Empirical analysis of the evolution of follower network: A case study on Douban
— **Source link** ☑

Junzhou Zhao, John C. S. Lui, Don Towsley, Xiaohong Guan ...+1 more authors

Institutions: Xi'an Jiaotong University, The Chinese University of Hong Kong, University of Massachusetts Amherst

Related papers:

- What is Twitter, a social network or a news media?

- The slashdot zoo: mining a social network with negative edges

- Trust management for the semantic web

- Measurement and analysis of online social networks

- Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters

Empirical Analysis of the Evolution of Follower Network: A Case Study on Douban

Junzhou Zhao*, John C. S. Lui†, Don Towsley‡, Xiaohong Guan* and Yadong Zhou*

*MOE KLINNS Lab, Xi'an Jiaotong University
Email: {jzzhao,xhguan,ydzhou}@sei.xjtu.edu.cn
†Department of Computer Science & Engineering, The Chinese University of Hong Kong
Email: cslui@cse.cuhk.edu.hk
‡Computer Science Department, University of Massachusetts at Amherst
Email: towsley@cs.umass.edu

Abstract—*Follower networks* such as Twitter and Digg are becoming popular form of social information networks. This paper seeks to gain insights into how they evolve and the relationship between their structure and their ability to spread information. By studying the Douban follower network, which is a popular online social network in China, we provide some evidences showing its suitability for information spreading. For example, it exhibits an *unbalanced bow-tie structure* with a large *out-component*, which indicates that the majority of users can spread information widely; the effective diameter of the *strongly connected component* is shrinking as the user base grows, which facilitates spreading; and the *transitivity* property shows that people in a follower network tend to shorten the path of information flow, i.e., it takes fewer hops to spread information. Also, we observe the following users' behaviors, a user's *following activity* decays exponentially during her lifetime and the following behaviors differ according to the age of the account. These findings provide a deep understanding on the evolution of follower networks, and can provide guidelines on how to build an efficient information diffusion system.

I. Introduction

In recent years, social networking sites(SNS) such as Facebook, Twitter, MySpace, LinkedIn, Flickr and Orkut, have become major social exchange platforms and are changing the way we interact with others. With the help of SNSs, people can contact friends, write online diaries, share information, track the latest news and so on. The popularity of SNSs has also attracted many researchers to study their topological properties [1], [2], evolution [3]–[5], user behaviors [6], [7] and the network dynamics [8].

In this paper, we study the Douban social network [9], a popular Chinese Web 2.0 website providing an exchange platform for reviews and recommendations on movies, books and music. Similar to Twitter and Digg, Douban is a "*follower network*" in which users can broadcast short messages to their followers. A distinct feature of a following relationship is that the person being followed can provide useful information to all his/her followers. In this way, information can flow from publishers to followers in a direction opposite to the following direction (or following relationship). Since the following relationship is directed, people who have following relationships may not know each other at all, and this is very different from the friendship relationship in other online social networks, say, Facebook.

We seek to understand the evolution of follower networks. In particular, *how does a follower network evolve? What is the relationship between network structure and information spreading? How do people establish their follower networks?* These are questions of fundamental importance as their answers would provide a better understanding of follower networks, and allow social networking service providers to improve their systems so to attract more users, and make their applications more scalable, or to help advertisers select potential targets and carry out effective advertisement, or to aid researchers to have a deeper understanding of users' following activities and construct an accurate model to characterize users' behaviors.

To facilitate our studies, we represent the Douban follower network as a time varying directed graph $G(t) = (V(t), E(t)), t = 0, 1, \cdots$, where $V(t)$ and $E(t)$ are the set of vertices and edges at time t respectively. A node $u \in G(t)$ can represent a user or an organization. A directed edge $e_t = (u, v)$ represents a *following relationship*: that node u follows node v at time t. Here, v is a *publisher* of u while u is called a *follower* of v.

