Physica A 392 (2013) 1712–1723

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Social network based microblog user behavior analysis

Qiang Yan *, Lianren Wu, Lan Zheng

Beijing University of Posts and Telecommunications, Beijing, China

A R T I C L E I N F O

Article history:
Received 3 February 2012
Received in revised form 25 September 2012
Available online 20 December 2012

Keywords:
Microblog
Social network
User behavior
Power-law distribution

A B S T R A C T

The influence of microblog on information transmission is becoming more and more obvious. By characterizing the behavior of following and being followed as out-degree and in-degree respectively, a microblog social network was built in this paper. It was found to have short diameter of connected graph, short average path length and high average clustering coefficient. The distributions of out-degree, in-degree and total number of microblogs posted present power-law characters. The exponent of total number distribution of microblogs is negatively correlated with the degree of each user. With the increase of degree, the exponent decreases much slower. Based on empirical analysis, we proposed a social network based human dynamics model in this paper, and pointed out that inducing drive and spontaneous drive lead to the behavior of posting microblogs. The simulation results of our model match well with practical situation.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Ever since the advent of Twitter in 2006, microblog has rapidly become an influential way to share information. Microblog started in China in 2007 and has made China the most active Microblog market all over the world. The number of microblog users in China, which amounts to 0.195 billion by June 2011, jumped by 208.9% in six months. Also, an open architecture of social network has been seen in the Chinese microblog world. Considering the characteristics of the microblog social network, we wonder what influence the microblog social network could have on the behavior of microblog users?

A lot of theoretical and empirical studies on human dynamics [1,2] have been published since Barabasi [3] proved that the distributions of interval time between sending and receiving Email or regular mail follow power-law distribution. In the mobile Internet environment, the interval time distributions of mobile calls [4], sending short message [5], posting and commenting on microblogs [6] follow a power-law distribution. The interval time distribution of sending instant messages and the total number distribution of instant messages also present power-law characters [7,8]. In order to explain the phenomena of power-law distribution in human dynamics, a lot of factors have been studied, including interest [9–12], activity level [13] and social identity [6,14], for example. In the area of personal mobile communications, scientific study revealed that social interactions between mobile phone users have an influence on their choice about mobile service [15], i.e. their consumption habits can be influenced by neighbors in their social network [16], and relational attributes are powerful predictors of customers' voice calling [17]. Studies conducted by Daniel Birke and G. Swann have also confirmed that the structure of social network has an impact on the consumer's choice about mobile communication service [18].

In the microblog world, is the relationship between microblog users closely related with their behavior? Considering the social network attribute of microblog users, the following questions will be answered in this paper. What are the characteristics of the relationship between microblog users? What is the relationship between the social network attributes of microblog users and their behavior? Could social network attributes be used to explain microblog users' behavior?

* Corresponding author. Tel.: +86 10 62281232; fax: +86 10 62281232.
E-mail address: buptyan@gmail.com (Q. Yan).

0378-4371/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2012.12.008

Table 1
The data format of relationship between microblog users.

User id1	User id2	Relation type(following/follower)
ID1	ID2	Following(User ID2 follows User ID1)
...	...	...
ID338	ID890	Follower(User ID338 follows User ID890)
...	...	...

Table 2
The data format of the basic information of microblog users.

User id	Following number	Follower number	Microblog number
ID1	50	10	30
...	...	...	...
ID338	890	340	1540
...	...	...	...

The remainder of the paper is organized as follows. Section 2 introduces the method and data used in this paper. The statistical results of the data are presented in Section 3. Section 4 introduces a social interaction based model and discusses the simulation results. Finally, the conclusions are outlined in Section 5.

2. Method and data description

2.1. Method

(1) Social network analysis

A social network is a set of social actors and their relationships. Social network analysis, which emphasizes the importance of the relationship between social actors, presents a special view in social science. Five social network attributes, degree, degree distribution, diameter, network clustering coefficient and centralization of a graph, will be discussed in this paper.

The relationship between microblog users can be divided into two kinds, the following and follower. If user A follows user B, then A becomes B's follower, and B is at the same time as A's following. Also, B can follow A, under this circumstance, A is B's following and B is A's follower.

The characteristics of behavior of microblog users with a different number of following and follower will be analyzed below.

(2) User behavior analysis

The analysis of microblog user behavior will be discussed in two directions. First, the analysis of interval time of people using microblog will be calculated to observe user behavior through the dimension of time. Second, the total number of microblogs posted by each user will be discussed to analyze user behavior from the dimension of quantity.

The time between two consecutive microblogs from the same user was denoted by interval time, and it was calculated by minute in this paper.

2.2. Data description

The data in this paper was collected from Sina microblog, the largest and the most active microblog website in China. According to the report published by Hitwise, Sina microblog accounts for about 0.63% of site visits in China in April 2011, i.e. one visit out of 158 site visits is to Sina microblog. Hence, the analysis of Sina microblog users can, to some extent, explain the microblog user behavior. Giving the fact that the number of Sina microblog users has reached hundreds of millions, non-probabilistic sampling is more practical than probabilistic sampling. Therefore, convenient sampling method and snowball sampling have been used in our case.

Two kinds of data set have been used in our paper. Data set 1 was used to analyze the social network attributes of microblog users. One microblog group was selected from Sina microblog in our study, and the relationships between its 1082 members have also been obtained. Specific information is shown below in Table 1.

Data set 2 was used to analyze microblog user behavior. In this data set, the basic information (as shown in Table 2) of 41,667 users was obtained by snowball sampling, and 345,096 microblogs (Table 3) were gathered by randomly selecting 291 users' microblogs, from the day when they started using Sina microblog.

3. Statistical results

3.1. Analysis of microblog social network

Social network analysis of our selected microblog group, 1082 users, was shown below. Fig. 1, social network graph of the studied microblog group, showed the relationship between the 1082 users. There are 2582 directed edges and 336 outliers

Table 3
The data format of microblogs.

User id	Microblog id	Date	Forward	Comment
ID1	Mid1	2011/1/15 13:21	1	4
...	...	...	...	...
ID2	Mid9	2011/2/15 11:21	0	0
...	...	...	...	...

