Topologies and Centralities of Replied Networks on Bulletin Board Systems

Qin Sen[1,2]*
Dai Guanzhong[2]
Wang Lin[2]
Fan Ming[2]
[1]*Hangzhou Dianzi University, School of Sciences, Hangzhou, 310018, China*
[2]*Northwestern Polytechnical University, College of Automation, Xi'an, 710072, China*

A new complex network model is studied on a Bulletin Board System (BBS) website. The model has as its vertices the identifications registered on the site, and an edge is linked if one article refers to another. Statistical characteristics of this network, such as degree distribution, average shortest-path length, and clustering coefficient are investigated. Centralities of this network including degree, betweenness, and closeness are presented. The model displays small-world characteristics with small average shortest-path lengths and a high clustering coefficient. It also exhibits scale-free topology but the exponent of power-law distribution does not exceed 2. The activity, control, and independence of communication on this network are stronger than that of some social networks. The phenomenon can be explained by the fact that the communication of viewpoints is very fast, and one person's change of view has a significant influence on other people using the BBS.

1. Introduction

In the continuing increase of research activity within physics and mathematics on the statistical characteristics of real-world complex networks, three typical models have been proposed and researched: random graph, small-world, and scale-free networks [1–5]. Because of the uniform linking probability for each pair of vertices, a random graph network is characterized by homogenous vertex degree distribution, which is a Poisson distribution [6]. Although this model has been discussed in detail and lots of theoretical results have been gained, applying this model to the research of topologies and characteristics of real-world networks is very difficult since the edges of almost all natural networks are not linked at random rigorously. Lots of available data indicate

*Present address: School of Sciences, Hangzhou Dianzi University, No.1, Street No.2, Xiasha Higher Education Zone, Hangzhou, Zhejiang Province, P.R. China, 310018. Electronic mail address: qinsen0425@gmail.com or qinsen0425@hdu.edu.cn.

that many social and technological networks have either small-world properties [7], scale-free topologies [8], or both.

Small-world networks have short average shortest-path lengths and high clustering coefficients. Disease spreading in small-world networks is faster than in random graph networks because there are some shortcuts that reduce average shortest-path lengths in the small-world networks [9]. If the vertex degree distribution of a network displays a power-law distribution, the network is a scale-free network. Most real-world scale-free networks have the characteristic that the exponent $\lambda \geq 2$ [1–3]. Scale-free networks have been shown to be robust to random failures of vertices, but vulnerable to failure of the highly connected vertices [10].

A social network is a set of people or groups each of which has connections to some or all of the others [11]. Each person is considered as a vertex and the connection between two people forms an edge. The centralities of social complex networks aim at uncovering the communication activity, control, and efficiency of vertices and entire networks. Three indices are proposed by Freeman in [12] to depict and compare the three abilities of networks: degree, betweenness, and closeness. By calculating and analyzing these indices we can estimate whether the network has a leading person (maybe a core clique) or not, and if it does, how the leader controls the whole network communication with other less significant people. Furthermore, finding the hubs of some specific networks, such as epidemic [9, 13, 14], rumor propagation [15], or terrorist networks [16], have very practical significance and interesting theoretical values.

Previous research on social networks mainly dealt with movie actor networks, scientific collaboration networks, citation networks, email networks, and bulletin board system (BBS) networks. Existing results focus on the statistical properties, centralities, and community finding of social networks. Watts and Strogatz [7] and Amaral *et al.* [17] showed that movie actor networks in the Internet Movie Database displayed small-world topologies, and Barabási and Albert [8] discovered the networks had a power-law distribution with an exponent of about 2.48. Newman [18, 19] studied the topologies of some scientific collaboration networks and found that their distributions roughly followed a power-law form. He calculated the measures of centrality such as betweenness and closeness, and gave a new algorithm for calculating betweenness for all vertices in time of order $O(mn)$ (m and n are the numbers of edges and vertices of a network, respectively). Using numerical data from citation networks, Redner [20] found that the number of papers with x citations $P(x)$ had a power-law distribution $P(x) \sim x^{-\alpha}$ with an exponent $\alpha \approx 3$. Ebel *et al.* [21] studied the topology of e-mail networks and the resulting network displayed a scale-free degree distribution and small-world behavior.