In general, a follower network differs from other kinds of existing networks in SNSs. First, follower network is different from friendship network such as Facebook, where relationships are undirected or symmetrical. In a follower network, relationships are usually directed and asymmetrical. Furthermore, the establishment of a following relationship is easier than that of a friendship, which must be confirmed by two parties. Secondly, a follower network differs from a citation network where papers rarely cite all other papers and old papers cannot cite new papers, and information only flows from old nodes to new nodes with rare reciprocity edges. Third, a follower network differs from a contact network such as Flickr, which is directed and has bidirectional edges, however it is not designed for information spreading like follower networks.

In this paper, we study the evolving properties of the Douban follower network. We summarize the results as follows.

• First, we find that the degree distribution has both power-law

and log-normal features. The Pearson correlation coefficient of the in-degree and out-degree of nodes is about 0.6, which indicates that nodes that have more publishers tend to have more followers. The *unbalanced bow-tie structure* with a large *out-component* and a small *in-component* indicates that follower networks have the potential to spread information efficiently.

• Second, we find that both the in-degree and out-degree show the effect of preferential attachment. They also exhibit transitivity features that provide evidence that the Douban follower network has the tendency to shorten the path of information diffusion.

• Third, during the lifetime of each user, the *following activity* decays exponentially with the age of a user. Users of different age groups show different following activities. Younger users are more likely to actively follow others, while older users are more likely to be followed.

The outline of this paper is as follows. We give a detailed description of Douban and the associated dataset in Section II. We analyze the network evolution and present its properties. Section IV and V present our analysis of the growth mechanisms and people's following behavior respectively. In Section VI, we discuss related works Section VII concludes.

II. DATASET DESCRIPTION

A. Douban and its follower network

Douban [9] is a Chinese Web 2.0 web site providing user review and recommendation services for books, movies and music. Douban was launched on March 6, 2005. It is also the largest online database for Chinese language books, movies and music. The users of Douban are mainly white collar workers and students who enjoy reading books, listening to music and watching movies. People use Douban to find favorable books, movies and music based on recommendations of other users. In short, Douban provides user services mainly by its powerful recommendation system.

Users are connected by the Douban follower network. Users can follow other users or be followed by others *freely*, i.e., user A can follow user B without the permission from B. Both followers and publishers are listed on a user's personal homepage and can be viewed by others without privacy restrictions. This follower/following mechanism is similar to Twitter and Digg.

There are approximately 6 million accounts in Douban as of August 14, 2009. Our Douban network dataset contains *all* users since its establishment till August 14, 2009. There are around 1.6 million accounts which have at least one publisher or follower. For each account, we know its registration date which is accurate up to the "day" granularity. For each edge, we know the exact creation time which is accurate up to the "second" granularity. Although Douban users can remove their publishers or followers freely, we believe this is rare and we do not consider its effect here. Detailed statistics of the Douban follower network are shown in Table I.

TABLE I
BASIC INFORMATION OF THE DATASET.

Time interval	2005-03-06∼2009-08-14
Nodes	$1,614,288$
Edges	$14,573,170$
Nodes in largest SCC*	$807,619(50.03\%)$
Edges in largest SCC	$13,277,614(98.81\%)$
Nodes in largest WCC**	$1,595,100(98.81\%)$
Edges in largest WCC	$14,558,294(99.90\%)$
Reciprocity edges	$10,100,854(69.31\%)$
90% effective diameter	6.55
Average clustering coefficient	0.10
Average in/out-degree	9

* The Strongly Connected Component (SCC), a directed subgraph in which any two nodes can reach to each other.
** The Weakly Connected Component (WCC), a directed subgraph in which any two nodes can reach to each other without considering direction of edges.

B. Accounts arrival process and account activity

We consider the accounts arrival process which describes how new accounts arrive (are created) over time. This will shed light on the popularity of a SNS. Figure 1a illustrates the number of accounts (nodes) versus time while Figure 1b illustrates the number of edges versus the number of nodes.