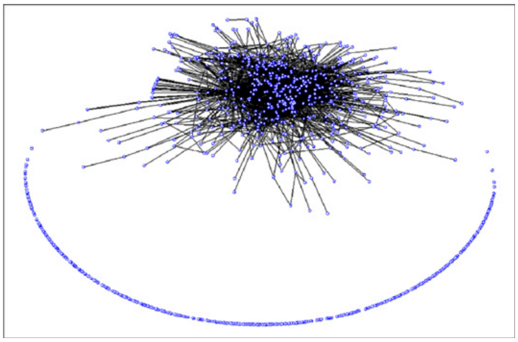

Fig. 1. Social network graph of sample microblog group.

Table 4
The statistical properties of microblog social network.

Social network type	Node	Average degree	Diameter	Average path length	Average clustering coefficient	Centralization of a graph
Sample microblog social network	1082	4.77	7	2.76	0.2062	0.3468
Reference random network	1082	4.77	11	4.61	0.0098	0.0198

Note: The statistical properties in the table were all calculated as an undirected network.

in the graph. The existence of outliers means that some microblog users in our sample microblog group do not follow others and yet are not followed by other microblog users.

The statistical properties of our sample microblog group have been calculated in Table 4. The comparison between microblog social network and a random network with the same number of nodes and average degree, which was created by Pajek, has also been shown in Table 4.

According to Table 4, compared to the reference random network, the microblog group has a shorter diameter of connected graph, average path length and larger average clustering coefficient, which means that the microblog social network is a closely connected group and has small-world characteristic. Also, the microblog social network has a higher centralization of a graph, which confirms that there is a centripetal tendency in the microblog social network.

The in-degree, out-degree and relative-degree distribution in double logarithmic coordinate has been shown in Fig. 2 and power-law distribution has also been observed. In the microblog world, most users would follow those they are interested in or be followed by people who are interested in them. As a result, only a small percentage of users will follow a large number of people or have many followers. Those people can thus have larger in-degree or out-degree. In particular, in-degree distribution, similar form of which can be observed in the study of blog networks [19], is different from out-degree distribution. The two scaling regimes observed in out-degree distribution coincide with the degree distributions of some online social networks [12,20–22]. That might be the consequence of the co-existence of both real social relations and virtual web relations. Besides, there is significant correlation between a node's in-degree and out-degree, with the relative coefficient being 0.597. That is because the relations between microblog users are reciprocal. If a user follows a lot of people, he or she is more likely to be followed by other users.

(Here we use Maximum Likelihood Estimation (MLE) method to estimate the slope γ)

3.2. Microblog user behavior analysis

3.2.1. The overall situation

The study of behavior of users posting microblogs based on data set 2 is discussed below. As can be seen in Fig. 3(a), the interval time distribution of people posting microblogs does not fit into normal distribution, however, it is a power-law distribution. For microblog users, posting microblogs is outburst behavior. People would frequently post microblogs after a long break. Shown in Fig. 3(b), only a small proportion of users actively post original microblogs, and most users tend to browse microblogs instead of posting them.

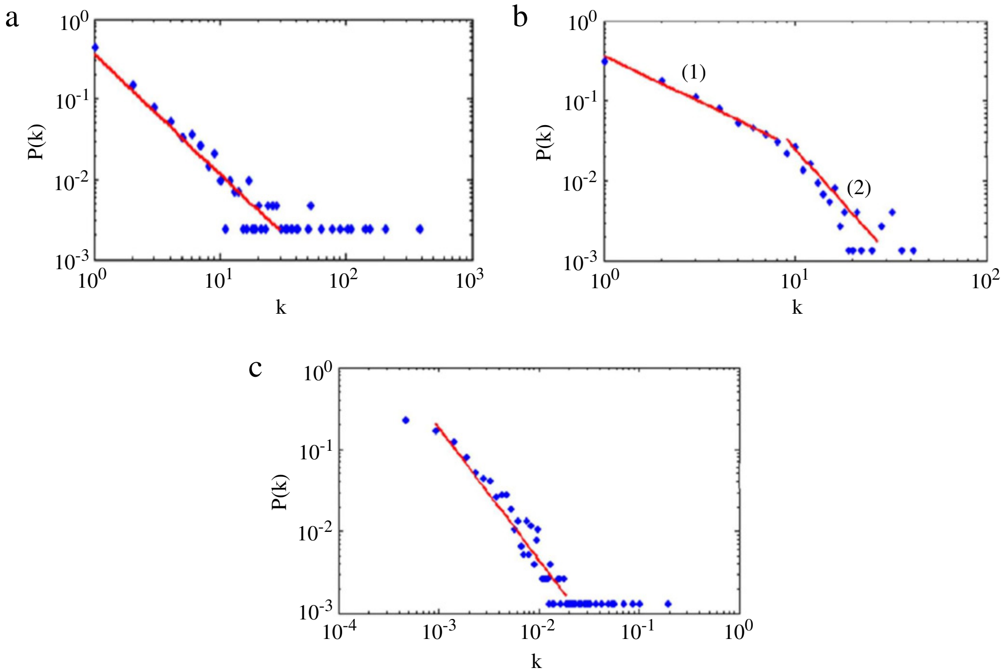

Fig. 2. The degree distribution of microblog social network (a) In-degree distribution, slope $\gamma = 1.49$, (b) Out-degree distribution, slope $\gamma 1 = 1.15$, $\gamma 2 = 2.70$, (c) Relative-degree distribution, slope $\gamma = 1.61$.

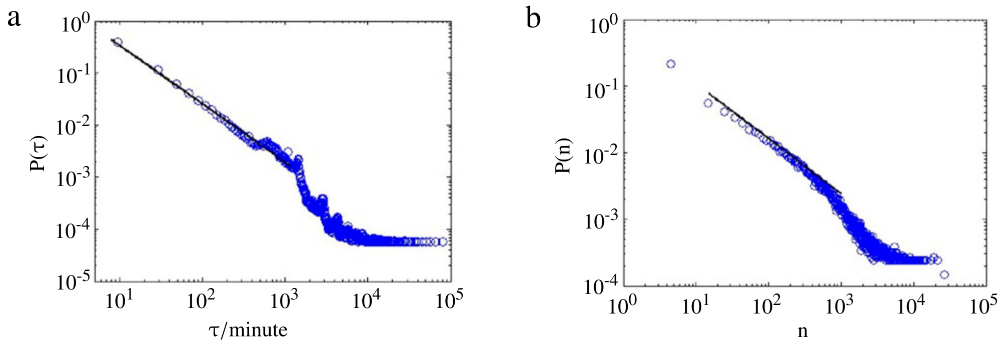

Fig. 3. The overall situation of microblog user behavior (a) The interval time distribution of users posting microblogs in double logarithmic coordinate, slope $\gamma = 1.12$, (b). The distribution of the total number of microblogs posted by per user in double logarithmic coordinate, slope $\gamma = 0.82$. Each sub-figure has a cut-off near the 10^3, this phenomenon is related to the rhythm of human life.

