An Analysis of Verifications in Microblogging Social Networks - Sina Weibo

Junting Chen† and James She*
*HKUST-NIE Social Media Lab
†*Dept. of Electronic and Computer Engineering
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
Email: {eejtchen, eejames}@ust.hk

***Abstract*—Sina Weibo (Weibo) is a fast growing microblogging social network with total user size closer to Twitter. Weibo adopts a mechanism to verify users, so that the public can identify true accounts of celebrities and official channels of certain organizations. The verification mechanism builds trust and authenticity to the source, and hence, stimulates people to actively participate on the platform. However, how the verifications affect the user behaviors in microblogging social networks have never been fully investigated. This paper analyzes the Weibo social network with verifications, by comparing the user microblogging behaviors between verified users and unverified users and studying the social network evovlements of these two group of users. In addition, a method is proposed to approximately reconstruct the network evolvement, and lower bound quasi-densification exponents for the social networks are found. Empirical evidence demonstrates that verifications play a significant role in motivating users to have more interactions in a social network.**

***Index Terms*—Weibo, Verification, Social Network, Network Evolvement**

I. Introduction

Sina Weibo is a Twitter-like microblogging service emerging in China for over two years. Weibo had reached 250 million users in November 2011, where 10% were from outside China mainland. It is now one of the dominating social networks in China and receiving increasing oversea attentions. An growing number of foreign celebrities and politicians like Bill Gates and Mayor Ed Lee start to use Weibo as a channel to reach Chinese audience.

Logging in to Weibo, users can share their status, follow other users' status and get followed by someone else (see Fig. 1 for a snapshot). Any message from the followed user will appear in the *follower*'s *timeline*, where *timeline* displays real time *statuses* from the users that one follows. Unlike social networks such as Facebook and LinkedIn, the user following relationship is not mutually reciprocal. One can follow another without being necessarily granted by the one being followed. On the other hand, when one gets followed, the user does not need to follow back. This property helps build up a relationship beyond acquaintance. The message shared by users in the microblogging, termed *status*, is limited to 140 characters. Meanwhile, hyperlinks, images, videos can attach to the main text as well. To response to a status, a

Figure 1. A snapshot of Sina Weibo

common practice is to *repost* it or *reply* to it. For reposting a status, an '@' sign followed by the author's user name will be added at the beginning of the status. This refers to the source of the reposted status. A few words can be added ahead to comment some ideas. Weibo supports replying to a status while reposting it at the same time. It automatically appends the original reposted status after the new comment and generates a new status. The reposted status gets appeared in timelines immediately, and the users mentioned after the '@' sign will get notified. Such mechanisms speed up the message propagation and interaction throughout the whole social network.

User behaviors and information diffusion in microblogging social network have caught a lot of attentions of researchers recently. Some research have been done on Twitter [1], [2], [3], [4], which is widely considered as the precursor of microblogging social network. However a deep research on Weibo is still missing. The motivation for studying Weibo is due to its profound verification mechanism. Weibo introduces

1545-0678/12 $26.00 © 2012 IEEE
DOI 10.1109/ICDCSW.2012.68

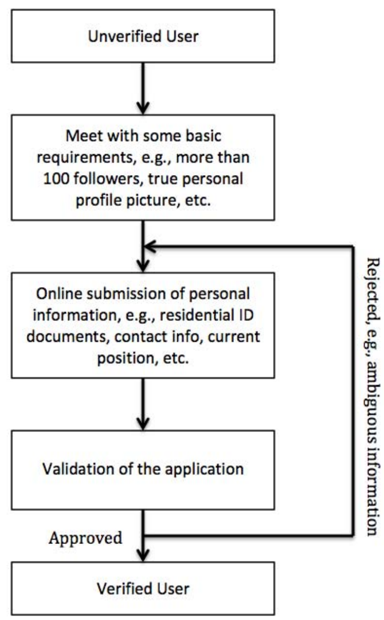

Figure 2. A flow chart for the verification procedure for an individual user

several levels of verifications for individual users as well as organization accounts. Anyone can apply for a verification. At the end of 2011, the total number of verified users has grown to 300,000. On the other hand, Twitter also has its verification. However, it is by closed invitation and only applies to a limited number of users. As a result, the verification mechanism affects a larger portion of users in Weibo than that in Twitter. Thus Weibo,with a broad verification mechanism, raises our research interest.