In Douban, the number of accounts increases exponentially during the first 25 months with a function $76632(e^{0.07t} - 1)$, and then linearly with a function $-878766 + 49359t$. The number of edges versus the number of nodes shows a linear growth at first, and then exponential growth as there are more nodes in the network. This transformation is due to the change of *account activity*. Not all accounts are active after registration. We say an account is *active* if it creates at least one following relationship during a given time interval, otherwise it is *inactive*. We define *account activity* as the fraction of active accounts of a given time interval.

In Figure 1a, we also plot the monthly *account activities* of Douban. One observes a big jump in account activity at the 36th month. This is due to the surge in popularity of Douban at that time (around the beginning of 2008). We will show how this increased account activity affects some network level properties in the next section.
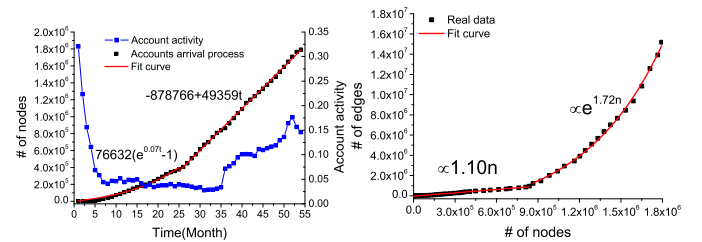
(a) Accounts arrival process and account activity.

(b) The number of edges versus the number of accounts.

Fig. 1. The nodes and edges arrival process. The 1st month is Mar. 2005 and the 54th month is Aug. 2009.

III. EVOLUTION OF THE DOUBAN FOLLOWER NETWORK

In this section, we study how different network level properties of the Douban follower network, such as the degree

distribution, degree correlation and component structure properties evolve over time so as to better understand its evolution.

A. Evolution of the degree distribution

Figure 2 illustrates the degree distributions of four snapshots: From the beginning of 2005 to the end of year 2006, 2007, 2008 and 2009 respectively. Both the in- and out-degree distribution reflect *heavy-tail* features and show both power-law and log-normal characteristics. It is difficult to say which is more dominant because there has been a long-standing debate about how to distinguish power-law from log-normal [10]. From the dataset, it looks like that the log-normal is more dominant because the CCDF plot is not a pure strict straight line. In the next section, we will show that there does exist preferential attachment effect which could lead to the power-law degree distribution.

Note that there is a noticeable change-point on the out-degree distribution curve of 2005-2008 at 1000, and the distribution shows a fast decline after this point. This is due to Douban's limitation that a person could only follow 1000 people at most before April 2008. But later this restriction is removed.
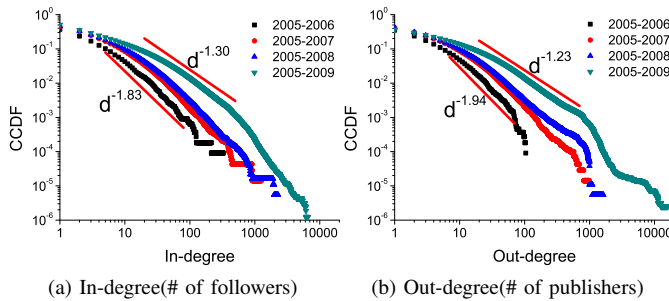
(a) In-degree(# of followers)   (b) Out-degree(# of publishers)

Fig. 2.   Evolution of the degree distribution. The y-axis is the Complementary Cumulative Distribution Function(CCDF) representing the probability $P(X > x)$.

B. Evolution of the degree correlation

In follower networks, the number of followers is a direct indicator of how famous people are. People can't force others to follow him, however there is a strong correlation between the number of publishers (out-degree) and the number of followers (in-degree). As the heat map shows in Figure 3a, from which we observe that the majority of the nodes are distributed around $y = x$. This indicates that users in a follower network are likely to have a comparable followers and publishers, or people have more publishers are also likely to have more followers. In order to show how this property is invariant over time, we compute the *Pearson correlation coefficient* between nodes' in-degree and out-degree of graph $G(t)$ at each time instance t.