3.2.2. Quantitative analysis based on relationship of following between microblog users

We will discuss the impact of the total number of follower and following on users' microblog using behavior below. First, let us focus on the mutual influence between the total number of following and microblogs.

The level of a user's interest in other users can be reflected by out-degree. In this study, we divided all the users into 6 groups according to out-degree from low to high, with approximately the same number of users and out-degree in each group. The information of grouping and total number distribution of microblogs in each group can be seen in Table 5 and Fig. 4. Fig. 4 shows that the total number distributions of microblogs of each group follow power-law approximately. The exponent of the distribution is decreasing with the increase of out-degree (or the level of active social interaction). Users post microblogs more frequently and post more microblogs with the increase of their active social interactions.

It also can be seen from Fig. 5 that the exponents decrease slower than expected at higher level of active social interaction. The relationship between users' following behavior and the total number of microblogs becomes less obvious when the number of users' following goes over a certain limit. Dunbar number shows that the total number of friends (communicating with each other at least once a year) a human can handle is about 150, no matter how many friends they have on their mobile phones or chat software. Fig. 5 also shows that the exponents decrease much slower when the number of average following exceeds 104. From the data analysis in our study, for an ordinary person, the relationship between users' following behavior and the total number of microblogs is obvious when the total number of following is within about 100.

Table 5
Information of dividing users according to active social interaction.

Group name	Total number of users	Average number of following	Slope γ
Group 1	6234	22	1.25
Group 2	6291	58	1.00
Group 3	6320	104	0.85
Group 4	6364	183	0.81
Group 5	6390	390	0.78
Group 6	6382	1406	0.72

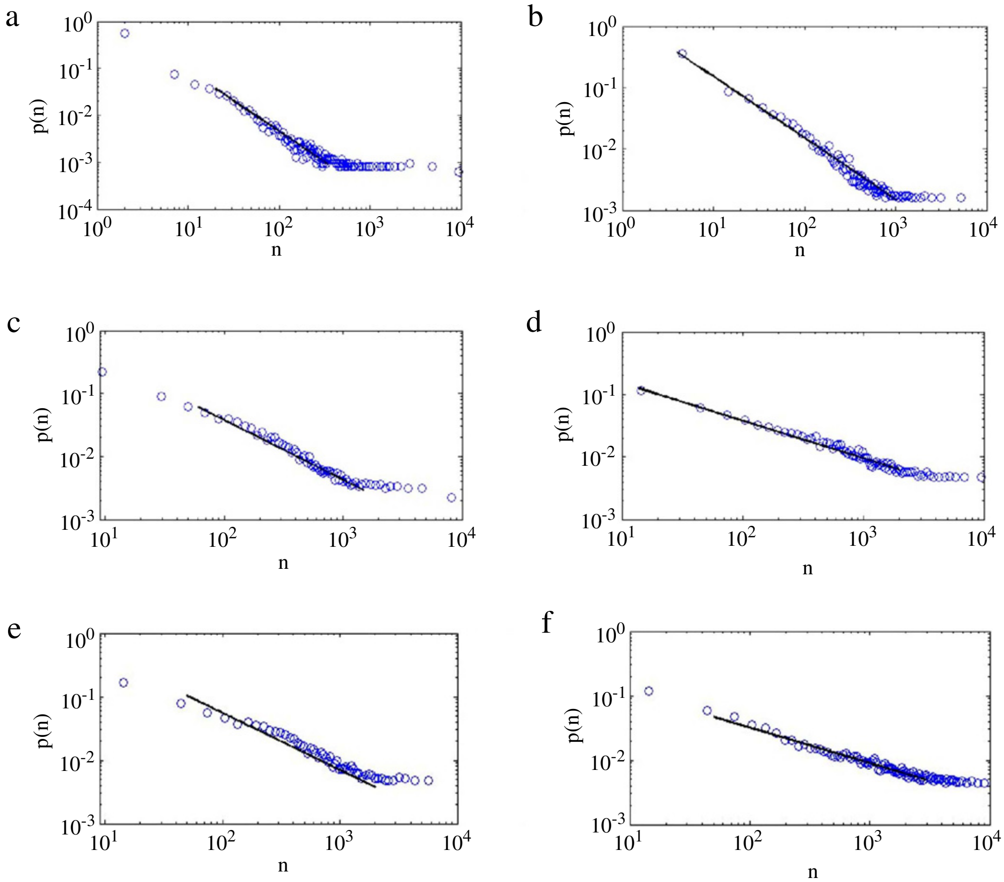

Fig. 4. The total number distribution of microblogs under different levels of active social interaction (a) The total number distribution of microblogs of group 1, slope $\gamma = 1.25$, (b) The total number distribution of microblogs of group 2, slope $\gamma = 1.0$, (c) The total number distribution of microblogs of group 3, slope $\gamma = 0.85$, (d) The total number distribution of microblogs of group 4, slope $\gamma = 0.81$, (e) The total number distribution of microblogs of group 5, slope $\gamma = 0.78$, (f) The total number distribution of microblogs of group 6, slope $\gamma = 0.72$.

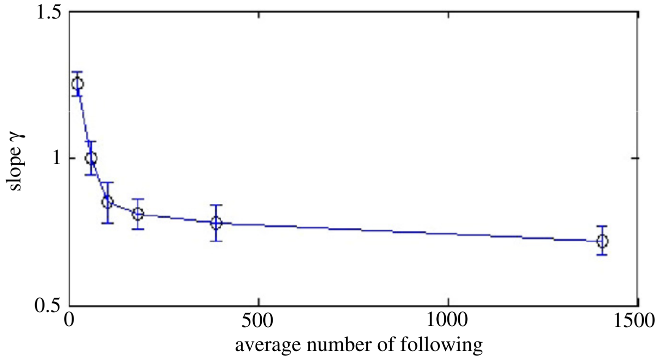

Fig. 5. The exponents of total number distribution of microblogs under different levels of active social interaction.

