

## Research Article

# A Comparison of Online Social Networks and Real-Life Social Networks: A Study of Sina Microblogging

**Dayong Zhang and Guang Guo**

*Department of New Media and Arts, Harbin Institute of Technology, Harbin 150001, China*

Correspondence should be addressed to Dayong Zhang; [zhdy@hit.edu.cn](mailto:zhdy@hit.edu.cn)

Received 10 November 2013; Revised 17 February 2014; Accepted 21 March 2014; Published 22 April 2014

Academic Editor: Guoqiang Hu

Copyright © 2014 D. Zhang and G. Guo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Online social networks appear to enrich our social life, which raises the question whether they remove cognitive constraints on human communication and improve human social capabilities. In this paper, we analyze the users' following and followed relationships based on the data of Sina Microblogging and reveal several structural properties of Sina Microblogging. Compared with real-life social networks, our results confirm some similar features. However, Sina Microblogging also shows its own specialties, such as hierarchical structure and degree disassortativity, which all mark a deviation from real-life social networks. The low cost of the online network forms a broader perspective, and the one-way link relationships make it easy to spread information, but the online social network does not make too much difference in the creation of strong interpersonal relationships. Finally, we describe the mechanisms for the formation of these characteristics and discuss the implications of these structural properties for the real-life social networks.

## 1. Introduction

In the past decade, the online social network has made new opportunities for communication and has revolutionized many aspects of our lives. As a new kind of online social networks, the Microblogging, such as Twitter and Sina Microblogging, has been developing rapidly in recent years. The online social network provides people with a public network platform, meets needs of computer-mediated communications, and rebuilds social connections. The structure and evolution of online social network have attracted the attention of researchers in different disciplines, including sociology, physics, and computer information technology [1, 2]. But large-scale and complex structure of online social network brings many difficulties in complete topological analysis. However, the emergence of complex network research [3–5] provides us with an effective method to study the online social network topology and information dissemination characteristics [6–11].

As for the mapping and expansion of the actual social networks on the Internet, online social networks have broken the spatiotemporal limitations and decreased communication

cost. Will the online social networks change the rules of actual social networks? The following studies reach fundamentally different conclusions: Backstrom and Boldi [12] studied the largest online social network that has ever been created (721 million active Facebook users and their 69 billion friendship links). The analyses of data proved that the average distance of Facebook has shortened from 5.28 in 2008 to 4.74 in 2011. When the search scale was narrowed to a country, it showed that most of the people were in fact only four apart. The result presented the “shrinking diameter” phenomena [13]. It suggested that the online social network made the relationship among people come closer. Six degrees of separation is inapplicable to the online social network. However, the statistics of Facebook's official website showed that the average number of Facebook active users' friends was 130 (<http://www.facebook.com/press/info.php?statistics>), which agrees well with Dunbar's number. Similarly, Gonçalves et al. [14] analyzed a dataset of Twitter conversations collected across six months involving 1.7 million individuals and found that users kept stable relationships with 100–200 users. The data is in agreement with Dunbar's result too. Thus, the

TABLE I: Summary of degree distribution.

$k$	Number of nodes	Proportion	<i>Extremum</i>
$1 \leq k < 10$	783	93.3%	$k_{\min} = 1$
$10 \leq k < 100$	50	6.00%	
$k \geq 100$	6	0.07%	$k_{\max} = 327$

“economy of attention” is limited in the online world by cognitive and biological constraints as predicted by Dunbar’s theory [15].

Has online social network changed our real-life social pattern? Why do such discrepancies exist between the studies? The differences are due to the different analytical content and analysis approach. The low cost of the online network forms a broader perspective, and the one-way link relationships make it easy to spread information, but the online social networks do not make too much difference in the creation of strong interpersonal relationships. These online social relationships, which may reflect the real-life social networks, create an unprecedented field to understand the characteristics of human networks [16, 17].