The value of Pearson correlation coefficient lies between -1 and 1, with 1 (-1) implies a positive (negative) correlation of two variables. We find that the Pearson correlation coefficient of Douban is always positive as shown in Figure 3b and fluctuates around 0.6 before the 35th month. After the change in

account activity at month 36, it drops to 0.2 but subsequently increasing monotonically.

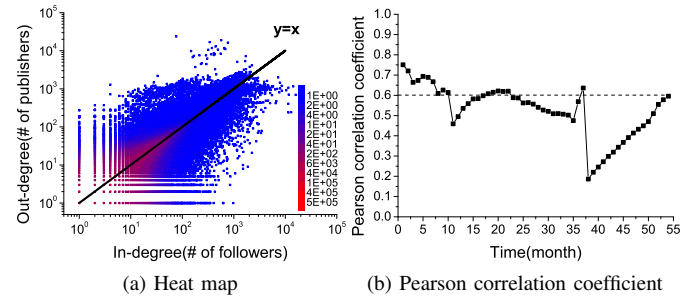(a) Heat map   (b) Pearson correlation coefficient

Fig. 3.   Degree correlation of Douban follower network

C. Evolution of the component structure properties

Douban follower network is dominated by a giant weakly connected component (WCC), which takes account of more than 90% of the entire graph. The largest WCC is characterized by a *bow-tie structure* [11], which is illustrated in Figure 4a. The middle component, labeled by "SCC", is the largest strongly connected component (SCC), in which any two users can reach each other. Along the following relationship direction, users in the in-component, labeled by "In", can reach users in the other two components. While users in the out-component, labeled by "Out", cannot reach users in the other two components. In other words, there does not exist *a path with following relationships*. So users in "Out" can't reach users in "SCC" or "In". Users in different components play different roles in information spreading. Note that a message can propagate in a follower network via *retweeting*, e.g., a user can re-post a message from people he/she follows. Therefore, the direction of information flow is *opposite* to the following relationship direction. So if a message is generated by a user who is in the out-component, it will have greater chance to spread to other users than a message generated by a user who is in the in-component.

The size of these three components as shown in Figure 4b. We find that the fractions of nodes in these three components are different. Generally, the size of the SCC component is the largest while the size of the in-component is the smallest. Because information spreading occurs in an *opposite* direction of following relationships, this *unbalanced bow-tie structure* indicates that the majority of users in Douban follower network have the potential to spread information widely.

Because the SCC is the largest component, and serves the role of a *relay* between in- and out-component, its diameter will affect the speed of spreading. From our measurement, we discover that as the network grows, the effective diameter of the SCC actually *decreased*, as shown in Figure 4c. *Effective diameter* [3], [12] is defined as the smallest distance such that at least 90% of the connected node pairs can be reached, which is considered to be more a robust measure than the standard diameter of graph and can reflect the truth of network diameter evolution. Note that in [3], [5], authors also reported the phenomenon of effective diameter shrinkage.
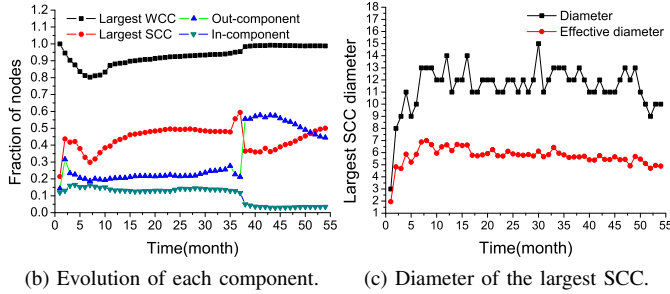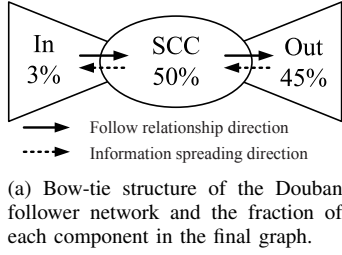
(a) Bow-tie structure of the Douban follower network and the fraction of each component in the final graph.

