

# A Growing Network Model with High Hub Connectivity and Tunable Clustering Coefficient

YIHJIA TSAI, PING-NAN HSIAO

Department of Computer Science and Information Engineering  
Tamkang University  
151 Ying-chuan Road, Tamsui, Taipei  
TAIWAN  
tsai@cs.tku.edu.tw, daniel@ssp.sinica.edu.tw

*Abstract:* - The paper investigates a model for social network based on a growing network. The model we propose can fit the following two basic properties: (1) ‘scale-free property’ which has power law degree distribution, and (2) ‘small-world property’ which has high clustering coefficient. In addition, the model is modified for high hub connectivity which is a new concept in scale-free network. This study also provided an analysis of empirical data.

*Key-Words:* - Social network model, growing network, scale-free, small-world, complex network.

## 1 Introduction

The signatures of complex network appear in many fields, such as biology [1], economy, computer science, and social networks [2]. In these fields, the social network models [3] [4] [5] [6] [7] have been used for the simulation of epidemiological transmission [8], rumor spreading [9] and the spread of computer viruses in email network [10]. Since social networks can be found in many aspects of our lives, it is worth to investigate how they work precisely. In this paper, we propose a scale-free property with tunable clustering coefficient social network model based on a growing network. Then we modify the network model according to  $s_{max}$  property. In addition, this study includes the analysis of real social network data.

### 1.1 Scale-Free

A scale-free network is inspired by empirical studies such as the Internet backbone network [11], web documents link network [12], and email network [13]. The characteristic property of scale-free network is that it consists of a few large connected nodes and many nodes with small connectivity. Barabási and Albert proposed an algorithm to construct a scale-free network [14]. The network grows by adding one new node at a time, denoted by  $t$ , that connects to  $m$  nodes already in the network by a probability function. The connecting probability depends on the node degree it was connected to. For node  $i$  with node degree  $k_i$ , the connecting probability is denoted by

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \quad (1)$$

The connectivity distribution  $pr(k)$  of the resulting network, which is defined by the probability of a node having  $k$  edges. The connectivity distribution for scale-free networks is a decaying function of  $k$ ’s power following

$$pr(k) = k^{-\gamma}. \quad (2)$$

This phenomenon is sometimes called power law distribution and  $\gamma$  is the tail index. Power law degree distribution is considered as a defining characteristic between real communication network and random network. The BA model for scale-free network construction with a preferential attachment strategy creates a scale-free network with tail index  $\gamma=3$ .

### 1.2 Small-World

The small-world network is known to have both small average network distance and large clustering coefficient. The average distance between network nodes, denoted by  $L$ ,

$$L = \frac{\sum_{i=1, j>i}^N \delta(i, j)}{N(N-1)/2}, \quad (3)$$

where  $\delta(i, j)$  is the distance of shortest path between node  $i$  and node  $j$ , and  $N$  is the number of nodes in the network. The  $L$  indicates the average distance to communicate from one node to the other.

The clustering coefficient describes how close among the neighbors of a node. The formula of calculating clustering coefficient of node  $i$  is the ratio of actual number of edges connecting these nodes in its  $k$  neighbors to the number of edges in a fully connected network of the  $k$  nodes, denoted by  $C_i$ ,

$$C_i = \frac{2E_i}{k_i(k_i - 1)}, \quad (4)$$

where  $E_i$  stands for the number of edges in the sub-network formed by node  $i$  and its  $E_i$  neighbors. The clustering coefficient of the entire network is defined as the average of all nodes

$$C = \frac{\sum_{i=1}^N C_i}{N}. \quad (5)$$

In 1998, Watts and Strogatz [15] proposed a small-world model that formally defines an algorithm to construct and analyze such network. The WS network model is based on a regular lattice network and a re-wiring probability  $p$ . Each edge in the base regular network is modified with probability  $p$  to re-connect to a different destination node. For  $p=0$  the original lattice network is unchanged. As  $p$  increases the network becomes increasingly disordered until for  $p=1$  all edges are rewired randomly. Increasing  $p$  increases randomness.