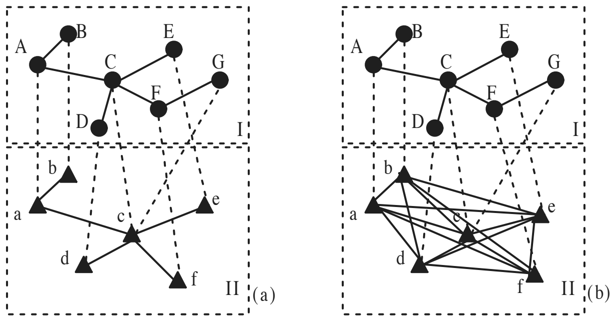

Figure 1. Example ID networks according to the relationship of replying to articles: (a) new replied network; (b) traditional replied network. (I) Replied network of one root article A and six replied articles. (II) ID network derived from the relationship of replied articles. An edge between two IDs in the new replied network indicates that opinions about a certain event are the same or different, shown as (a). For instance, although article C and G are different replied articles, they have the same ID c so there is only one edge between IDs c and f. However, an edge in the traditional network indictates that IDs enjoy the same interests on the same topic making the network shown as (b) a completed graph.

Kou *et al.* [22] constructed replied networks such that the identifications were considered as vertices and articles with replies set up the edges (this is called the "traditional reply network" in the following), and showed that these networks displayed small-world behavior and scale-free distributions. In this paper, we study a new replied network for a BBS website. The vertices in this network are the identifications (IDs) registered on the site, and there is an edge between two IDs if one replies in their article to another, as shown in Figure 1(a). We compare the new and traditional networks from the aspect of their topologies and physical significance. Then statistical properties of the new replied network, such as degree distribution, average shortest-path length, and clustering coefficient are analyzed. We discover that the model displays small-world behavior with short characteristic path lengths and a high clustering coefficient. It also has scale-free topology, but the exponent of power-law distribution does not exceed 2. The centralities indices of this network are calculated and we discover that all indices in this replied network are larger than other social networks, for example, terrorist networks.

The remainder of this paper is organized as follows. We describe the research data and preprocessing in section 2. In section 3, differences between the two replied networks are analyzed. Statistics and results about the new replied network are shown, the centralities indices are gained, and characteristics of the communication of IDs in the network are discussed. In section 4, we give a summary.

2. Data and preprocessing

The data used in this paper was downloaded from the Current Affairs Board on China Forum at http://bbs.people.com.cn/bbs/. Each article on China Forum contains the current pen-name (corresponding to only one ID registered on the website, but one ID allows using two or more pen-names), title, replied relationship, date and time, content and title of the replied articles, number of votes, agreement or disagreement, and so on. The articles posted from February 1, 2003 through July 31, 2003 are used to form the original articles' networks month-by-month. Statistical data on the corresponding replied networks are given in Table 1.

We construct the replied network according to IDs and the replied relationship of articles to gain the effective network. The effective network throws off loops and repeated edges from the original network. Then we draw the largest connected component (L-subgraph) out of the effective network. The subsequent analyses and calculations are based on the topologies of the L-subgraphs.