In this paper, we study the topological characteristics of Sina Microblogging and try to explain its characteristics formation mechanism by comparing it with the real-life social networks. This paper is organized as follows. Section 2 describes our data collection on Sina Microblogging. Then, we conduct topological analysis of the Sina Microblogging network and analyze the mechanisms for the formation of the structure in Section 3. In Section 4, we study the mixing patterns on the Sina Microblogging network. In Section 5, we focus on the analysis of node importance by betweenness centrality. In Section 6, we conclude.

## 2. Sina Microblogging Data Collection

In August 2009, Sina Microblogging began its trial service and became the most popular Microblogging service in China, with more than 500 million users as of 2013. Sina blog and Sina news provide good capital bases and natural advantages for the success of Sina Microblogging. Due to these advantages, Sina Microblogging developed the characteristics of We Media that spread news faster than any other media. In order to study the structure of Sina Microblogging, we develop a spider with Python for Sina Microblogging. The snowball [18] crawl algorithm has been applied in the spider. We collected profiles of 3441 users on Sina Microblogging until November 11, 2011. Isolated points are excluded, so that filter leaves only 839 nodes. In order to protect privacy, we only keep the users’ ID number in the original data information about the users. Based on the “following” and “followed,” we construct an undirected network which has 2112 links and analyze its basic characteristics.

## 3. Analysis of Network Structure

We begin our analysis of Sina Microblogging from the following aspects. In Section 3.1, first, we calculate the degree

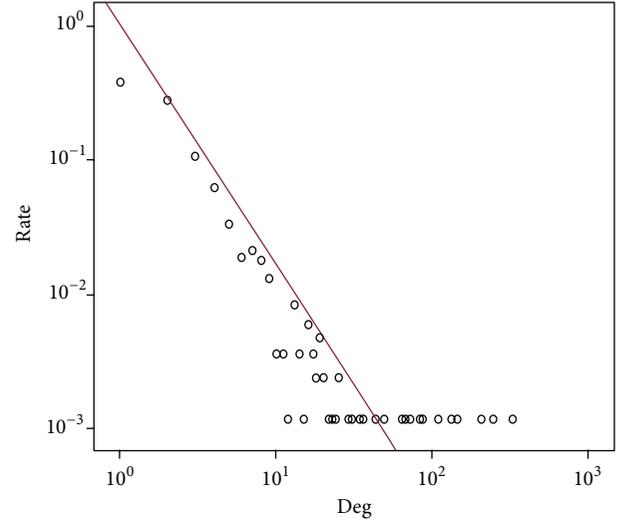


FIGURE 1: Degree distribution for nodes in Sina Microblogging.

distribution. Then, we evaluate the average path length and clustering coefficient and study the reason why Sina Microblogging possesses obviously small-world property in Section 3.2.

**3.1. Power-Law Node Degrees.** In Table 1, we summarize the degree distribution of Sina Microblogging. We find that only a few nodes of the network have a high degree. The degree distribution is very unevenly distributed.

We begin the analysis of Sina Microblogging topology by looking at its degree distributions. Networks of a power-law degree distribution,  $(k) \propto k^{-\gamma}$ , where  $k$  is the node degree and  $\gamma \leq 3$ , attest to the existence of a relatively small number of nodes with a very large number of links. Many social networks satisfy power-law degree distribution [19, 20, 23], and a few networks obey stretched exponential distribution [24] which is defined as  $p(k) = be^{-(k/k_0)^\epsilon}$ .

The degree distribution of Sina Microblogging is shown as in Figure 1. The  $y$ -axis represents the frequency of degree. It satisfies a power-law distribution with an exponent of 1.75; the goodness of fit is 0.712.

The networks which meet a power-law distribution have scale-free property, and such networks are called scale-free networks. Therefore, Sina Microblogging is a scale-free network. The real-life social networks are usually scale-free networks. The scale-free network is caused by the growing and preferential mechanisms that the new nodes tend to connect with hub nodes, and this phenomenon is called the Matthew effect.