(b) Evolution of each component.

(c) Diameter of the largest SCC.

Fig. 4. Evolution of component structure.

IV. GROWTH MECHANISMS

In this section, we seek to understand the underlying mechanisms that drive the growth of a follower network. In particular, we want to see whether the network's growth is due to preferential attachment [13], [14] which can generate power-law characteristics in the degree distributions. We also analyze the *transitivity* feature, which characterizes the following relationships between three users.

A. Preferential attachment

As shown in Figure 2, one may argue that the degree distribution is more like a log-normal than power law. In here, we show that there exists a preferential attachment effect.

Since each edge in a follower network has a direction, we divide preferential attachment into *preferential creation* and *preferential reception* to study which nodes are more likely to create edges and which are more likely to receive edges. Preferential creation, denoted by $p_{cre}(d)$, is the probability that a node with attribute d creates an edge. Preferential reception, denoted by $p_{rec}(d)$, is the probability that a node with attribute d receives an edge. The attributes of a node can be its in-degree, out-degree and total degree. We extend the preferential attachment metric for an undirected network in [4], and $p_{cre}(d)$ and $p_{rec}(d)$ are expressed as follows:

$$p_{cre}(d) \propto \frac{\sum_{i=1}^{|E|} \mathbf{1}(e_i = (u,v) \wedge Attr_{t_{i-1}}(u) = d)}{\sum_{i=1}^{|E|} |\{u : Attr_{t_{i-1}}(u) = d\}|}, \quad (1)$$

$$p_{rec}(d) \propto \frac{\sum_{i=1}^{|E|} \mathbf{1}(e_i = (u,v) \wedge Attr_{t_{i-1}}(v) = d)}{\sum_{i=1}^{|E|} |\{u : Attr_{t_{i-1}}(u) = d\}|}, \quad (2)$$

where $e_i = (u,v)$ is the i-th edge(order by creation time) of the graph pointing from u to v. t_i is the creation time of e_i. $Attr_t(u)$ represents the attribute of node u at time t. The indicator function $\mathbf{1}(B) = 1$ if condition B is true, otherwise $\mathbf{1}(B) = 0$.

We find that the attributes "in-degree", "out-degree" and "total degree" all exhibit some form of preferential attachment property and the both probabilities have a form of d^α. In Figure 5 we only show the results of preferential creation for out-degree and preferential reception for in-degree. The others are similar but with different value of α. After a normalization to make sure that $\sum_d p_{cre}(d) = 1$ and $\sum_d p_{rec}(d) = 1$, the values of α are shown in Table II. The attributes, out-degree for preferential creation and in-degree for preferential reception have the largest exponents, which indicate that preferential creation is more sensitive to nodes' out-degree and preferential reception is more sensitive to nodes' in-degree. Furthermore, both exponents are very close to 1, indicating that preferential attachment does exist during the growth of the Douban follower network.
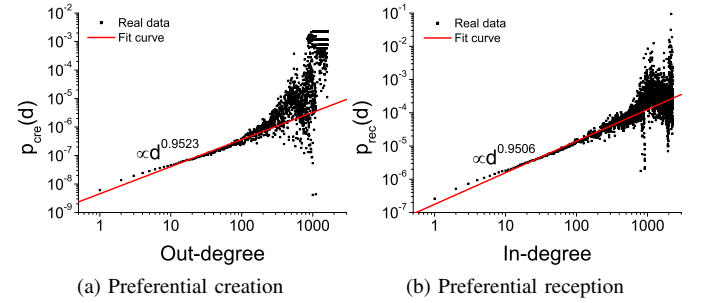