Table 6
Information of dividing users according to passive social interaction.

Group name	Total number of users	Average number of Follower	Slope γ
Group 1	6457	6	1.50
Group 2	6240	23	1.35
Group 3	6237	57	1.20
Group 4	6337	137	0.95
Group 5	6330	450	0.60
Group 6	6380	128,525	0.45

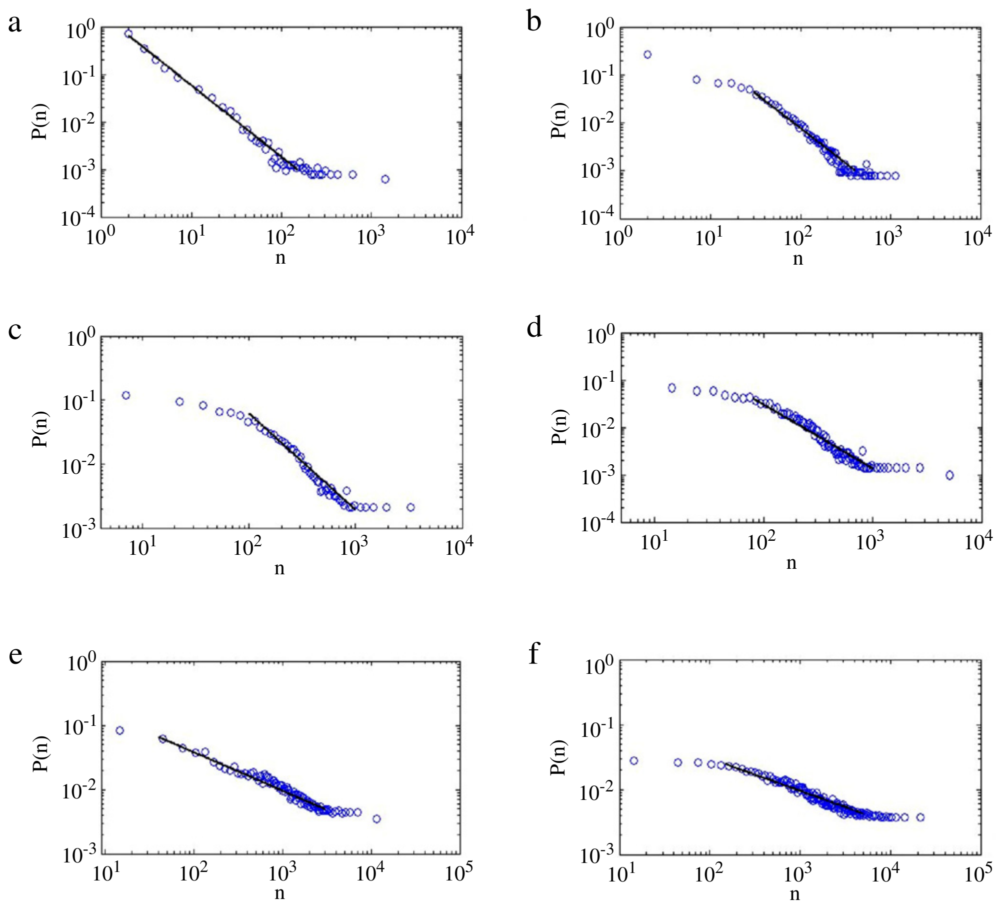

Fig. 6. Total number distribution of microblogs under different levels of passive social interaction (a) Total number distribution of microblogs of group 1, slope $\gamma = 1.50$, (b) Total number distribution of microblogs of group 2, slope $\gamma = 1.35$, (c) Total number distribution of microblogs of group 3, slope $\gamma = 1.20$, (d) Total number distribution of microblogs of group 4, slope $\gamma = 0.95$, (e) Total number distribution of microblogs of group 5, slope $\gamma = 0.60$, (f) Total number distribution of microblogs of group 6, slope $\gamma = 0.45$.

3.2.3. Quantitative analysis based on the relationship of being followed between microblog users

The number of followers (in-degree) reflects the level of a user's passive social interaction received. We will divide all the users into 6 groups according to in-degree from low to high, with approximately the same number of users and level of followers in each group. The information of grouping and the distribution of total number of microblogs in each group can be seen in Table 6 and Fig. 6. As shown in Fig. 6, the distributions of groups follows a power-law distribution. The exponent of the distribution is decreasing with the increase of the number followers (the level of passive social interaction). Fig. 7 reveals that the distribution of total number of microblogs is not obvious when in-degree exceeds 57. That is because there is a limit of users handling their social relationship. The relationship between the behavior of users' being followed and the total number of microblogs is obvious when the total number of following is within about 57.

3.2.4. Microblog user behavior analysis based on social interaction

The character of microblog use behavior from the perspective of both active interaction and passive interaction has been discussed above. In order to analyze the relationship between social interaction and microblog use behavior as a whole,

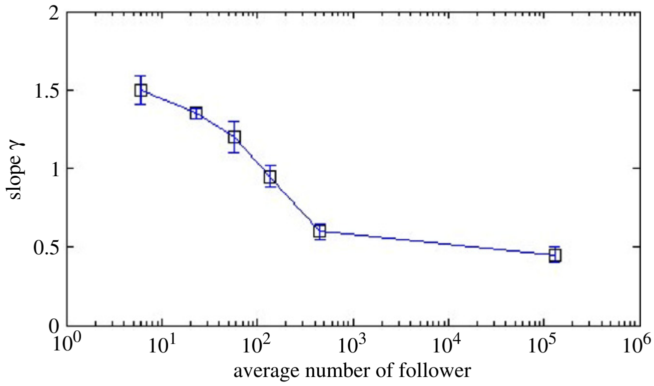

Fig. 7. The exponent of total number distribution of microblogs under different levels of passive social interaction.

Table 7
Information of dividing users according to social interaction.

Group name	Total number of users	Social interaction level	Slope γ
Group 1	6257	35	1.50
Group 2	6298	97	1.30
Group 3	6328	195	0.95
Group 4	6335	393	0.70
Group 5	6330	1,333	0.55
Group 6	6433	128,252	0.48

we define a social interaction coefficient to measure the whole social interaction between microblog users. Specifically, the social interaction coefficient of a user is the sum of the number of following and followers, i.e. the sum of in-degree and out-degree.