### 1.3 S-metric

In 2005, Li et al. [16] defined an "s-metric", that measures the extent to which high-degree nodes connect to other high-degree nodes. Let  $g$  be an undirected graph with edge-set  $\mathcal{E}$ , and let degree of node  $i$  be  $d_i$ , then define the metric

$$s(g) = \sum_{(i,j) \in \mathcal{E}} d_i d_j, \quad (6)$$

and the  $s_{max}$  is the maximum value of  $s(g)$ . The value of  $s(g)/s_{max}$ , denoted by  $S$ , describes the level of a large high-degree nodes connected to another large high-degree nodes, and it is more close to 1 if more high-degree nodes are connected in the network. In this paper, we use the algorithm, shown in Table 1, to implement the  $s_{max}$  in simple graph. The time complexity of the algorithm is  $O(n^2 \lg n)$ , since there is a sorting algorithm for square edge set.

## 2 Empirical Networks

Many real world social networks such as the collaboration network of movie actors [17], email communication of students [7], and the co-authorship network of neuroscientists [3], show both the property of the small-world and that of the scale-free. We collected a real data from the People Fisher Website, denoted by PFW, at September 2006 [18], which is a web site of a blog style. Everyone can create a blog, share photos, and make new friends in the web site. We analyzed the user data with their setting of friends. If user A considers user B as his/her friend while user B also sets user A as his/her friend in their web pages, then one edge between node A and B is added. The total number of users on the web site is 2789, and the largest connected network consists of 2738 nodes and 9333 edges.

Fig. 1 shows the degree distribution of the PFW network, and it shows the scale-free property obviously. Table 2 is the resulting average distance  $L$ , clustering coefficient  $C$ , tail index  $\gamma$  and s-metric  $S$ .

In the PFW network, the average distance  $L$  is very small, since the maximum node degree, the creator of the website, is 2629. The creator sets about 96% of the total users to be his friends. Therefore, in terms of s-metric, this causes the results in the maximum node greater than the second node by a large amount.

Table 1 The algorithm of  $s_{max}$  in simple graph.

- |  |
|--|
| Step 1. Let edge set $\mathcal{E}' = \bigcup_{i=1, j>i}^N d_i d_j$   |
| Step 2. Sort edge set $\mathcal{E}'$ in decreasing order.  |
| Step 3. Construct a new graph $g'$ with $\mathcal{E}'$ in order without multi-link, and keep its degree sequence constant. |
| Step 4. Calculate $s(g')$ for the new graph, which is $s_{max}$ .  |

Table 2 The empirical data of the PFW network.

Node $N$	2738
Average degree $k$	6.817
Tail index $\gamma$	1.292
Clustering coefficient $C$	0.268
Average distance $L$	2.073
S-metric $S$	0.839

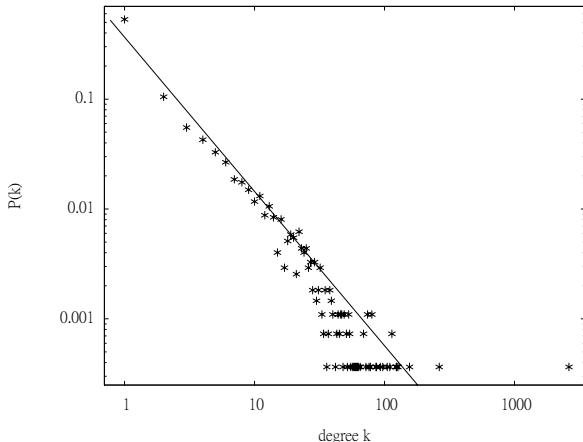


Fig. 1 The log-log degree distribution plot of the PFW network, the slope of the line is -1.292.

### 3 New Model

We proposed a model based on the BA model to describe social network topology. The BA model constructs a scale-free network with decreasing clustering coefficient when network size  $N$  increases. We add a parameter  $f$  to increase the clustering coefficient by increasing the probability of friendship in the network. For example, when you know a nice person, you may want to know his/her friend. At last, when the network is constructed, we then add extra edges to satisfy the average degree of the real network that is what we need.