Weibo opens the verification application to the public. Taking individual users for example, the basic requirements for verification include using a true photo as profile picture, using a true name as user name and having followers more than 100. After fulfilling these requirements, users can go through an online system to submit their applications. The system then asks for some personal information, such as documents of residential identification, contact, position, company name, and a brief user description. Finally, Weibo verifies the information and approves the verification. Fig. 2 shows a flow chart of the verification procedure.

Verifications provide a way for the public to identify trusted source. For celebrities and organizations, they can have an official and reliable way to promote themselves on the Internet, the virtual world. However, little has been investigated on the impacts of such verifications in literature. Does verifications have a significant influence on users' behaviors in online communities? What is the statistical difference between verified users and unverified users? Do the verified users have more influence? The goal of this work is to study the characteristics of different groups of users under the verification mechanism in Weibo.

The research is based on a set of first-hand data crawled from Weibo. It contains 251,988 user profiles, 2,563,079 social relationships, and 535,249 user statuses. Two key contributions in this paper are summarized in the following:

1): Three different types of influence measurements, i.e., indegree influence, repost influence and reply influence, are applied to analyze user characteristics. Verified users are justified to have more influence than unverified users under all the three influence measurements.

2): A method for approximately reconstructing the network evolvement is proposed, which does not rely on the past information of relationships in the social network. A quasi-densification law is proposed to analyze how the sub-networks get denser over time when the whole network contains multiple class of users. The law applies to analyze subgraphs with directional edges as well. The methods provides a lower bound of the quasi-densification exponent as well as a lower bound of the growth rate of the social network. The network evolvement analysis shows that verified users are highly motivated in interacting with others and generating contents.

The paper is organized as follows. Section II explains the process of data collection. Section III introduces the method used in analyzing the verification. In Section IV, empirical results are presented and the comparisons between verified users and unverified users are carried out. Section V proposes some discussions and future works based on the findings from the study. Finally, conclusions are drawn in Section VI.

II. Data Source

A. Data Collection

The social network data was collected from 7th Dec to 13th Dec in 2011, based on the Application Programming Interfaces (APIs) provided by Weibo. As a summary, 251,988 user profiles, 2,563,079 followers and following relationships had been crawled. In addition, 535,249 statuses from 2,736 randomly selected users were included in the dataset. Within the dataset, there are 210,579 unverified users and 41,409 verified users. The data collection process started from a randomly selected user, obtaining the user's profile and the user's follower and following trees. Then it repeatedly did the data collection from each of the followers and followings, obtaining one's user profile as well as one's follower and following trees. The same was done whenever a new user came to the dataset. If the network was fully connected without isolated nodes, the collection process could reach every user whenever it ran for a sufficiently long time (assuming the network did not change during the data collection).

Note that the analyses shall be carried out by comparing the groups of verified users and unverified users, relying on only the statistical behaviors of individual users. Therefore, there is no need to crawled all the data from the Weibo network.

B. Dataset Refinement

To make a precise assessment of the network, spam users are removed from the dataset. In general, two types of spam users

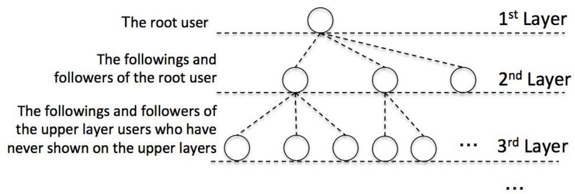

Figure 3. The tree representation of the data collection process.

had been removed, *inactive users* and *robots*. *Inactive users* referred to users who had posted less than 10 posts since they registered. For those users, they might be new to the platform or had already left the platform. Their data was meaningless to our network statistics analysis. *Robot* users referred to users who had followed more than 2,000 users but had less than 10 following back. A robot user was usually maintained by advertisers, who tried to seek following back in order to boost up the size of audience. As this paper aims to compare between verified users and unverified users, such behavior is out of the scope and creates noise to the results. After filtering the outliers, there were 167,499 users that remained, where 127,086 were unverified users and 40,413 were verified users.

C. Discussion on the Date Collection Process

As only a portion of data was collected from Weibo, a proper data collection process is important. Fig. 3 illustrates a tree representation of the data collection process. Specifically, the first layer is the root user we started with. The second layer is the followers and following of the root user. The deeper layers contain users who are followers and followings from the upper layer users but who have never appeared on all the upper layers. Note that the deepest layer corresponds to the largest geodesic distance that exists in the subgraph, and there is no connection between the root user and the users on the bottom layer.