**3.2. The Small-World Property.** The study of “small-world” networks has become a key to understanding the societal structure, ever since Stanley Milgram’s famous “six degrees of separation” experiment. In his work, he reports that any two people could be connected on average within six hops from each other. Watts and Strogatz [3] have revealed two important characteristics of “small-world” networks:

TABLE 2: Small-World property compared with different social networks.

Social network	Average path length	Clustering coefficient
Sina Microblogging	2.794	0.374
Random network	4.162	0.006
Cyworld [19]	3.2	0.16
Renren [20]	3.48	0.2
Mixi [21]	5.53	0.1215
Co-author network [22]	4.6	0.066

(1) the “small-world” networks have small characteristic path lengths, like random graphs; (2) the “small-world” networks can be highly clustered, like regular lattices. The network measurement studies have shown that the real networks are mostly small-world, especially social networks [13, 21].

The concept of average path length for undirected networks is well known; the measures are related by

$$L = \frac{1}{(1/2)N(N+1)} \sum_{i \geq j} d_{ij}, \quad (1)$$

where  $i, j$  are two nodes,  $N$  is the number of nodes in the network, and  $d_{ij}$  denote the shortest distance between  $i$  and  $j$ .

The average path length is closely connected to the characteristics of the network, such as connectivity, reachability, and transferring latency. A short average path length facilitates the quick transfer of information and reduces costs.

Another property of “small-world” networks is the clustering coefficient. It is the measure of the extent to which one’s friends are also friends of each other. The local clustering coefficient for a node is then given by the proportion of links between the nodes within its neighborhood divided by the number of links that could possibly exist between them. The local clustering coefficient for undirected graphs can be defined as

$$C_i = \frac{2E_i}{k_i(k_i - 1)}, \quad (2)$$

where  $k_i$  is the degree of node  $i$  and  $E_i$  is the number of edges between neighbors of node  $i$ .

The clustering coefficient for the whole network is given by Watts and Strogatz [3] as the average of the local clustering coefficients of all the nodes; the measures are related by

$$C = \frac{1}{N} \sum_i C_i. \quad (3)$$

The networks with the largest possible average clustering coefficient are found to have a modular structure.

The results concerning average path length and clustering coefficient are displayed in Table 2. Compared with the same scale random network, the average path length of Sina Microblogging is shorter than random network (the average path length of random network is defined as  $L \propto$

$\ln N / \ln k$ ), and the clustering coefficient is much greater than the random network (the clustering coefficient of random network is defined as  $\langle k \rangle / N$ ). Thus, we confirm that Sina Microblogging has the small-world property.

Renren, Cyworld, and Mixi are the largest and oldest online social networking services in China, South Korea, and Japan, respectively. Coauthor network is a kind of real-life social network. By contrast, the average path length of Sina Microblogging has a shorter average path length and a greater clustering coefficient. It is revealed that Sina Microblogging’s small-world property is the most obvious. The striking small-world phenomenon indicates that there are rich local connections in Sina Microblogging network, the nodes are linked closely, and information dissemination is more efficient.

We analyze the mechanisms for the formation of small-world property from the following two aspects. First is the user’s behavior aspect. In 2011, for many users, the main reasons they use social network out there are the latest developments of friends (73.4%), keeping in touch with old friends (73.1%), and documentation of life and feeling (67.5%), and the main purposes of Microblogging users are getting information (58.1%), following celebrities (57.6%), discussing the hot topics, and personal experience (52.3%), according to the data of “the user behavior research of SNS and microblogging in China.” The Microblogging users tend to use the Microblogging to record personal feelings, share the news, and find groups with similar interests, and the traditional SNS users show themselves, contact with friends in real life, and expand social circle on SNS. Consequently, the former are oriented to information exchange, but the latter stress interpersonal communication.