The average degree  $K$  in BA model is

$$K = \frac{2mt}{t + m_0}, \quad (7)$$

when  $t$  is large enough, the average degree  $K$  will nearly equal to  $2m$  which is almost even number. In the real world, the degree of most networks will not be even, so we add extra edges to fit the average degree  $k$  of the real network. The number of extra edges should be added to the network is

$$N(k - K)/2 = N(k - \frac{2mt}{t + m_0})/2. \quad (8)$$

When adding the extra edges  $(i, j)$  which follows the weight of  $s_{max}$  probability:

$$P(i, j) = \frac{d_i d_j}{\sum d_i d_j}, \quad (i, j) \notin \varepsilon, \text{ and } i \neq j. \quad (9)$$

It will make the large hub-nodes have large probability to connect to large ones rather than small degree nodes connect to other small nodes. The

reason is that each person has only limited time, so he/she would rather choose to meet a famous person than others. Furthermore, if we hope the negotiation will be more successful in a social network, the famous person is likely to know more other famous persons. For example, in a company, there are usually some departments. If the leaders of the departments not direct communicate with other leaders, the negotiation of the social network will not smooth. Thus, the high-degree nodes of a social network should have more large probability to connect to other ones.

To avoid the “winners take all” phenomenon [19] in adding extra edges phase, we add a parameter  $x$  to adjust the weight of  $s_{max}$ . The probability  $x$  is for preferential attachment of  $s_{max}$  and  $1-x$  is for uniform attachment. So (9) will be modified to

$$P(i, j) = x \frac{d_i d_j}{\sum d_i d_j} + (1-x) \frac{1}{\sum (i, j)}, \quad (i, j) \notin \varepsilon, \text{ and } i \neq j. \quad (10)$$

In adding one new node and its edges, we define its preferential attachment as PA, and the attachment of making friend as FA. Table 3 shows the comparison with our model, the growing model proposed by Holme and Kim [4], and the dynamic model proposed by Tsai, Lin and Hsiao [7] when adding one new node. The main difference of three models is whom is the FA connecting to. The step by step diagram of our model when adding one new node shows in Fig. 2.

In fact, the tunable clustering coefficient methods of the three models are not too different, and we do not compare them in detail. The only thing that we want to do is to show that we do not need use the same way to control the clustering coefficient.

Table 3 The comparison with three models when adding one new node. The “ $\leftarrow$ ” means FA connects to the neighbor node of PA or FA.

	The steps of PA and FA
Our model	PA <sub>1</sub> $\leftarrow$ FA <sub>1</sub> , PA <sub>2</sub> , PA <sub>3</sub> $\leftarrow$ FA <sub>2</sub> $\leftarrow$ FA <sub>3</sub> $\leftarrow$ FA <sub>4</sub> , ...
Growing model [4]	PA <sub>1</sub> $\leftarrow$ FA <sub>1</sub> , PA <sub>2</sub> , PA <sub>3</sub> $\leftarrow$ FA <sub>2</sub> , FA <sub>3</sub> , FA <sub>4</sub> , ...
Dynamic model [7]	PA <sub>1</sub> $\leftarrow$ FA <sub>1</sub> , PA <sub>2</sub> , PA <sub>3</sub> , FA <sub>2</sub> , FA <sub>3</sub> , FA <sub>4</sub> , ...