From Table 1, we find that the effective network retains about 60% of the vertices but only 12% of the edges from the original IDs network each month. The number of edges in the effective network is obviously less than that of the original network. This phenomenon can be explained by our emphasis on the opinions, not the interests of IDs, in the new connecting mode and there are many loops and repeated edges, as discussed in section 3.1. On the other hand, the L-subgraph retains about 99.9% of the edges and 98.7% of the vertices from the effective network, indicating that the L-subgraph has most of the information of

Month	articles_o	IDs_o	articles_e	IDs_e	articles_L	IDs_L
February	75142	3017	9983	1939	9968	1912
March	116940	3936	13132	2401	13118	2375
April	119146	4053	14092	2533	14085	2519
May	146560	4443	16978	2750	16960	2715
June	141378	5001	15872	2885	15848	2839
July	146574	4497	16278	2625	16250	2572

Table 1. Statistical data on replied networks month-by-month. The items articles_o and IDs_o represent the number of the articles and IDs of the original network, and the rest may be deduced by analogy to the effective network and L-subgraph. The effective network is undirected, acyclic, and has no repeated edges. The L-subgraph is the largest connected component of the effective network.

the effective network. This is why we discuss the properties of the new replied network based on the topologies of the L-subgraphs, not of the original networks.

3. Statistics and centralities

In this section, we first point out the differences between the traditional and new replied networks. Then, statistics such as degree distributions, average shortest-path lengths, and clustering coefficients from six months of data are described. Finally, the centralities of these models are calculated.

3.1 Differences between the two replied networks

The model proposed in this paper uses a different connecting mode than the traditional replied network on a BBS. In the traditional replied network, an edge is connected between two IDs when one of them replies to an article posted by the other. For a root article, the ID will be linked with all other IDs who have replied to this article or its sublayer of articles. Therefore, the subnetwork containing the set of these IDs is a complete graph, as shown in Figure 1(b). This linking mode indicates that these IDs have conversed on the same topics and they are more likely to enjoy the same interests [22]. The mode discussed in this paper only considers the relationship of straightforward replied articles and there is a smaller number of edges than in the traditional mode. It indicates that two IDs have communicated with each other on the same topic and that they have different or similar opinions. Therefore, the connecting mode of the new replied network reflects the communication of viewpoints since one person never replies to another article unless they have a different or similar opinion about a certain topic.

Although there are fewer edges in the new replied network than in the traditional network, the topologies of the former exhibit scale-free characteristics and have small-world behaviors (shown in sections 3.2 and 3.3) which is the same as the latter [22]. However, for the new network, the exponents of the power-law distributions are obviously less than 2, even smaller than the traditional replied network. This can be seen by comparing the data of Table 2 with the exponents of degree distributions on six other boards in Table I from [22].

3.2 Degree distributions are scale-free with $\lambda<2$

Degree distribution is an important characteristic of complex networks. An obvious characteristic of a scale-free network is that its degree distribution follows the power-law distribution, that is, the distribution follows the formula

$$P(k) \sim k^{-\lambda}, \quad (1)$$

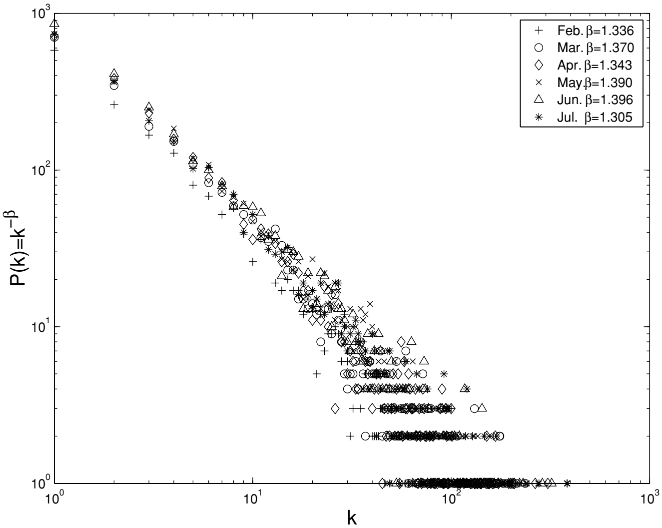

Figure 2. Degree distributions for replied networks of six months. All of them follow scale-free forms but the exponents are obviously less than 2. The range of exponents of these degree distributions is $[1.3, 1.4]$ actually.

where $P(k)$ gives the number of vertices having exactly k edges and λ is the exponent of degree distribution.