There are two reasons for implementing such data collection process. First of all, it makes full use of the efficacy of Weibo APIs while no white listed IPs are available. Second of all, it automatically forms a subgraph that has dense interconnection among users. Such dense interconnected network makes the network evolvement analysis and the quasi-densification exponent estimation reliable (see Section IV-C and IV-C3).

Moreover, this method is insensitive to the selection of root user. Note that as the layer goes deeper, the users there have less correlation to the root user, as there is no connection from the users at the bottom layer back to the root user. While the lower layers dominates the user sizes, the majority users in the dataset share no direct connection with the root users. Thus the root user would not introduce significant bias to the dataset, and, as the dataset goes large, the impact of the choice of root user vanishes. A further justification is also discussed in Section V, showing that the process managed to reach a large variety of users randomly.

III. Methodology for Analyzing the Impacts of Verifications

Verifications separate users into two major groups, verified users and unverified users. The impacts of verifications are shown through the different characteristics between the two groups of users. In this paper, three categories of measurements are applied for comparisons. The activity intensity shows the overall distributions of intensities of user activities in microblogging. The influence analysis studies an individual user's capability in motivating others. Finally, the network evolvement analysis shows how these two groups of users grows over time.

A. Activity Intensity of Microblogging

Three parameters from user profile are directly used to characterize the user activity intensity, namely, the number of followers, the number of followings and the number of statuses. Empirical complementary cumulative distribution functions (CCDF) are used to show the distributions of the numbers of followers, followings and statuses, respectively. The role of CCDF is to show an overall chance for a user to have followers (following/statuses) more than n, i.e., $P_i(n) = \frac{1}{N}\sum_{i=1}^{N} 1(F_i > n)$, where $1(\cdot)$ is the indicator function, F_i is the number of followers (followings/statuses) for user i and N is the total number of users.

B. Influence Measurement

Influence reflects how one user interacts with others in the social network and unveils the capability for one to motivate another. Different influence measurements view the user capabilities and user relationships from different aspects. In the following, three different influence measurements applied in the rest of the paper are summarized.

- **Indegree influence:** It is defined as the number user's followers [1], [5]. A user with a larger number of followers should have a greater influence power, since the message that the user broadcasts has a larger direct audience.
- **Repost influence:** It is defined as the ratio of the total aggregated number of reposts over the total number of user's statuses. For example, if a user have 10 statuses, among which only one status has replies and the number of replies is 100. Then the repost ratio is $100/10 = 10$. For a microblogging social network, the user connections are sometimes built on common interest. A user with insightful thoughts can attract a large number of audience who are strangers. The repost influence indicates one's capability in leading a discussion topic.
- **Reply influence:** It is defined as the ratio of the total aggregated number of replies over the total number of user's statuses. The reply behavior is slightly different from repost, since the replied message does not go to others' timeline and hence it sounds more private to the owner of the microblog. While the ones who repost one's statuses might be strangers, the ones who reply to the blogger are most likely to be acquaintance. Thus the reply

Verified Users			Unverified Users		
User ID	Followers	Description	User ID	Followers	Description
微博小秘书	20, 106, 712	Weibo customer service	冷笑话精选	7, 344, 640	A broadcast channel for selective jokes
姚晨	15, 215, 866	Movie star	veggieg	6, 007, 471	Singer
小S	14, 350, 501	TV star	微博搞笑排行榜	5, 734, 078	A broadcast channel for hot jokes
谢娜	12, 533, 831	TV star	吹神	4, 426, 986	Singer
蔡康永	12, 231, 408	TV star	精彩语录	4, 150, 809	Collections of famous quotations
杨幂	12, 136, 668	Movie star	全球热门排行榜	4, 140, 154	A collection of hot gossips
赵薇	11, 890, 813	Movie star	星座秘语	3, 788, 400	Descover the astrology
何炅	11, 406, 089	TV star	星座爱情001	3, 752, 686	Astrology and love
王力宏	11, 024, 485	Singer	全球时尚	3, 722, 425	Global fashion
微博客服	10, 730, 411	Weibo customer service	时尚经典语录	3, 713, 212	Classical and fashion quotations

Table I
SUMMARY OF INFORMATION FOR TOP-10 INFLUENTIALS

influence reflects the social acquaintance and closeness between the user and the user's friends (or fans).