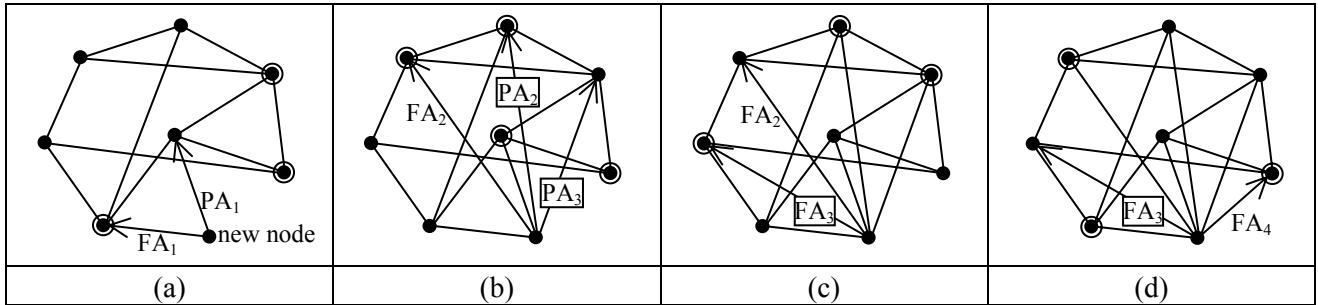


Fig. 2 The double circle nodes are the neighbor nodes of PA or FA. (a)FA<sub>1</sub> connects to the neighbor node of PA<sub>1</sub>; (b)FA<sub>2</sub> connects to the neighbor node of PA<sub>3</sub>; (c)FA<sub>3</sub> connects to the neighbor node of FA<sub>2</sub>; (d)FA<sub>4</sub> connects to the neighbor node of FA<sub>3</sub>.

#### 4 Simulations and Discussion

We use our model to simulate the proposed social network model and compare with the PFW experimental results. Table 4 summarized the comparison. In the simulation, we use (9), that is  $x=1$  in (10), in the proposed model since the  $S$  of the PFW network is too large, and the suitable  $f$  in our model for the PFW network is 0.35. Fig. 3 shows the degree distribution of our model for the PFW network with  $f=0.35$ , and it shows the scale-free property. Fig. 4 demonstrates the  $S$  in two conditions: constructed network before extra edges added, and extra edges added by (9) for  $s_{max}$  and by (10) for random when  $x=0$  with different  $f$ . We find that our model can raise the  $S$  significantly. Fig. 5 shows the clustering coefficient relationship with the same parameters of Fig. 4, and this is absolutely for large  $f$  made large  $C$ . Fig. 6 shows the tail index relationship with the same parameters of Fig. 4, and the more  $f$  makes, the less  $\gamma$  is.

At the same time, we use other parameters to simulate the social network. Fig. 7 shows the results of  $S$  with  $N=2000$ ,  $m=3$  and  $f=0.10$  for various additional edges. We find that more extra edges added by (10) with larger  $x$ , the higher the  $S$  will become than by random when  $x=0$ . In the right portion of this figure, the breadth of increasing is lower than left. The reason is that the more additional edges added by  $s_{max}$  property cause the most large high-degree nodes to be connected, and this is less helpful for the  $S$  by connecting nodes with small degree.

Fig. 8 shows the clustering coefficient  $C$  with the same parameters of Fig. 7. When we add additional edges by (10) with larger  $x$ , the  $C$  will increase slowly; but if we add by (10) with  $x$  below 0.3, the result is decreasing  $C$ . The reason is that the more large hub-nodes connected make the neighbor nodes of the large-hub nodes will be large ones.

Fig. 9 demonstrates the tail index  $\gamma$  with the same parameters of Fig. 8. We find that the more additional edges added by (10) with larger  $x$ , the lower  $\gamma$  will become than by (10) with small  $x$ . The reason is that (10) will make large hub-nodes become larger when  $x=1$ , and the position of these nodes will be on the right portion in log-log degree distribution plot, and that causes the slope of line, that is  $- \gamma$ , to appear flatter than before, thus resulting in smaller  $\gamma$ .

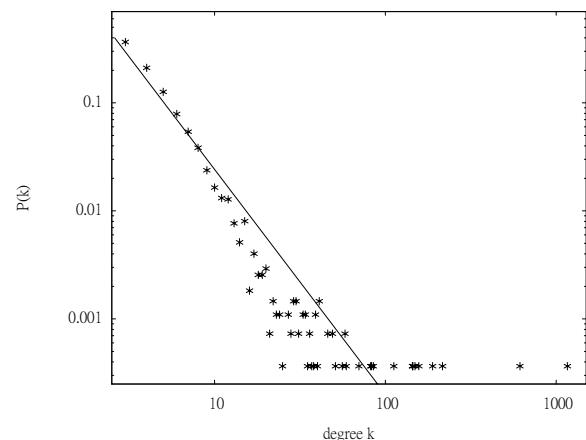


Fig. 3 The log-log degree distribution plot of our model with  $N=2738$ ,  $k=6.817$ ,  $f=0.35$ . The slope of the line is  $-1.290$ .