The key evolutive mechanisms of a scale-free network are continuous growth of new vertices and a preferential attachment for new edges [4]. Dorogovtsev and Mendes [23] generalized the Barabási and Albert model of growing networks and found the exact degree distribution $P(k) \sim k^{-(2+a)}$ with $a > 0$, that is, the exponent of power-law distribution $\lambda \geq 2$. The exponents of almost all real-world networks exceed 2 according to several studies [1–4, 23].

The degree distributions of the new replied network for six months are all scale-free, but their exponents are less than 2, as shown in Figure 2 and the second row of Table 2. On the one hand, the scale-free topologies of the new replied network indicate that hub vertices exist in these networks, which are the core people on the BBS. The preferential attachment mechanism is represented by the number of people that reply to the articles released by the core people. On the other hand, these exponents of scale-free networks are even smaller than those in the traditional replied network, when compared with the data in [22].

The lower exponents indicate that there are more hubs in our replied network than in the traditional network or other social and technological networks. The hubs in our replied network are some highlights paid attention to by almost all IDs.

Month	λ	L	L_{rand}	C	C_{rand}
February	1.336	3.273	3.223	0.245	0.005456
March	1.370	3.298	3.236	0.229	0.004653
April	1.343	3.281	3.244	0.253	0.004441
May	1.390	3.231	3.131	0.245	0.004603
June	1.396	3.338	3.296	0.247	0.003934
July	1.305	3.263	3.096	0.277	0.004915

Table 2. Statistical data on the replied networks month-by-month.

3.3 New replied network is a small-world network

The path length L_{ij} is the number of edges in a shortest path between two vertices v_i and v_j. And the average shortest-path length of vertex v_i is $L_{v_i} = \sum_j L_{ij}/N$. Then the average shortest-path length of a network is

$$L = \frac{1}{N}\sum_{i=1}^{N} L_{v_i}. \tag{2}$$

Let $\langle k\rangle$ be the mean degree of a random network, then its average shortest-path length [24] is

$$L_{\text{rand}} = \frac{\log(N)}{\log(\langle k\rangle)}. \tag{3}$$

For a selected vertex v_i with its degree k of a complex network, if there actually exist z_{v_i} edges among these k neighbors, the clustering coefficient of v_i can be obtained by the formula $C_{v_i} = 2z_{v_i}/[k(k-1)]$. C_{v_i} is the ratio of the actual number of edges that are among the given vertex's neighbors, to the theoretical number of edges. The clustering coefficient C of a network is

$$C = \frac{1}{N}\sum_{i} C_{v_i}. \tag{4}$$

The clustering coefficient of a random network [24] can be calculated by the formula

$$C_{\text{rand}} = \frac{\langle k\rangle}{N-1}. \tag{5}$$

Table 2 lists the average shortest-path length L and the clustering coefficient C of the new replied network on the Current Affairs BBS, together with the corresponding characteristics L_{rand} and C_{rand} of a random network.

It is obvious that the new replied network shows small-world behaviors in the resulting high C and small L values. Small L values indicate that the viewpoints can be communicated rapidly and easily from one

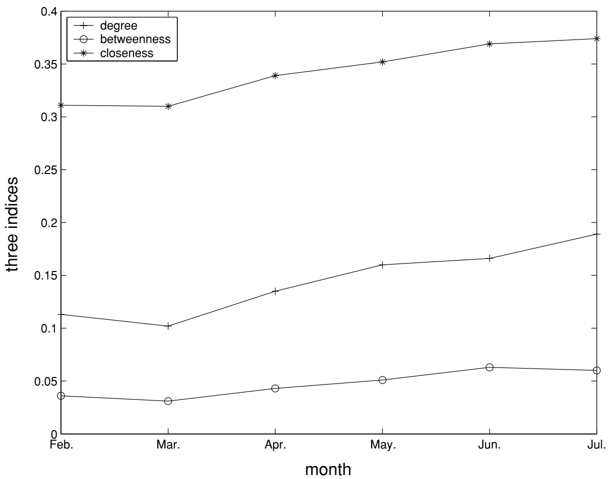

Figure 3. Degree, betweenness, and closeness indices of the new replied networks month-by-month.

person to another. This shows that personal viewpoints not only can be changed very easily during a discussion, but also that they work quickly on others' opinion. The mechanism of the new replied network presenting large C values is that the communications on a BBS can be processed conveniently. This is an inherent property of the replied network on a BBS.