Note that the three influence measurements represent different characteristics of one's involvement in a social network, reconstructing a full angle of one's influence capability.

C. Reconstruction of Network Evolvement

The network evolvement processes tell some characteristics of the user groups, such as entrance barriers. However, it is difficult to reconstruct the processes on how one user or even a whole network evolve with more and more followers and followings, since the time for building a connection between two users is not recorded. In the following, two methods are proposed to overcome this challenge.

1) Estimate the evolvement for the whole network: Our procedure starts from the current data and calculates the total number of followings and statuses on the whole network. It then traces back to see which users should disappear (not yet registered) at some certain time instances ahead. It proceeds with subtracting the number of the corresponding users' followers, followings and statuses from the current total values. For example, a user *A* joined the network two years ago. To estimate the number of followers of user *A* after 100 days from then, the number of followers of user *B*, who has recently joined for 100 days, is used. Using the current data of users who joined in different time reconstructs an estimation on how the number of followers, followings and statuses grow for user *A*. According to this method, an upper bound values of followers, followings and statuses at any past time are constructed. Therefore an estimated network evolvement process is reconstructed even when the relationship history of the social network is unavailable.

Note that the real network actually grows faster then the estimated one, as upper values are used as a guess of the true value in the past. As a result, the estimation gives a lower bound of the growth rate. It is worth-wise to point out that, although an estimation is made, a fair comparison between verified users and unverified users is still available, since the same assumptions are applied to both sides.

2) Estimate the Densification Exponent: Finding the network densification exponent [6] involves finding the ratios between the number of users and the number of followers and followings in a series of time periods. As we do not have the past information of the network, the following procedure is used to reconstruct an approximated the densification process: *a)* Starting from the present time T, the total number of users $N(T)$ and total number of followers and followings $E(T)$ are found; *b)* For a previous time $T - \triangle t$, find $N(T - \triangle t)$ by subtracting from $N(t)$ the users who had not joined the network at that time, and $E(T - \triangle t)$ by subtracting from $F(T)$ all those users' follower and following relationships; *c)* Repeat *b* for determining all $N(T - i\triangle t)$ and $E(T - i\triangle t)$ for $i = 1, \ldots, \left\lfloor \frac{T}{\triangle t} \right\rfloor$.

Note that, as the estimated $E(t)$ is always large than the actually number of followers and followings, it results in estimating a lower bound densification exponent (see Fig. 9 as an illustration).

IV. EMPIRICAL STUDIES OF THE IMPACTS OF VERIFICATIONS IN WEIBO

In this section, empirical studies are presented based on the methods we discussed in Section III and utilizing the data we crawled from Weibo. Comparisons between verified users and unverified users are shown in terms of user activity intensity, influence and network evolvement.

A. Activity Intensity Studies

In the following, the CCDFs of of the number of followers, followings and statuses for verified users and unverified users are shown and compared.

1) Followers: Fig. 4a shows the CCDFs of the numbers of followers. Over a half of verified users have more than or close to 10,000 followers. However, the majority of unverified users (with probability larger than 0.8 from the figure) only have less than 1,000 followers. The two curves differ quite a lot, and in most of the probability region, the followers of verified users are even 10 times more than that of unverified users. Furthermore, the curve for unverified users seems to

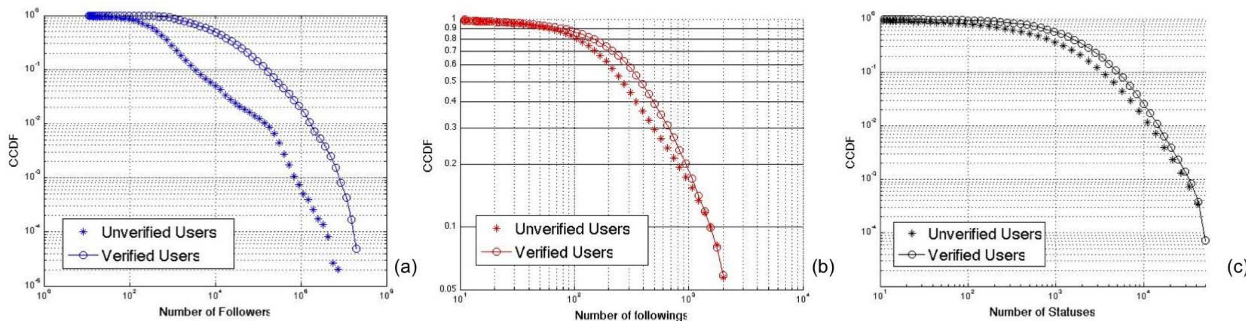

Figure 4. The CCDFs of the number of followers, followings and statuses for verified users and unverified users.