Table 4 The comparison between the PFW and our model.

	PFW	The different parameter $f$ in our model with parameter $x$ set 1 for all the cases.								
		$f=0.05$	$f=0.10$	$f=0.15$	$f=0.20$	$f=0.25$	$f=0.30$	$f=0.35$	$f=0.40$	$f=0.45$
$C$	0.268	0.106	0.130	0.157	0.184	0.214	0.245	0.276	0.307	0.338
$\gamma$	1.292	1.557	1.509	1.468	1.416	1.374	1.328	1.285	1.238	1.194
$L$	2.073	3.133	3.112	3.110	3.088	2.975	2.973	2.972	2.875	2.820
$S$	0.839	0.816	0.795	0.758	0.734	0.721	0.717	0.715	0.713	0.711

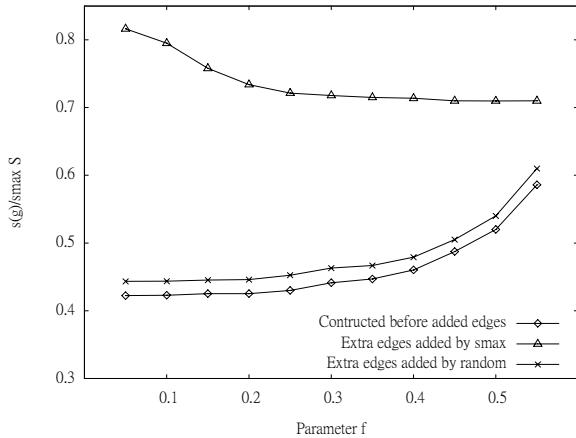


Fig. 4 The value of  $S$  with  $N=2738$ ,  $f=0.05\sim0.55$ . The average degree  $K$  by (7) in the constructed network before extra edges added is 5.991, and the  $k$  of the others after edges added by (8) is 6.817.

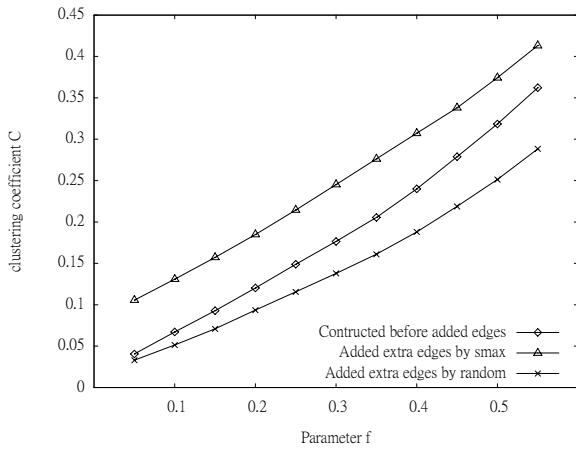


Fig. 5 The clustering coefficient  $C$  with  $N=2738$ ,  $f=0.05\sim0.55$ . The average degree  $K$  by (7) in the constructed network before extra edges added is 5.991, and the  $k$  of the others after edges added by (8) is 6.817.

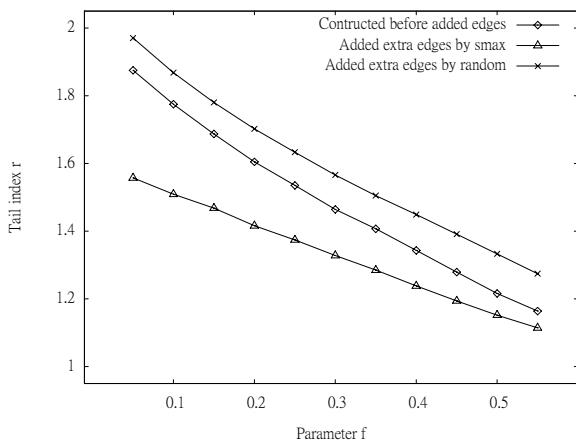


Fig. 6 The tail index  $\gamma$  with  $N=2738$ ,  $f=0.05\sim0.55$ . The average degree  $K$  by (7) in the constructed network before extra edges added is 5.991, and the  $k$  of the others after edges added by (8) is 6.817.