3.4 Centralities of new replied network

We estimate the importance of each vertex of a social network and its significance to the network's resiliency to attack and the information spread. There are two main ways to achieve this aim. One is by analyzing the network topologies, for example, studying the degree distribution of the networks. The other is by calculating the centralities. In order to quantify the centralities, different quantities are defined in the context of social networks, such as degree, betweenness, closeness, and press centrality [25]. In this paper, we introduce the three indices of centralities—degree, betweenness, and closeness—to investigate the activity, control, and efficiency of communication of the replied network on a BBS.

We calculate the three centrality indices by formulas given in [12], and compare the data with the corresponding indices of a terrorist network given in [16]. Degree, betweenness, and closeness indices month-by-month are shown in Figure 3. There is a rough trend of the three indices to increase month-by-month. This indicates that the abilities of communication activity, control, and efficiency on the BBS are enhancing steadily. This phenomenon may have two reasons. On the one hand, important people on the BBS, such as board owners or star members, release some influential articles to attract more replied articles so that they can control the direction of discussion on the BBS. On the other hand, some people reply to these articles in order to enhance the browsing numbers of their own articles, which enhances overall the communication activity and efficiency.

According to the results of a terrorist network calculated by Krebs in [16] degree, betweenness, and closeness indices of the whole terrorist network are 0.081, 0.032, and 0.052, respectively. These indices are less than the corresponding values of the replied network. However, the core person in a terrorist network has much larger centrality indices than degree indices, with values of 0.361, 0.588, 0.587 (data of the core person, Mohamed Atta). So we discover that for the entire replied network, there is stronger communication activity, control, and efficiency of communication than the terrorist network because replying to articles is less dangerous than communication between terrorists. However, for the core people, it is weaker in controlling the communities of other people than the hubs of a terrorist network. This phenomenon accords with the result that the new replied network has more hubs than other social networks as discussed in section 3.2.

4. Discussion and conclusions

In this paper we have studied replied networks with a new connecting mode in which there is an edge between two identifications when one replies within an article to another article. The new network has fewer edges than the replied network with traditional connecting mode, so there are more hubs than in the latter network. Furthermore, as shown in section 3.2, although they display scale-free topologies, the exponents of degree distributions in the new network are very little, only in the range of 1.3 to 1.4.

Average shortest-path lengths and clustering coefficients of the networks are calculated in this paper, and we show that the topologies of the networks have small-world characteristics with smaller average shortest-path lengths and a higher clustering coefficient. If we consider the new network as an opinion network, and the traditional network as an interest network, we find that both have small-world and scale-free topologies, the same as many social networks. So not only the people's interests, but also their opinions, are accepted, taken in, and transferred on the bulletin board system very easily.

Centralities indices of the networks including degree, betweenness, and closeness are studied in detail. Because of the convenient way of replying to an article and discussing with other people on a BBS, these networks strongly exhibit communication activity, control, and efficiency of communication. However, more hubs with scale-free characteristics makes the control and efficiency of communication of the core people weaker than that of the terrorist network.

The small-world and scale-free topologies, along with the strong centralities of the new replied network, ensure that they can be controlled by means of releasing influential articles by the core people. At the same time, the ability to change someone's viewpoint is very important in order to constrain the spread of radical ideas. It may be possible to change some peoples' thinking, by communication, before they become radicals or even terrorists. As long as we master the opinions of the core people or become the core person on a BBS, and increase constantly the control of communication of the core people by releasing new influential articles, it is possible to not only control the direction of discussion, but also to prevent some people from accepting radical viewpoints on a bulletin board system.