We will divide all the users into 6 groups according to social interaction coefficient from low to high, with approximately the same number of users and level of social interaction in each group, as can be seen in Table 7. Fig. 8 shows that the total number distributions of microblogs in most groups follow power-law distribution. The exponent of the distribution is decreasing with the increase of the level of social interaction. Fig. 9, describing the relation between social interaction coefficient and the exponent of the distribution, exhibits that the exponent of the total number distribution of microblogs is negatively correlated with the social interaction coefficient. The exponent decreases fast at low level of social interaction while slow at high level of social interaction.

As can be seen in Fig. 9, there are relationships between microblog using behavior and social interaction. The exponent of total number distribution of microblogs decreases quickly at lower levels of social interaction, when microblog user behavior is closely related with their following behavior. However, at higher levels of social interaction, the change of exponent of total number distribution of microblogs is no longer obvious and microblog use behavior changes little with the change of following behavior. The relationship between microblog use behavior and users' following behavior is less obvious. Nowadays, personalized recommendation has been widely used to encourage microblog users to follow other users in the operation of microblog. According to the empirical analysis above, this technique can be useful within a certain limit of social interaction. As a result, microblog operators should improve the existing personalized recommendation technique to help users find people they are interested in. On the other hand, other methods should be considered to encourage users with a high level of social interaction to use microblog more frequently.

3.2.5. Goodness-of-fit tests

In the paper, we adopt the method of Maximum Likelihood Estimation (MLE) to estimate the power-law exponents [23,24]. However, the method of MLE just allows us to fit a power-law distribution to a given data set and provide estimates of the slope γ. They tell us nothing about whether the power law is a plausible fit to the data. Regardless of the true distribution from which our data were drawn, we can always fit a power law. Therefore, we need some way to tell whether the fit is a good match to the data.

Here, we use the Kolmogorov–Smirnov (KS) statistic to test the power-law hypothesis quantitatively [25,26]. The main procedure is as follows:

First, we fit the empirical data to the power-law model using the method of maximum likelihood estimation and calculate the KS statistic for this fit. Second, we generate a large number of power-law distributed synthetic data sets with slope γ and lower bound $x_{\min}$ equal to those of the distribution that best fits the observed data. Third, we fit each synthetic data set individually to its own power-law model and calculate the KS statistic for each one relative to its own model. Then we simply count the fraction of the time that the resulting statistic is larger than the value for the empirical data. This fraction

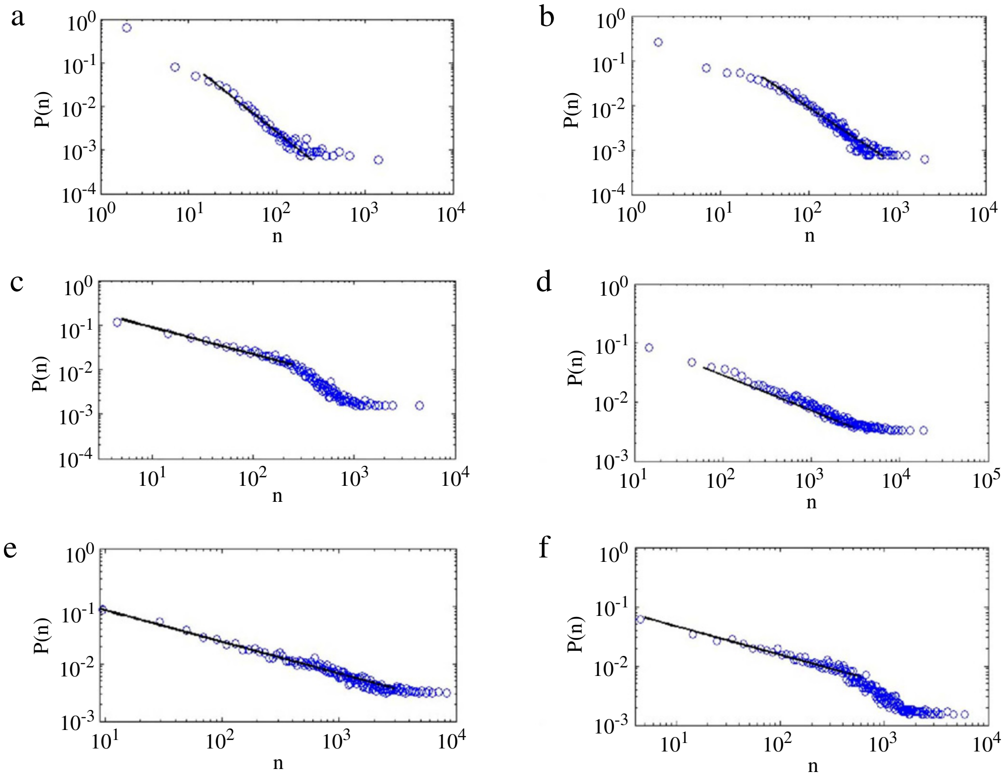

Fig. 8. Total number distribution of microblogs under different levels of social interaction (a) Total number distribution of microblogs of group 1, slope $\gamma = 1.50$, (b) Total number distribution of microblogs of group 2, slope $\gamma = 1.30$, (c) Total number distribution of microblogs of group 3, slope $\gamma = 0.95$, (d) Total number distribution of microblogs of group 4, slope $\gamma = 0.70$, (e) Total number distribution of microblogs of group 5, slope $\gamma = 0.55$, (f) Total number distribution of microblogs of group 6, slope $\gamma = 0.48$.

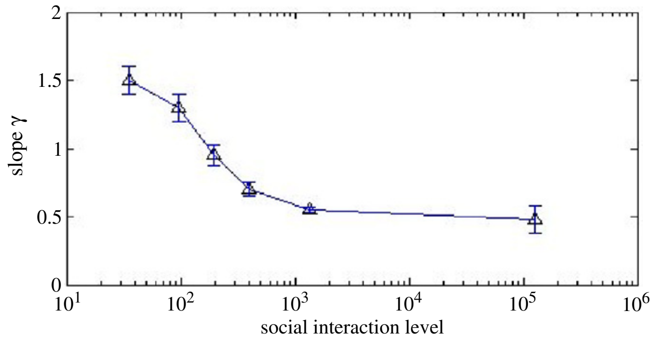

Fig. 9. The exponent of total number distribution of microblogs under different levels of social interaction.

is the p-value. If p is large (close to 1), then the difference between the empirical data and the model can be attributed to statistical functions alone; if it is small ($p < 0.1$), the model is not a plausible fit to the data. As shown in the Tables 8–10, we give the p-value that quantifies the plausibility of the hypothesis, along with a variety of generic statistics for the data such as mean, standard deviation, and maximum value.