Second is the aspect of users’ link mode. The main difference between the Renren, Cyworld, and Mixi networks and Microblogging is the directed nature of Microblogging relationship. In Renren, Cyworld, and Mixi, a link represents a mutual agreement of a relationship, while on Sina Microblogging a user is not obligated to reciprocate followers by following them. Thus, a path from a user to another may follow different hops or not exist in the reverse direction. Both the internal links (strong ties) and one-way following links (weak ties) are in Microblogging. And the Microblogging’s distinction between different types of interactions allows us to get more information: personal interactions are more likely to occur on internal links (strong ties) and events transmitting new information rely more on one-way following links (weak ties) [25].

In conclusion, Microblogging is easier to form the user centric We Media than traditional SNS; therefore, it possesses small-world property that can facilitate the flow of information.

In order to gauge the correlation between clustering coefficient and the degree, we plot the clustering coefficient of nodes ( $y$ ) against the degree of nodes ( $x$ ) in Figure 2 and bin the clustering coefficient in log scale. If the correlation between clustering coefficient and the degree accords with  $C(k) = k^{-a}$ , we can consider that the network is of apparent hierarchical structure. We observe that the clustering coefficient varies inversely as the degree. It satisfies a power-law

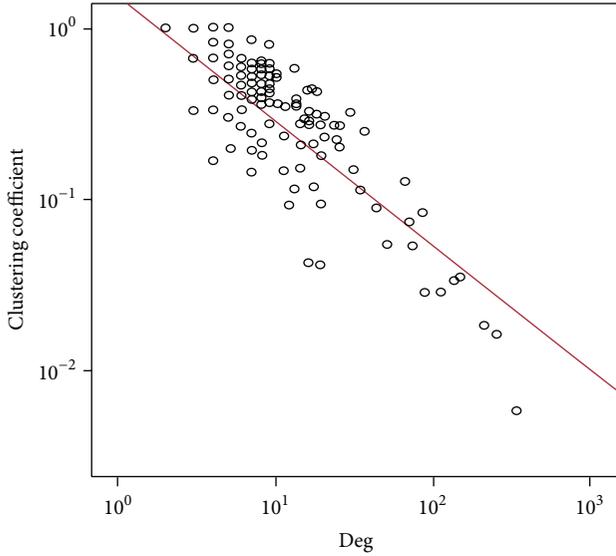


FIGURE 2: Correlations between clustering coefficients and degrees.

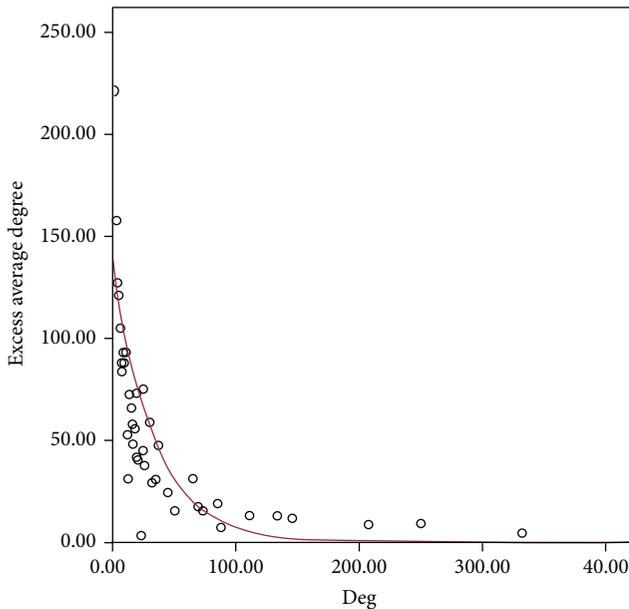


FIGURE 3: Correlations between excess average degrees and degrees.

distribution with an exponent of 0.733, and the goodness of fit is 0.712. Thus, the Sina Microblogging has apparent hierarchical structure, where vertices divide into groups that further subdivide into groups of groups and so forth over multiple scales.