(a) Preferential creation

(b) Preferential reception

Fig. 5. Preferential attachment

TABLE II
THE VALUE OF α VERSUS NODE ATTRIBUTE.

| d^α	In-degree	Out-degree	Total degree
$p_{cre}(d)$	0.7078	0.9523	0.8240
$p_{rec}(d)$	0.9506	0.7299	0.8332

B. Transitivity features

In social networks, friends of my friends are likely to be my friends. This is known as the *transitivity phenomenon* [15]. When relationships are asymmetrical as in a follower network, we want to find out whether transitivity exists.

A *closed following triad* is a following relationship between three people ordered by time. Since each edge has two possible directions, there are eight types of closed following triads, which are shown in Figure 6a. In a *random* network, if the probabilities of each following direction between any two persons are equal, then the eight types of closed following triads should have equal or comparable probabilities. We explore whether this is true for the Douban network. If not, then we examine which types of triads are more likely.

We extract all of the closed following triads of the first 35 months. Figure 6b illustrates their respective probabilities and we observe that they occur with *different* frequencies. Generally, $T1$ and $T5$ have higher percentage than others while $T7$ is the least likely. This indicates that relation in a follower network has the *transitive property* [16]. If we use $\mathbf{R}$ to denote the relation in a follower network, and $a\mathbf{R}b$ represents that a follows b, then *transitivity* implies that for
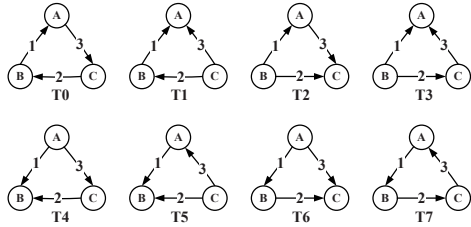
(a) Eight types of closed following triads. A, B and C represent three people. 1, 2 and 3 labeled on the edge represent the creation order.
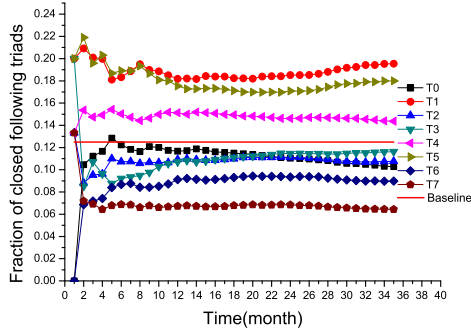
(b) Fractions of the closed following triads versus time.

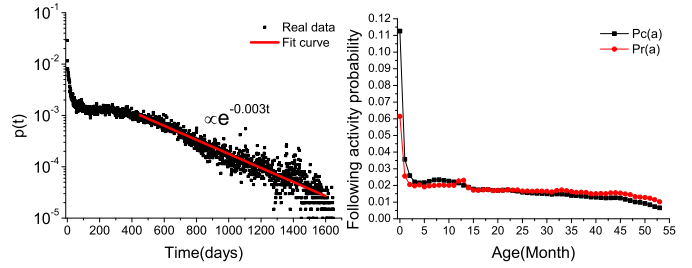Fig. 6.   Closed following triads and evolution of their fractions.

(a) Lifetime following activity

(b) Following activity versus account age

Fig. 7.   Features of users' following activity.

three users x, y and z, if $x\mathbf{R}y$ and $y\mathbf{R}z$, we have $x\mathbf{R}z$. Since $T7$ violates this transitivity property (i.e., it has $C \to A$ instead of $A \to C$), that is why it occurs with the least frequency. In contrast, $T1$ and $T5$ satisfy the transitivity property so they occur with higher frequencies.

Another reason why $T1$ gets the largest fraction may be because the establishment of the edge (C, A) can shorten the information path length: instead of information flowing from A to B and then to C, information now flow from A to C directly. This indicates that users in a follower network have a tendency to *shorten the information flow path*.