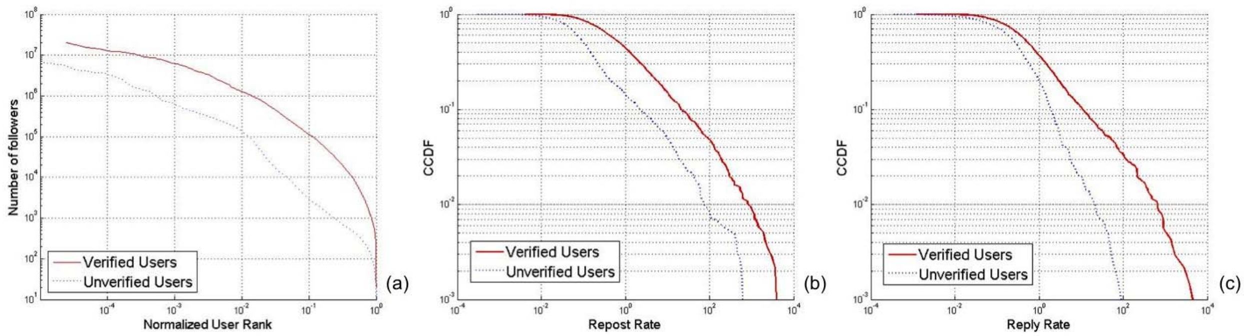

Figure 5. The number of followers versus the normalized user ranks.

be less 'regular' than that of the verified users. One possible suspicion is that, the statistics of verified users, although have a smaller size of dataset, is less disturbed by spam users.

2) Followings: Fig. 4b shows the CCDFs of the numbers of followings. Unlike the curves of followers, the CCDF of followings for verified users and unverified users are quite identical. One possible reason is that, while one could be followed by a limited number of followers, the number of followings is still limited by one's energy on absorbing information.

3) Statuses: Fig. 4c shows the CCDF of the number of user statuses for verified users and unverified users. Similarly, the two curves do not differ too much. It shows that, although the verified users have on average more audience, the user verification system does not motivate people to speak more in general. On the other hand, a trivial conclusion is that the reason for a user being followed is not that he/she speaks more.

In summary, verified users and unverified users have similar activity intensities in microblogging in terms of following people and generating statuses. However, verified users tend to attract much more followers than unverified users on average.

B. Influence Analysis

In this section, indegree influence, repost influence and reply influence are studied by investigating user's statuses.

1) Indegree Influence: A PageRank-like demonstration is applied to show the distribution of users' followers. Users are ranked according to their indegree influence in descent order, and the ranks are linearly normalized to $[0, 1]$ with 0 indicating the highest rank and 1 indicating the lowest.

Fig. 5a shows the users' indegree influence versus their ranks. Verified users have higher indegree influence than unverified users.

2) Repost Influence and Reply Influence: Fig. 5b and Fig. 5c show the empirical CCDFs on repost rate and reply rate for both verified users and unverified users. The figures show that for the majority of users, the repost rate and reply rate are less than 1, while, verified users have repost rate and reply rate 10 times larger than that of unverified users, in most of the probability region. According to the discussion in Section III-B, statistically, verified users are shown to have a higher capability in leading discussion topics and a larger number of friends (or fans).