The structure is caused by Sina Microblogging's fans mechanism. Basing on the user interest and the celebrity effect, Sina Microblogging builds many kinds of fans groups. Celebrities and professionals in a certain area who have more resources and influence will get more attention, and these users only keep internal links with peers or slightly lower level users and so forth until the hierarchical structure is formed.

## 4. The Mixing Patterns

Mixing patterns refer to systematic tendencies of one type of nodes in a network to connect to another type. There are three types of mixing patterns: assortative network, disassortative network, and neutral network. Similar vertices tend to connect to each other in assortativity network, and nodes of low degree are more likely to connect with nodes of high degree in disassortative network. A lot of empirical studies [26, 27] had revealed that the real-life social networks trend to assortativity, opposite of the online social networks. In real-life social networks, the ordinary people want to get along with the celebrity, while the celebrity tends to make the acquaintance of peers; therefore, the ordinary people have less opportunity to integrate into the circles of the celebrity. In contrast, the ordinary people can easily get connected with the celebrity; the celebrities are also willing to show their influence by the number of fans in the online social network. Thus, the online social networks trend to be disassortative networks.

We cannot understand the mixing pattern intuitively, so we introduce the concept of excess average degree. The excess degree is the number of edges leaving the vertex other than the one we arrived along. This number is one less than the degrees of the vertices themselves. The excess average degree is defined as

$$\langle k_{nn} \rangle (k) = \frac{1}{q_k} \sum_{k'=k_{\min}}^{k_{\max}} k' e_{kk'}. \quad (4)$$

The excess average degree is positive in assortative network and negative otherwise. Figure 3 plots the curve of the excess average degree ( $y$ ) against the degree ( $x$ ) as the red line. We see significant negative correlation. Hence, the Sina Microblogging is a disassortative network.

To better understand the meaning of disassortative network, we illustrate the network connections of the node ID 975 in Figure 4. In connected component, different colors represent different communities. We find that the node ID 975 connects with three hubs of community and shows disassortativity.

## 5. Analysis of Node Importance

Social networks are discrete systems with a large amount of heterogeneity among nodes. Measures of centrality direct at a quantification of nodes' importance for structure and function. The most direct measure of centrality is the degree centrality; that is, the node with greater degree is the most important one. In addition, betweenness centrality is also a measure of node's centrality in a network and can be used to measure the influence a node has over the spread of information through the network. It is equal to the number of the shortest paths from all vertices to all others that pass through that node.

In Figure 5, we plot the correlation between node betweenness centralities and degrees. Then, we compute that the average of node betweenness centralities is 33.62 and the correlation coefficient is 0.96, indicating that there are

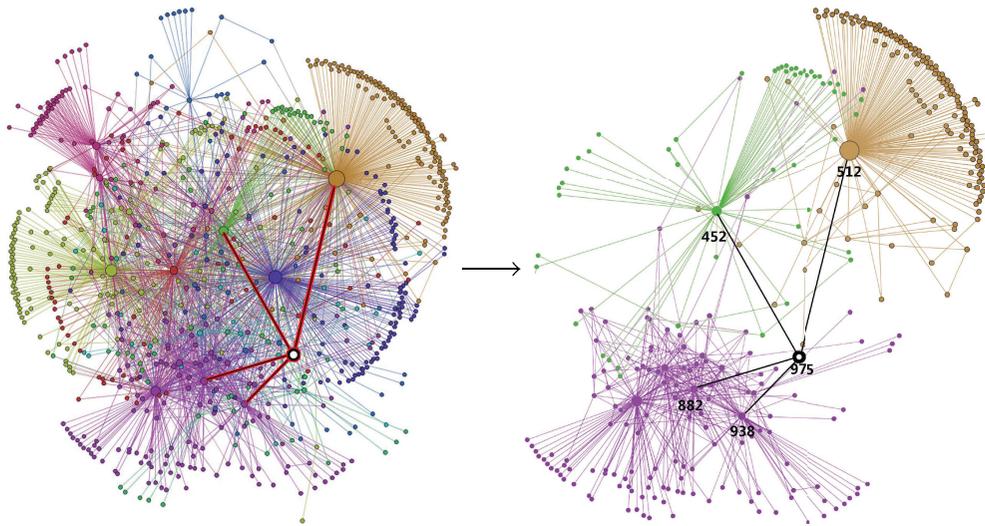


FIGURE 4: The network connections of the node ID 975.