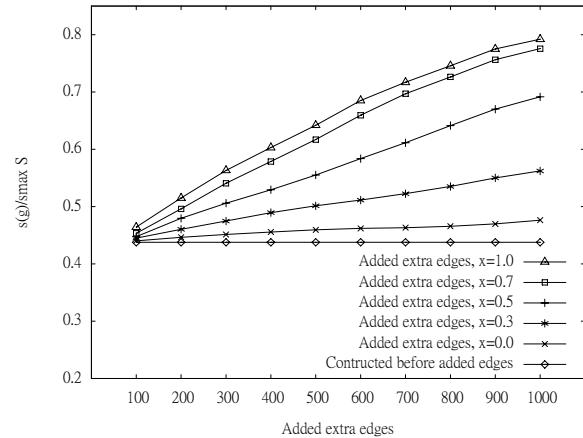


Fig. 7 The  $S$  with  $N=2000$ ,  $m=3$ ,  $f=0.10$ , and the extra edges added are from 100 to 1000. The parameter  $x$  is from 0 to 1.

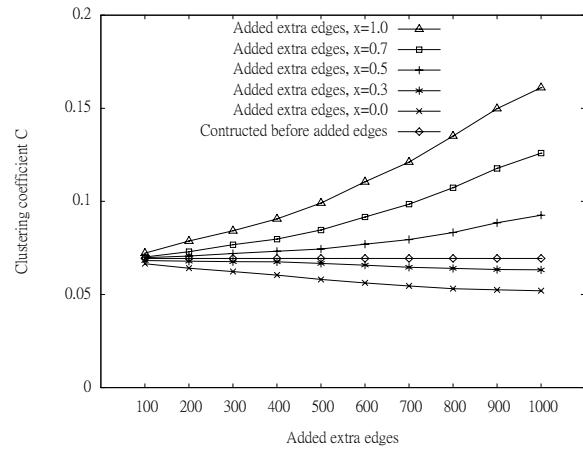


Fig. 8 The clustering coefficient  $C$  with  $N=2000$ ,  $m=3$ ,  $f=0.10$ , and the extra edges added are from 100 to 1000.

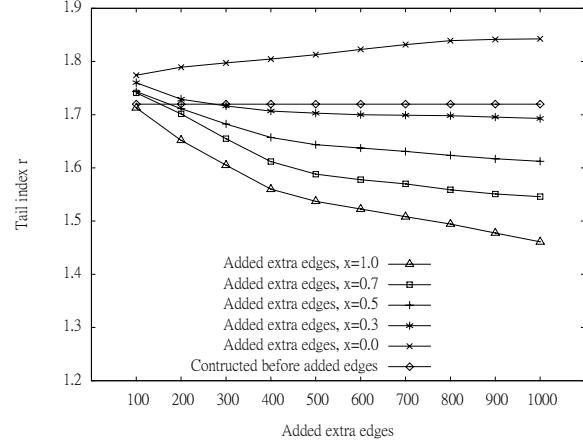


Fig. 9 The tail index  $\gamma$  with  $N=2000$ ,  $m=3$ ,  $f=0.10$ , and the extra edges added are from 100 to 1000.

## 5 Conclusions and Future Work

In this paper, we present a social network model that combines the degree distribution property of a scale-free network model and the clustering coefficient

property of a small-world network. In the connectivity of high-degree nodes, our model can provide very high value when parameter  $x=1$  than random when  $x=0$ . In addition, the simulation result of the proposed model closely matches the empirical network degree distribution and the clustering coefficient. Although the shortest path and s-metric can not fit perfectly, since the maximum node of empirical network is too large, our model still provide a good way to tune the level of high hub connectivity.