4. Model and simulation

The analysis above reveals that the social interaction between microblog users is related with their using behavior. A high level of social interaction leads to more microblogs, and this might be one of the mechanisms that make the power-law distribution of the total number distribution of microblogs. Therefore, building a model from the perspective of social interaction will be discussed in this section.

Table 8
Basic parameters of the data sets described in Section 3.2.2, along with their power-law fits and the corresponding p-values (statistically significant values are denoted in **bold**) (here and after n is the number of data in each group and x is the value of data).

Quantity	n	$\langle x\rangle$	σ	$x_{\max}$	$x_{\min}$	slope γ	p-value
Group 1	6234	58.84	326.53	10,713	2 ± 1	1.25	**0.44**
Group 2	6291	149.30	613.43	27,355	9 ± 2	1.00	**0.62**
Group 3	6320	309.61	637.86	12,246	9.5 ± 2	0.85	**0.53**
Group 4	6364	534.60	915.10	16,935	14.5 ± 4	0.81	**0.68**
Group 5	6390	847.87	1376.80	20,658	14 ± 7	0.78	**0.25**
Group 6	6382	1251.20	1935.66	24,005	22 ± 11	0.72	**0.42**

Table 9
Basic parameters of the data sets described in Section 3.2.3, along with their power-law fits and the corresponding p-values.

Quantity	n	$\langle x\rangle$	σ	$x_{\max}$	$x_{\min}$	slope γ	p-value
Group 1	6457	11.22	52.11	2,325	3 ± 1	1.50	**0.29**
Group 2	6240	50.59	89.92	1,662	4.5 ± 1	1.35	**0.36**
Group 3	6237	168.28	277.06	7,762	9.5 ± 2	1.20	**0.09**
Group 4	6337	418.48	560.44	13,093	9.5 ± 4	0.95	**0.17**
Group 5	6330	902.16	1255.45	20,658	5 ± 2	0.60	**0.20**
Group 6	6380	1692.48	2386.30	30,240	14.5 ± 5	0.45	**0.34**

Table 10
Basic parameters of the data sets described in Section 3.2.4, along with their power-law fits and the corresponding p-values.

Quantity	n	$\langle x\rangle$	σ	$x_{\max}$	$x_{\min}$	slope γ	p-value
Group 1	6257	16.78	57.84	2,325	4.5 ± 1	1.50	**0.36**
Group 2	6298	74.65	131.76	2,183	3 ± 1	1.30	**0.41**
Group 3	6328	214.37	331.24	7,762	9 ± 2	0.95	**0.05**
Group 4	6335	490.95	696.82	13,477	12 ± 4	0.70	**0.56**
Group 5	6330	953.62	1464.99	20,658	20 ± 7	0.55	**0.67**
Group 6	6433	1408.66	2238.40	30,240	42 ± 13	0.48	**0.03**

4.1. Model based on social interaction

Based on the empirical analysis above, considering the relationship between social interaction and microblog behavior of microblog users, we proposed that social interaction between microblog users has an influence on their using behavior. Users with close engagement with their follower or following are more likely to post microblogs. The behavior of their followings is the spontaneous drive that makes microblog users post microblogs, which is a kind of relatively static influence. On the other hand, the forwarding and commenting behavior from other users are the inducing drive and the corresponding influence is relatively dynamic. Under the influence of those two factors, the distributions of microblog user behavior are power-law distributions. The details of our assumption were assumed to be as follows.

(1) The confluence of inducing drive and spontaneous drive leads to the behavior of posting microblogs.
(2) The inducing drive lies into that microblog user behavior is influenced by the forwarding and commenting behavior of their friends (follower and following in microblog world). More forwards and comments lead to more microblogs.
(3) The spontaneous drive means that users with more microblog friends are more likely to post microblogs.

According to the assumptions above, a social network has been built to describe the social relationship between microblog users. To simplify the model, we built an undirected and unweight social network based on BA model [27] to simulate microblog social network. The BA model, designed to explain the mechanism of power-law distribution, is a scale-free network proposed by Barabasi and Albert to characterize the two important properties of real network, growth and preferential attachment. Growth means that the size of the real network is increasing. Preferential attachment means that the new node is more likely to connect with nodes with larger degree. In contrast, the microblog social network has the same properties. First, the size of the microblog social network is increasing all the time. According to the statistics report published by the China Internet Network Information Center (CNNIC) on June 19, 2011, microblog has become the fast growing Internet application mode in China, with the growth of its users at 208.9%. After registering to use microblog, users would not only follow people they know from real life, but also the stars in the microblog world, i.e., nodes with larger degree in the microblog social network. As a result, the BA model can be used to simulate the microblog social network.

Based on the microblog social network simulated by BA model, our model, social interaction based microblog using a behavior model, can be explained below.

(1) s_i^t stands for the state of node i at time step t. $s_i^t = 0$ means that node i does not post a microblog at time step t, while $s_i^t = 1$ means that node i posts a microblog at time step t. At the first time step in our simulation, the states of r nodes are randomly selected to be changed to 1. States of the rest nodes are set to 0.

(2) At time step t, the forwarding and commenting behavior of user i have an impact on its neighboring nodes' microblog using behavior. As a result, the states of those nodes might change at time step $t+1$. $E(i, t)$ stands for the influence that node i could have on its neighboring nodes at time step t.

$E(i, t) = c_i^* s_i^t$, c_i stands for the authority of node i, i.e. the social influence of node i in the whole social network, and can be quantified by the relative point centrality. $c_i = k_i/(N-1)$, k_i is the degree of node i, and N is the size of simulated microblog social network. Under this circumstance, if a user is an isolated node in the social network, and does not interact with other users, its influence in the social network is zero.