The difference between $T1$ and $T5$ is that the targets of C are B's publishers in $T1$ and followers in $T5$. The larger fraction of $T1$ implies that users are more likely to follow the publishers of a user than followers. This raises an interesting question: Are publishers more attractive than followers? To answer this we randomly choosing 1000 users and calculating the probability that the user chooses followers or publishers to follow. We carry the experiment for 50 times and obtain the result that probability in choosing publishers is 9.1% *higher* than choosing followers. This indicates that publishers are more attractive than followers in a follower network.

V. FEATURES OF PEOPLE'S FOLLOWING ACTIVITY

Let $t_r(u)$ denote the registration time of user u in the network. After this time, a user becomes active: either following other users or being followed. We use *following activity* to quantify a user's activity. We study various features of users' following activity and characterize user behavior after he/she joins a follower network.

A. Following activity during lifetime of a user

We explore the following activity of a user during his/her lifetime. Let n_{ut} denote the number of followers of user u on the t-th day after $t_r(u)$, and $p(t)$ is the probability that a user follows a person on the t-th day after $t_r(u)$. We express $p(t) = \frac{\sum_u n_{ut}}{\sum_{u,t} n_{ut}}$. Figure 7a illustrates $p(t)$ and shows that it is generally exponentially decreasing. Specifically, we observe *three-phases* in a user's following activity. A burst of activity occurs during the first 10 days, and then the following activity becomes stable and for approximately 200 days. Then the following activity decreases exponentially with a small decay rate 0.003.

B. Following activity versus account age

We explore following activities of accounts as a function of age. In particular, we ask whether or not younger accounts are more likely to follow others than older accounts.

The age of an account u at time t is defined as $age(u,t) = t - t_r(u)$. The probabilities that an account with age a follows or is followed are denoted by $p_c(a)$ and $p_r(a)$. They are calculated as follows,

$$p_c(a) \propto \frac{\sum_{i=1}^{|E|} \mathbf{1}(e_i = (u,v) \wedge age(u,t_i) = a)}{\sum_{i=1}^{|E|} |\{u : age(u,t_i) = a\}|}, \quad (3)$$

$$p_r(a) \propto \frac{\sum_{i=1}^{|E|} \mathbf{1}(e_i = (u,v) \wedge age(v,t_i) = a)}{\sum_{i=1}^{|E|} |\{u : age(u,t_i) = a\}|}. \quad (4)$$

We find that accounts with different ages show different features, as shown in Figure 7b. Both p_c and p_r show are large during the first few days after the registration, which is consistent with the result in Figure 7a. After that, younger accounts and older accounts exhibit different following activities. In particular, we have $p_c(a) > p_r(a)$ for $a < 11$, and $p_c(a) < p_r(a)$ for $a > 24$. This implies that younger accounts, whose ages are less than 11 months, have a higher probability to follow others than be followed. Older accounts, whose ages are greater than 24 months, have a larger probability of being followed than following others. This indicates that the main activities for younger accounts are to establish connections by actively following others, while older accounts are more likely to receive followers passively.

945

VI. Related work

The power-law degree distribution and small-world effect are two well known findings over the past few decades when researchers study complex networks. Various network models [17] are proposed to explain why complex networks have these properties, e.g., the ER model [18], the preferential attachment model [13], the small-world model [19] and so on. But the majority of these works are based on *static networks* or a small number of snapshots, e.g., the number of nodes is fixed in the ER model and the small-world model while defining edge connection or rewiring rules. A. Mislove *et al* [20] analyze the growth of Flickr based on 104 daily snapshots of the Flickr contact network. Y. Y. Ahn *et al* [1] analyze the evolution of degree distribution, clustering coefficient and diameter of Cyworld using three years snapshots of the Cyworld social network. Static or snapshots of networks can characterize the *macroscopic* properties of networks, but they cannot characterize microscopic properties of networks, such as the reasons for link formation and node connectivity behavior. Several other researchers analyze the continuous time evolution of graphs. J. Leskovec *et al* [3], [4] analyzed the evolution of several real graphs. They observed that these graphs are densifying over time and the average distance between nodes are shrinking. They also analyze the preferential attachment and locality of edge attachment's effect. R. Kumar *et al* [5] study the structure evolution of Flickr and Yahoo! 360. Their measurements reveal the segmentation of these networks in three regions: singletons, isolated communities and a giant component. There are mainly two differences between our work and theirs. Firstly, the follower network we consider is a directed graph and growth mechanisms are inherently different between directed and undirected networks. Secondly, follower network is associated with information diffusion and we analyze the relationship between follower network structure and information spreading.