As is justified by the empirical results, verified users are more influential under all the three influence measurements. They usually attract more followers and their statuses catch much more public attention. As a result, verified users have

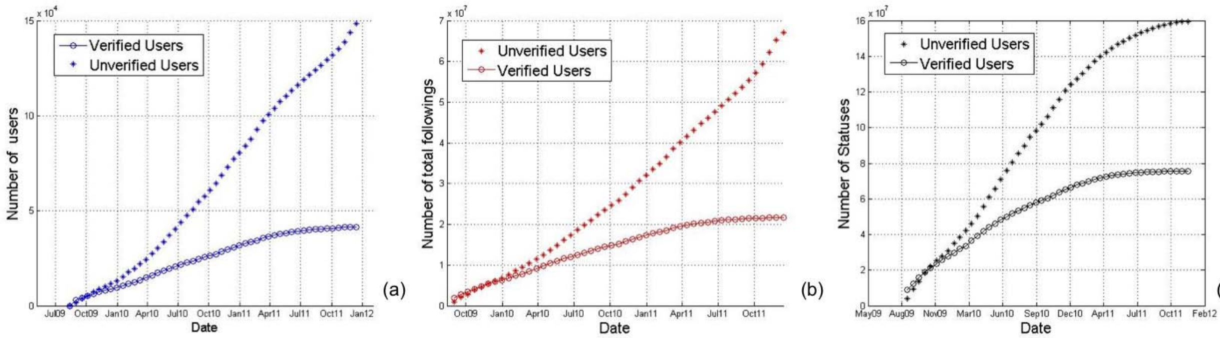

Figure 6. Growth of the total number of users, followings and statuses of a subgraph in Weibo.

the capability to initiate intensive interactions among users.

3) The Top Influentials: The information of top 10 influential people based on indegree influence is summarized in Table I. For verified users, 8 IDs are from individual users and they are all celebrities, while the other two IDs are official customer service channels. For unverified users, only 2 IDs are from individual users. It is believed that they are the famous Chinese singers Faye Wong and Eason Chan, respectively. However, although their influence in China is no less than those celebrities on the left column, their followers are much less than those with verifications. On the other hand, the 8 other unverified IDs are broadcast channels for specialized topics. Those IDs never share personal statuses, but publish selective hot messages in their specialized areas. It is obvious that people follow those accounts only because of their useful and interest content.

Table I provides some evidence to the power of verifications. For celebrities, verifications gain a lot more of influence for them. Without influence, social media channels with interesting content attract more followers than individual users.

C. Network Evolvement Analysis

Network evolvement analysis compares the networks of two groups of users over time. In the following, the network-wide growth and individual growth of social network are investigated. Densification law introduced by [6] is also studied under verified users and unverified users.

1) Growth of Users, Followings and Statuses in a Macroscopic View : The date that a user sign up to the platform is recorded on user's profile. In our dataset, the first user registered on 14 August 2009 and the latest one registered on 13 December 2011. Fig. 6 shows how the user size, the total followings and the total statuses grow within this period. The number of unverified users first grew slowly and then grew in an increasing rate. On the other hand, the number of verified users first grew quite fast, but then slowed down in recent months.

It needs to emphasize that, the actual growth rate of followers, followings and statuses can be arbitrarily larger than which is reported in Fig. 6. However, as a comparison between the two group of users, the estimated growing processes with lower bound growth rate are still useful.

2) Growth of Followers, Followings and Statuses in an Individual Point of View: Fig. 7 show the distributions of numbers of followers, followings and statuses for different users who had joined Weibo for different period of time (up to 14 Dec 2011). We refer to the time period that a user had joined Weibo (up to 14 Dec 2011) as the user's *Weibo age*.

In Fig. 7a, the distribution of numbers of of followers for unverified users grew almost linearly according users' Weibo age within the first year after joining Weibo. For unverified users with Weibo age greater than one year, the number of followers became saturated and remained below 1,000 ultimately. On the other hand, for verified users, the number of followers grew quite fast and kept growing. Verified users with older Weibo ages usually show a larger number of followers. The average number of followers reaches 100,000 for second year verified users.

Fig. 7b shows the distribution of numbers of followings for verified users and unverified users according to their Weibo ages. For verified users, the number of followings kept growing according to their Weibo ages within the first year, but got saturated in around 500, starting from the second year.

An interesting phenomenon happens to unverified users. The number of followings increased at the first 100 days and then continually dropped according to Weibo ages. This phenomenon suggests that the newly joined users (users who joined in the recent 100 days before 14 Dec 2011) are much more passionate to follow others than the users who joined Weibo a long time ago. In addition, this serves as an evidence that user behaviors change a lot of time and the behaviors between fresh users and users with an older Weibo age may diverse significantly.

In Fig. 7c, the distribution of number of statuses from verified users grows faster than that of unverified users according to Weibo ages. The verified users behave statistically more active in the social network.