(3) At time step $t+1$, the behavior of node j would change because of the spontaneous drive and inducing drive it received at time step t. If node j has k_j neighboring nodes, the spontaneous drive is from its social network properties and the inducing drive comes from the behavior of its k_j neighboring nodes. Therefore, node j will have k_j+1 influencing sources and the effect of each source on node j is $1/(k_j+1)$ of the source's original influence. For example, if the original influence that node i can have on its neighboring nodes is $E(i, t)$, the influence that node j could receive from node i would be $E(i, t)^* 1/(k_j+1)$.

(4) At time step $t+1$, the total influences of the k_j neighboring nodes on node j are

$$I_{j1} = \sum_{i=1}^{k_j} \left(\frac{1}{k_j+1}\right)^* c_i^* s_i^t.$$

(5) At time step $t+1$, the spontaneous drive that node j received is $I_{j2} = k_j/(k_j+1)$. Here we use k_j to stand for the influence of social interaction on the microblog using the behavior of node j. The spontaneous drive on node j only occupies $1/(k_j+1)$ of its original influence.

(6) Finally, the total influence node j received at time step $t+1$ is $I_j = I_{j1} + I_{j2}$. In all the N nodes, randomly select M nodes which receive a high level of total influence, and change their status to 1. It means that those nodes will post microblogs while the remaining users will not perform any activity.

4.2. Model simulation and explanation

In the process of simulating microblog social network based on BA model, the total number of microblog users was set to 10,000. The step size in our simulation is 1000 and the number of users posting microblogs at the first step was $r = 50$. To compare with the empirical data in Table 7, the degree was set to 35, 97, 195, 393 and 1333 respectively. The simulation results are illustrated as Fig. 10, and compared with empirical data in Fig. 11.

As can be seen in Figs. 10 and 11, the total number distribution of microblogs can be simulated as a power-law distribution. While increasing the average degree of the simulated social network, the level of social interaction increases. At the same time, the exponent of total number distribution of microblogs decreases. This coincides with the actual occurrence.

5. Conclusions

Microblog has recently become an influential Internet Service for information publishing. The relationship between the social relation of microblog users and their behavior has been explored in this paper. Based on analysis of data collected from Sina microblog, it has been proved that the microblog social network has a shorter diameter of connected graph, average path length, higher average clustering coefficient, and presents small-world characteristic. The degree distributions of in-degree, out-degree and relative-degree of microblog users are power-law. The distribution of the interval time and total number of users posting microblogs also obey a power-law distribution. By dividing users in different groups according to their following and being followed behavior, the analysis results show that the exponent of the total number distribution of microblogs is negatively correlated with the degree of each user. Particularly, when degrees of users are limited within about 100, this correlation is more obvious.

Based on the empirical analysis above, we proposed a social network based human dynamics model in this paper, and pointed out that inducing drive and spontaneous drive led to the behavior of posting microblogs. The simulation results of our model match well with the practical situation.

Regarding the mechanisms of user's online behavior, the interest driven model and social identity driven model have also been proved in previous studies. Thus there may be multiple driving mechanisms for the user's behavior. The Sina microblog has just experienced about 3 years of development. In the early stage of development, Sina adopted the marketing strategy which mainly focused on extending users scale. And building up the relations between users is a good approach to realize the strategy. So we think social relations have a greater impact on the users' behavior at this stage. However, as illustrated in this paper, the impact of social relation decreases with the increasing of the users' followers and followings. When the social network of microblog users becomes relatively stable, online behaviors of users become their habits. At this stage, interest driven and social identity driven models would be more reasonable. Thus, we think that there may be different dominant mechanisms for users' behavior at different development stages of microblog.

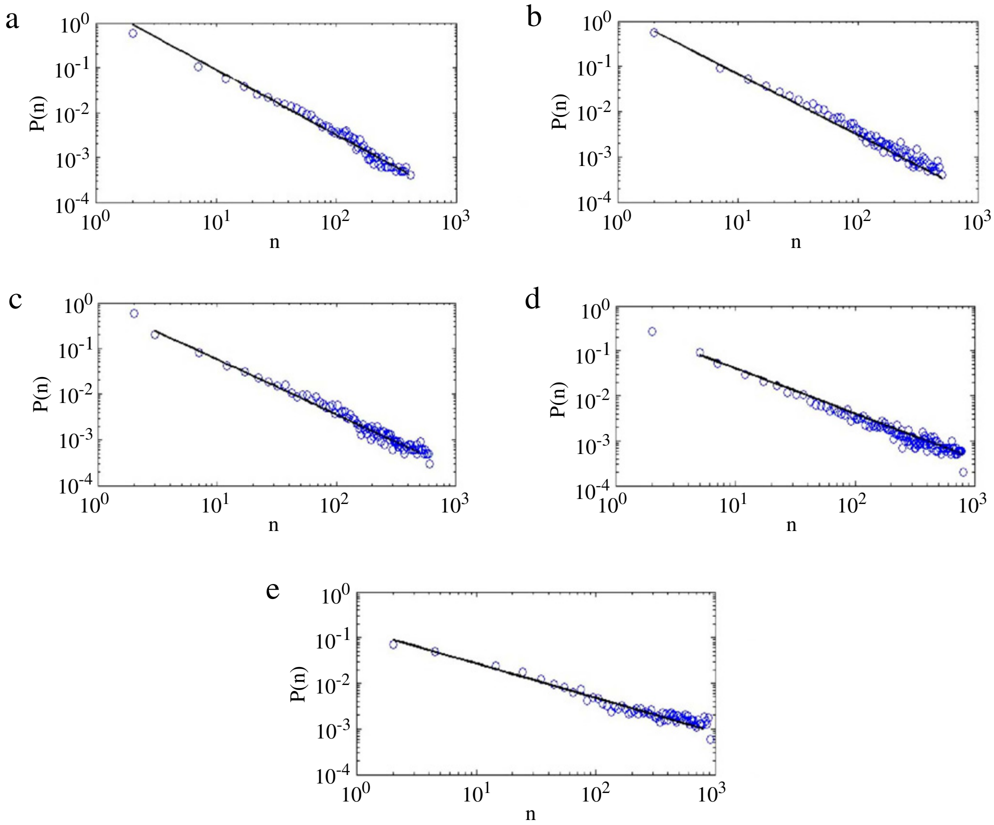

Fig. 10. The model simulation results under different level of social interaction. (a) Total number distribution of microblogs with the average degree at 35, slope $\gamma = 1.45$, (b) Total number distribution of microblogs with the average degree at 97, slope $\gamma = 1.35$, (c) Total number distribution of microblogs with the average degree at 195, slope $\gamma = 1.20$, (d) Total number distribution of microblogs with the average degree at 393, slope $\gamma = 1.00$, (e) Total number distribution of microblogs with the average degree at 1333, slope $\gamma = 0.75$.