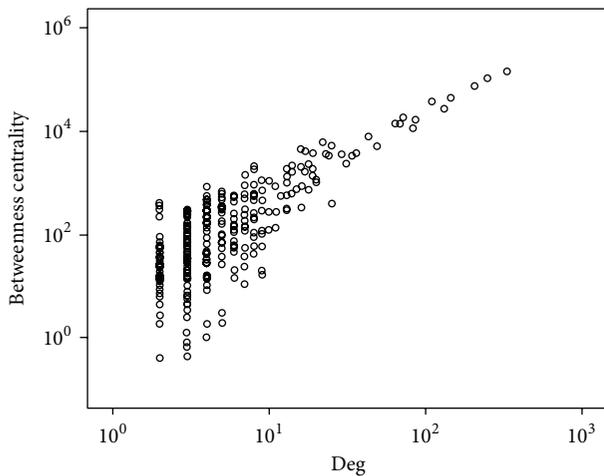


FIGURE 5: Correlations between node betweenness centralities and degrees.

strong and positive correlations between node betweenness centralities and degrees. However, the special case discussed here is the one in which the node connecting to several groups has high betweenness centrality. We rank users by betweenness centrality and find three nodes both in the top 5% list and with a degree less than 10. They connect with all communities in the network and act as bridges between the different communities. The result is consistent well with the structural holes theory that was advanced by sociologist Ronald Burt in real-life social network study.

## 6. Conclusions

In this paper, we have studied the structural properties of Microblogging ever created (839 active Sina Microblogging users and their 2112 social relations) from several viewpoints.

First of all, we have found a power-law distribution, a short average length, and a high clustering coefficient in

its topology analysis, which are all compatible with known characteristics of other online social networks and real-life social networks. In order to illuminate the mechanisms for the formation of small-world property, we have studied the difference between Sina Microblogging and traditional social networks from the aspect of users' behavior and the users' link mode and found that Sina Microblogging can easily form the user centric We Media. Therefore, it possesses small-world property that can facilitate the flow of information. Then, we calculated the correlation between clustering coefficients and degrees and showed that Sina Microblogging has apparent hierarchical structure, and we have found that Sina Microblogging trends to be disassortative network, which all mark a deviation from real-life social networks. Moreover, we analyzed the betweenness centralities of intermediary nodes and confirmed that the intermediary nodes can control the spread of information. Last but not least, our work is only the first step towards exploring the difference between the online and real-life social networks. Much work still remains.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This work was partly supported by the National Science Foundation of China under Grant no. 70903016 and the National Science & Technology Support Program under Grant no. 2012BAH81F03.

## References

- [1] R. Corten, "Composition and structure of a large online social network in the netherlands," *PLoS ONE*, vol. 7, no. 4, Article ID e34760, 2012.