The social network analysis can be used in business for improving knowledge creation and

#### Reference:

- [1] R. Dobrescu, Modeling complex biological systems using Scale Free Networks, *Proceedings of the 5th WSEAS International Conference on Non-Linear Analysis, Non-Linear Systems and Chaos*, 2006, pp. 57-62.
- [2] S. H. Strogatz, Exploring complex networks, *Nature* 410, 2001, pp. 268-276.
- [3] A.-L. Barabási, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, T. Vicsek, Evolution of the social network of scientific collaborations, *Physica A* 311, 2002, pp. 590-614.
- [4] P. Holme, B. J. Kim, Growing scale-free networks with tunable clustering, *Physical Review E* 65, 2002, 026107.
- [5] K. Klemm, V. M. Eguiluz, Growing scale-free networks with small-world behavior, *Physical Review E* 65, 2002, 057102.
- [6] R. Toivonen, J.-P. Onnela, J. Saramäki, J. Hyvönen, K. Kaski, A Model for social networks, *Physica A* 371, 2006, pp. 851-860.
- [7] Y. Tsai, C.-C. Lin, P.-N. Hsiao, Modeling Email Communications, *IEICE Transactions on Information and Systems*, Vol.E87-D No.6, 2004, pp. 1438-1445.
- [8] L. A. Meyers, B. P. Pourbohloul, M. E. J. Newman, D. M. Skowronski, R. C. Brunham, Network theory and SARS: predicting outbreak diversity, *Journal of Theoretical Biology* 232, 2005, pp. 71-81.
- [9] Y. Moreno, M. Nekovee, A.F. Pacheco, Dynamics of rumor spreading in complex networks, *Physical Review E* 69, 2004, 066130.
- [10] M. E. J. Newman, S. Forrest, J. Balthrop, Email networks and the spread of computer viruses, *Physical Review E* 66, 2002, 035101.
- [11] M. E. Crovella, M. S. Taqqu, A. Bestavros, Heavy-tailed probability distributions in the World Wide Web, *In A Practical Guide To Heavy Tails*, Chap.1, 1998, pp.3-25.
- [12] R. Albert, H. Jeong, A.-L. Barabási, Internet: Diameter of the World-Wide Web, *Nature* 401, 1999, pp. 130-131.
- [13] H. Ebel, L.-I. Mielsch, S. Bornholdt, Scale-free topology of e-mail networks, *Physical Review E* 66, 2002, 035103.
- [14] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, *Reviews of Modern Physics*, Vol.74, 2002, pp. 48-97.
- [15] D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks, *Nature* 393, 1998, pp. 440-442.
- [16] L. Li, D. Alderson, J. C. Doyle, W. Willinger, Towards a theory of scale-free graphs: definition, properties, and implications, *Internet Mathematics*, 2(4), 2005, pp. 431-523.
- [17] A.-L. Barabási, R. Albert, Emergence of Scaling in Random Networks, *Science* 286, 1999, pp. 509-512.
- [18] <http://www.peoplefisher.com/>
- [19] D. M. Pennock, G. W. Flake, S. Lawrence, E. J. Glover, and C. L. Giles, Winners don't take all: Characterizing the competition for links on the web, *Proceedings of the National Academy of Sciences*, 99(8), 2002, pp. 5207-5211.
- [20] M. Huysman, V. Wulf, IT to Support Knowledge Sharing in Communities, Towards a Social Capital Analysis, *Journal of Information Technology*, Vol. 21, 2006, pp. 40-51.
- [21] S. Mocanu, S. Taralunga, Immunization strategies for networks with scale-free topology, *Proceedings of the 5th WSEAS International Conference on Non-Linear Analysis, Non-Linear Systems and Chaos*, 2006, pp. 45-50.
- [22] A. Abbott, H. Pearson, Fear of human pandemic grows as bird flu sweeps through Asia, *Nature* 427, 2004, pp. 472-473.