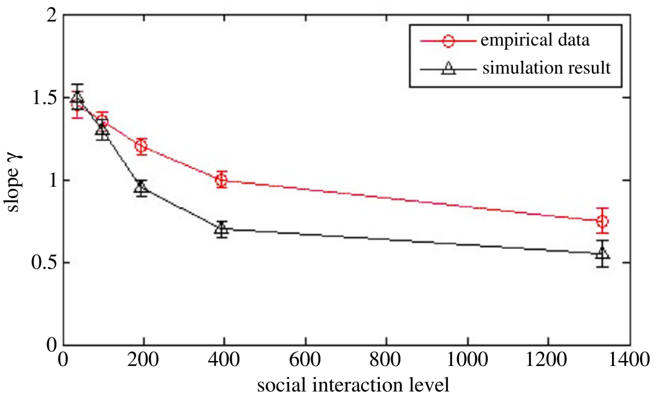

Fig. 11. The comparison between the empirical data and simulation results.

Acknowledgments

This work was supported by the Program for New Century Excellent Talents in University (NCET-11-0597), and the Fundamental Research Funds for the Central Universities (2012RC1002).

References

[1] R.D. Malmgren, D.B. Stouffer, A.E. Motter, L.A.N. Amaral, A Poissonian explanation for heavy tails in e-mail communication, Proceedings of the National Academy of Sciences 105 (47) (2008) 18153–18158.
[2] R.D. Malmgren, D.B. Stouffer, A.S.L.O. Campanharo, L.A.N. Amaral, On universality in human correspondence activity, Science 325 (5948) (2009) 1696–1705.
[3] A.L. Barabási, The origin of bursts and heavy tails in human dynamics, Nature 435 (7039) (2005) 207–211.
[4] J. Candia, M.C. González, P. Wang, T. Schoenharl, G. Madey, A.L. Barabasi, Uncovering individual and collective human dynamics from mobile phone records, Journal of Physics A: Mathematical and Theoretical 41 (22) (2008) 224015.
[5] W. Hong, X.P. Han, T. Zhou, B.H. Wang, Heavy-tailed statistics in short-message communication, Chinese Physics Letters 26 (2) (2009) 028902.
[6] Q. Yan, L.L. Yi, L.R. Wu, Human dynamic model co-driven by interest and social identity in microblog community, Physica A: Statistical Mechanics and its Applications 319 (2012) 1540–1545.
[7] F. Luo, J.M. Yang, Z.H. Li, Study of human dynamics from QQ group message, Journal of South China University of Technology (Social Science Edition) 13 (4) (2011) 14–19.
[8] G.X. Chen, X.P. Han, B.H. Wang, Multi-level scaling properties of instant-message communications, Physics Procedia 3 (5) (2010) 1897–1905.
[9] X.P. Han, T. Zhou, B.H. Wang, Modeling human dynamics with adaptive interest, New Journal of Physics 10 (7) (2008) 073010.
[10] Q. Yan, L.R. Wu, Impact of bursty human activity patterns on the popularity of online content, Discrete Dynamics in Nature and Society 2012 (2012) 872908.
[11] M.S. Shang, G.X. Chen, S.X. Dai, B.H. Wang, T. Zhou, Interest-driven model for human dynamics, Chinese Physics Letters 27 (4) (2010) 048701.
[12] A. Grabowski, Human behavior in online social systems, The European Physical Journal B — Condensed Matter and Complex Systems 69 (4) (2009) 605–611.
[13] T. Zhou, H.A.T. Kiet, B.J. Kim, B.H. Wang, P. Holme, Role of activity in human dynamics, Europhysics Letters 82 (2) (2008) 28002.
[14] Q. Yan, L.R. Wu, L.L. Yi, Research on the human dynamics in the mobile communities based on social identity, Discrete Dynamics in Nature and Society 2012 (2012) 672756.
[15] K. Christine, B. Martin, Identification of influencers — Measuring influence in customer networks, Decision Support Systems 46 (1) (2008) 233–253.
[16] K. Christine, B. Martin, Leveraging Network Effects for Predictive Modeling in Customer Relationship Management, in: Proceedings of the 15th Annual Workshop on Information Technologies & System Las-Vegas USA, 2005, p. 78.
[17] X.C Yang, X.H. Zhang, W.H. Shi, The impact of relational attributes on mobile phone user's communication behavior, Journal of BUPT (Social Science Edition) 10 (5) (2008) 43–47.
[18] B. Daniel, G. Swann, Network effects and the choice of mobile phone operator, Journal of Evolutionary Economics 16 (1) (2006) 65–84.
[19] W. Bachnik, S. Szymczyk, et al., Quantitative and sociological analysis of blog networks, Acta Physica Polonica B 36 (10) (2005) 3179–3191.
[20] A. Grabowski, N. Kruszewska, R.A. Kosiński, Properties of on-line social systems, The European Physical Journal B 66 (1) (2008) 107–113.
[21] Z.D. Zhao, H. Xia, M.S. Shang, T. Zhou, Empirical analysis on the human dynamics of a large-scale short message communication system, Chinese Physics Letters 28 (6) (2011) 068901.
[22] N.N. Li, N. Zhang, T. Zhou, Empirical analysis on temporal statistics of human correspondence patterns, Physica A: Statistical Mechanics and its Applications 387 (25) (2008) 6391–6394.
[23] L. Wasserman, All of Statistics: A Concise Course in Statistical Inference, Springer-Verlag, Berlin, 2003.
[24] H. Bauke, Parameter estimation for power-law tail distributions by maximum likelihood methods, European Physical Journal B 58 (2007) 167–173.
[25] A. Clauset, C.R. Shalizi, M.E.J. Newman, Power-law distribution in empirical data, SIAM Review 51 (4) (2009) 661–703.
[26] M.L. Goldstein, S.A. Morris, G.G. Yen, Problems with fitting to the power-law distribution, European Physical Journal B 41 (2004) 255–258.
[27] A.L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.