VII. Conclusion

In this paper, we study the evolution of Douban follower network. We analyze the network evolution and reveal some interesting properties, including degree distribution, degree correlation and component structure properties. Preferential attachment exists both for in-degree and out-degree and the features of transitivity do exist, which indicates that there are biases when people follow others. Also, evolution of follower network reflects some features of people's following behavior, such as the decaying of following activity and differences of people's follow behavior with different account age. A deep understanding of follower network is important because the follower network is designed for information flow. Our work reveals some evidences that follower network is suitable for information spreading. For example, the unbalanced bow-tie structure with a large out-component, the shrinkage of the effective diameter, and the features of transitivity show the shortening of the information flow path length.

References

[1] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of topological characteristics of huge online social networking services," in *WWW'2007*, 2007, pp. 835–844.

[2] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *WWW'2010*, 2010, pp. 591–600.

[3] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," in *KDD'2005*, 2005, pp. 177–187.

[4] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, "Microscopic evolution of social networks," in *KDD'2008*, 2008, pp. 462–470.

[5] R. Kumar, J. Novak, and A. Tomkins, "Structure and evolution of online social networks," in *KDD'2006*, 2006, pp. 611–617.

[6] F. Schneider, A. Feldmann, B. Krishnamurthy, and W. Willinger, "Understanding online social network usage from a network perspective," in *IMC'2009*, 2009, pp. 35–48.

[7] J. Jiang, C. Wilson, X. Wang, P. Huang, Y. Dai, and B. Y. Zhao, "Understanding latent interactions in online social networks," in *IMC'2010*, 2010.

[8] M. G. Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," in *KDD'2010*, 2010, pp. 1019–1028.

[9] Douban, *www.douban.com*.

[10] M. Mitzenmacher, "A brief history of generative models for power law and lognormal distributions," *Internet Mathematics*, vol. 1, pp. 226–251, 2008.

[11] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph structure in the web," *Computer Networks*, vol. 33, no. 1-6, pp. 309–320, June 2000.

[12] S. L. Tauro, C. Palmer, G. Siganos, and M. Faloutsos, "A simple conceptual model for the internet topology," in *GLOBECOM'2001*, vol. 3, 2001, pp. 1667–1671.

[13] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[14] A. Vazquez, "Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations," *Phys. Rev. E*, vol. 67, no. 5, p. 056104, May 2003.

[15] G. Kossinets and D. J. Watts, "Empirical analysis of an evolving social network," *Science*, vol. 311, no. 5757, pp. 88–90, January 2006.

[16] E. W. Weisstein. Transitive. [Online]. Available: http://mathworld.wolfram.com/Transitive.html

[17] A. Goldenberg, A. X. Zheng, S. E. Fienberg, and E. M. Airoldi, "A survey of statistic network models," in *Foundations and Trends in Machine Learning*, vol. 2(2), 2002.

[18] P. Erdös and A. Rényi, "On random graphs," in *Publications Mathematics*, vol. 6, 1959, pp. 290–297.

[19] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, pp. 440–442, 1997.

[20] A. Mislove, H. S. Koppula, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Growth of the flickr social network," in *WOSN'2008*, 2008, pp. 25–30.