3) Estimate the Densification Exponent: Densification law was introduced in [6] to predict the relationship between the dynamics of the number of nodes and the total degrees of

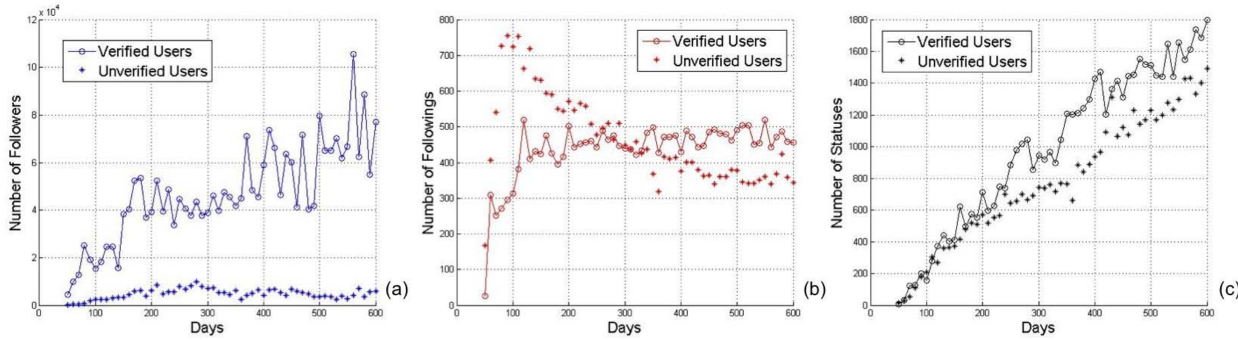

Figure 7. Distributions of numbers of followers, followings and statuses for different users according to the period of time the users had joined Weibo (up to 14 Dec 2011).

the network. It predicts that the network evolves following the densification law as $E(t) = cN(t)^{\alpha}$, where $E(t)$ is the total number of edges (user relationships, e.g., followers and followings), $N(t)$ is the total number of nodes (users), c is some constant and α is the densification exponent.

However, it is always difficult to obtain the whole social graph of an online social network, as the graph is usually huge in dimension and fast changing in topology. Also, it is difficult to use a subgraph to estimate the densification property of the whole network. The reason is that when a subgraph is cut from a whole network, some interconnections (e.g. the links from inside the subgraph to outside the subgraph) are pruned, significantly affecting the ratio between the number of nodes and the number of edges. Fig. 8 illustrates this phenomenon. Cutting node 1 affects 4 edges, while cutting node 2 affects only 1 edge. Hence the nodes over edges ratio in the subgraph may diverges from that of the entire network.

To tackle this problem, denote $E_i(t)$ as the number of edges connecting nodes within the subgraph (solid lines) and $E_o(t)$ as the number edges connecting nodes from inside the subgraph to outside the subgraph (dash lines). A *quasi-densification law* is proposed as follows,

$$\overline{E}(t) = c\overline{N}(t)^{\overline{\alpha}}$$

where $\overline{E}(t) = E_i(t) + 0.5E_o(t)$ is the effective number of edges, $\overline{N}(t)$ is the total number of nodes in the subgraph and $\overline{\alpha}$ is the associated *quasi-densification exponent*.

The coefficient 0.5 in the expression of $\overline{E}(t)$ is chosen as follows. In a complete graph, every two nodes, if they are connected, share one edge. However, an outgoing edge (dash line) connects with only one node in the subgraph, and hence it should be half significant over an internal edge (solid line). An other interpretation is that, consider all the edges are with directions. An outgoing edge (dash line) corresponds to a following relationship, an incoming edge (dash line) corresponds to a being followed relationship, and a bi-directional edge (solid line) corresponds to a mutually following relationship. The nodes over effective edges ratio measures the relationship between the number of users and the number of following and followed back behaviors among users.

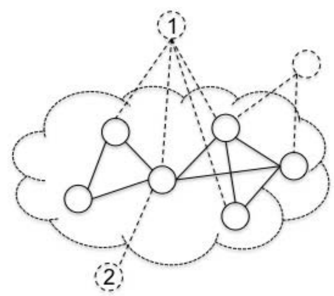

Figure 8. A subgraph (inside the cloud) in the entire network.