Acknowledgments

We thank the fellows of Control and Network Institute for helpful discussions and checking the consistency of this paper. This paper was partially supported by the National High-Tech Research and Development Plan of China.

References

[1] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of Networks," cond-mat/0106144 (2001).

[2] R. Albert and A. L. Barabási, "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics*, **74** (2002) 47–97.

[3] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, **45** (2003) 167–256.

[4] Xiao Fan Wang and Guan Rong Chen, "Small-World, Scale-Free and Beyond," *IEEE Circuits and Systems Magazine*, **3** (2003) 6–20.

[5] S. N. Dorogovtsev and J. F. F. Mendes, "The Shortest Path to Complex Networks," cond-mat/0404593 (2004).

[6] S. Janson, T. Luczak, and A. Rucinski, *Random Graphs* (Wiley & Sons, New York, first edition, 2000).

[7] D. J. Watts and S. H. Strogatz, "Collective Dynamics of 'Small-World' Networks," *Nature(London)*, **393** (1998) 440–442.

[8] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, **286** (1999) 509–512.

[9] C. Moore and M. E. J. Newman, "Epidemics and Percolation in Small-World Networks," *Physical Review E*, **61** (2000) 5678–5682.

[10] R. Albert, H. Jeong, and A.-L. Barabási, "Error and Attack Tolerance of Complex Networks," *Nature(London)*, **406** (2000) 378–382.

[11] M. E. J. Newman, "Assortative Mixing in Networks," *Physical Review Letters*, **89** (2002) 208701.

[12] L. C. Freeman, "Centrality in Social Networks: Conceptual Clarificaiton," *Social Networks*, **1** (1978/79) 215–239.

[13] V. M. Eguíluz and K. Klemm, "Epidemic Threshold in Structured Scale-Free Networks," *Physical Review Letters*, **89** (2002) 108701.

[14] R. Pastor-Satorras and A. Vespignani, "Epidemic Spreading in Scale-Free Networks," *Physical Review Letters*, **86** (2000) 3200–3203.

[15] D. H. Zanette, "Dynamics of Rumor Propagation on Small-World Networks," cond-mat/0110324 (2001).

[16] V. E. Krebs, "Uncloaking Terrorist Networks," *First Monday*, **7** (April, 2000).

[17] L. A. N. Amaral, A. Scala, M. Barthélémy, *et al.*, "Classes of Small-World Networks," *Proceedings of the National Academy of Sciences USA*, **97** (2000) 11149–11152.

[18] M. E. J. Newman, "Scientific Collaboration Networks. I. Network Construction and Fundamental Results," *Physical Review E*, **64** (2001) 016131.

[19] M. E. J. Newman, "Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality," *Physical Review E*, **64** (2001) 016132.

[20] S. Redner, "How Popular Is Your Paper? An Empirical Study of the Citation Distribution," *The European Physical Journal B*, **4** (1998) 131–134.

[21] H. Ebel, L.-I. Mielsch, and S. Bornholdt, "Scale-Free Topology of E-mail Networks," *Physical Review E*, **66** (2002) 035103(R).

[22] Kou Zhongbao and Zhang Changshui, "Reply Networks on a Bulletin Board System," *Physical Review E*, **67** (2003) 036117.

[23] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, "Structure of Growing Networks With Preferential Linking," *Physical Review Letters*, **85** (2000) 4633–4636.

[24] J. Davidsen, H. Ebel, and S. Bornholdt, "Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks," *Physical Review Letters*, **88** (2002) 128701.

[25] K. A. Lehmann and M. Kaufmann, "Decentralized Algorithms for Evaluating Centrality in Complex Networks," Technical Report of the Wilhelm-Schickard-Institut, WSI-2003-10, *ISSN 0946-3852* (2003).