- [2] R. E. Wilson and S. D. Gosling, "A review of Facebook research in the social sciences," *Perspectives on Psychological Science*, vol. 7, no. 3, pp. 203–220, 2012.
- [3] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [4] R. Albert, H. Jeong, and A.-L. Barabási, "Diameter of the world-wide web," *Nature*, vol. 401, no. 6749, pp. 130–131, 1999.
- [5] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [6] M. Barthélemy, "Spatial networks," *Physics Reports*, vol. 499, no. 1, pp. 1–101, 2011.
- [7] D. Centola, "The spread of behavior in an online social network experiment," *Science*, vol. 329, no. 5996, pp. 1194–1197, 2010.
- [8] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of topological characteristics of huge online social networking services," in *Proceedings of the 16th International World Wide Web Conference (WWW '07)*, pp. 835–844, May 2007.
- [9] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *Proceedings of the 19th International World Wide Web Conference (WWW '10)*, pp. 591–600, April 2010.
- [10] Q. Yan, L. R. Wu, and L. Zheng, "Social network based microblog user behavior analysis," *Physica A: Statistical Mechanics and Its Applications*, vol. 392, no. 7, pp. 1712–1723, 2013.
- [11] W. G. Yuan, Y. Liu, J. J. Cheng, and F. Xiong, "Empirical analysis of microblog centrality and spread influence based on Bi-directional connection," *Acta Physica Sinica*, vol. 62, no. 3, Article ID 038901, 2013.
- [12] L. Backstrom and P. Boldi, "Four degrees of separation," in *Proceedings of the 3rd Annual ACM Web Science Conference*, pp. 33–42, 2012.
- [13] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: densification and shrinking diameters," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, Article ID 1217301, 2007.
- [14] B. Gonçalves, N. Perra, and A. Vespignani, "Modeling users' activity on twitter networks: validation of Dunbar's number," *PLoS ONE*, vol. 6, no. 8, 2011.
- [15] R. I. M. Dunbar, "Social cognition on the Internet: testing constraints on social network size," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1599, pp. 2192–2201, 2012.
- [16] G. Miller, "Social scientists wade into the tweet stream," *Science*, vol. 333, no. 6051, pp. 1814–1815, 2011.
- [17] R. Lex and B. Kovacs, "A comparison of email networks and off-line social networks: a study of a medium-sized bank," *Social Networks*, vol. 34, no. 4, pp. 462–469, 2011.
- [18] S. H. Lee, P.-J. Kim, and H. Jeong, "Statistical properties of sampled networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 73, no. 1, Article ID 016102, 2006.
- [19] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of topological characteristics of huge online social networking services," in *Proceedings of the 16th International World Wide Web Conference (WWW '07)*, pp. 835–844, May 2007.
- [20] F. Fu, X. Chen, L. Liu, and L. Wang, "Social dilemmas in an online social network: the structure and evolution of cooperation," *Physics Letters A: General, Atomic and Solid State Physics*, vol. 371, no. 1-2, pp. 58–64, 2007.
- [21] K. Yuta and N. Ono, "gap in the community-size distribution of a large-scale social networking site," *Physics and Society*, <http://arxiv.org/abs/physics/0701168>.
- [22] M. Tomassini and L. Luthi, "Empirical analysis of the evolution of a scientific collaboration network," *Physica A: Statistical Mechanics and Its Applications*, vol. 385, no. 2, pp. 750–764, 2007.
- [23] K.-I. Goh, Y.-H. Eom, H. Jeong, B. Kahng, and D. Kim, "Structure and evolution of online social relationships: heterogeneity in unrestricted discussions," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 73, no. 6, Article ID 066123, 2006.
- [24] H. Hu, D. Han, and X. Wang, "Individual popularity and activity in online social systems," *Physica A: Statistical Mechanics and Its Applications*, vol. 389, no. 5, pp. 1065–1070, 2010.
- [25] P. A. Grabowicz, J. J. Ramasco, E. Moro, J. M. Pujol, and V. M. Eguiluz, "Social features of online networks: the strength of intermediary ties in online social media," *PLoS ONE*, vol. 7, no. 1, Article ID e29358, 2012.
- [26] T. A. B. Snijders, G. G. van de Bunt, and C. E. G. Steglich, "Introduction to stochastic actor-based models for network dynamics," *Social Networks*, vol. 32, no. 1, pp. 44–60, 2010.
- [27] J. Ugander and B. Karrer, "The anatomy of the facebook social graph," *Social and Information Networks*, <http://arxiv.org/abs/1111.4503>.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