Note that, when the subgraph becomes a complete graph, then $E_o(t) \equiv 0$, and the effective number of edges $\overline{E}(t)$ becomes the true number of total edges. Moreover, the quasi-densification law deals more general network with directional edges, while the densification law in [6] only takes care of non-directional edges. In addition, the quasi-densification law enables an densification analysis of subgraphs split from a complete network.

In this section, subgraphs of verified users and unverified users are formed. Quasi-densification exponent are estimated for these two subgraphs. However, estimating the exponent $\overline{\alpha}$ needs to know the past information of relationships in the network. In the following, the past total number of followers and followings are estimated using the techniques discussed in Section III-C. The estimation gives a lower bound of the quasi-densification exponent $\overline{\alpha}$.

Fig. 9 shows the power-law characteristics of the growing number of users and relationships (sum of the numbers of followers and followings). For the subgraph of verified users, the quasi-densification exponent $\overline{\alpha}_v = 0.8745$, while for that of unverified users, $\overline{\alpha}_u = 1.1312$. The unverified users' network grows more aggressively. This result is consistent with Fig. 7 and the corresponding discussion in Section IV-C2, where the newly joined unverified users are shown to be much

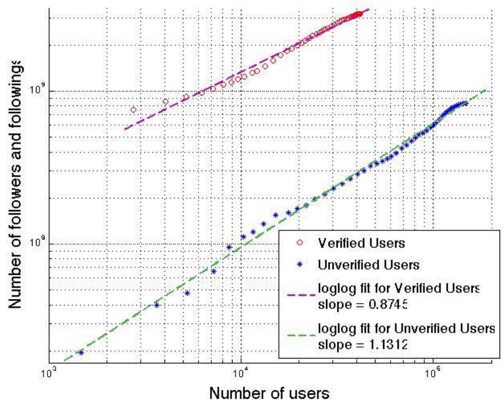

Figure 9. The network evolvement

more passionate to follow others.

V. Discussions and Future Works

The above empirical results have compared verified users and unverified users over several characteristics from Weibo microblogging social network. The group of verified users have denser connections than unverified users, and hence, they tend to be more influential. Throughout this study, there are two interesting findings:

1): Through the data collection process, the users joined in 14 Aug 2009 and 13 Dec 2011 are both collected in our dataset. Interestingly, the first date is quite close to the date that Weibo started to provide service (Aug 2009), while the latter one is exactly the date we stopped crawling data. This result validates that our data collection process was able to reach the whole network randomly.

2): The network evolvement analysis is an approximation based on the assumption that the characteristics of user behaviors does not change over time. Hence it is interesting to compare the estimated evolvement processes with the true ones and evaluate the gap between them, which leads to the future work of a further analysis of verifications.

VI. Conclusion

In this paper, Weibo, a Twitter-like microblogging social network with verifications was investigated. Based on the user data crawled from Weibo, the characteristics of user behaviors were analyzed. Activity intensity analysis, influence measurements and network evolvement analyses were carried out to compare the statistical difference between verified users and unverified users. Empirical results showed that the verifications stimulated people to follow verified users, get deep involved in social interactions and actively participate in the online social activities.

VII. Acknowledgement

We thank Shougang Shi from Huawei-HKUST lab, Hong Kong, for his help in implementing the APIs for data collection.

References

[1] A. Leavitt, E. Burchard, D. Fisher, and S. Gilbert, "The Influentials: New Approaches for Analyzing Influence on Twitter," *Webecology Project*, Sep. 2009.

[2] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proceedings of the third ACM international conference on Web search and data mining*, ser. WSDM '10. New York, NY, USA: ACM, 2010, pp. 261–270. [Online]. Available: http://doi.acm.org/10.1145/1718487.1718520

[3] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *WWW '10: Proceedings of the 19th international conference on World wide web*. New York, NY, USA: ACM, 2010, pp. 591–600.

[4] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring User Influence in Twitter: The Million Follower Fallacy," in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*. [Online]. Available: http://an.kaist.ac.kr/˜mycha/docs/icwsm2010_cha.pdf

[5] S. Ye and S. F. Wu, "Measuring message propagation and social influence on twitter.com," in *Proceedings of the Second international conference on Social informatics*, ser. SocInfo'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 216–231. [Online]. Available: http://dl.acm.org/citation.cfm?id=1929326.1929342

[6] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, ser. KDD '05. New York, NY, USA: ACM, 2005, pp. 177–187. [Online]. Available: http://doi.acm.org/10.1145/1081870.1081893