," *Physica A*, vol. 389, no. 24, pp. 5832–5837, 2010.

[18] A. Grabowski, N. Kruszewska, and R. A. Kosiński, "Properties of on-line social systems," *European Physical Journal B*, vol. 66, no. 1, pp. 107–113, 2008.

[19] G. Csanyi and B. Szendroi, "Structure of a large social network," *Physical Review E*, vol. 69, no. 3, Article ID 036131, 5 pages, 2004.

[20] J. Yu, Y. Hu, M. Yu, and Z. Di, "Analyzing netizens' view and reply behaviors on the forum," *Physica A*, vol. 389, no. 16, pp. 3267–3273, 2010.

[21] S. N. Dorogovtsev and J. F. F. Mendes, "Effect of the accelerating growth of communications networks on their structure," *Physical Review E*, vol. 63, no. 2, Article ID 025101, pp. 1–4, 2001.

[22] D. G. Xenikos, "Modeling human dialogue-the case of group communications in trunked mobile telephony," *Physica A*, vol. 388, no. 23, pp. 4910–4918, 2009.