

# A Social Network Model Based on Topology Vision

Ping-Nan Hsiao

Research Center for Humanities and Social Sciences, Academia Sinica  
128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan  
daniel@ssp.sinica.edu.tw

**Abstract.** There are many researchers proposed social network models in recent years, and most of them focus on clustering coefficient property of a small-world network and power law degree distribution of a scale-free property. In social network topology, we observed the network is consisted of many nodes with small connectivity and a few high-degree nodes. In the small connectivity part, there are many nodes which have only one degree. Most of past social network models can not generate this part. In this paper, we proposed a social network model based on topology vision and with tunable high hub connectivity. At the same time, we suggested a new characteristic of social network, condensed clustering coefficient, to replace the original clustering coefficient. Finally, this study also includes the analysis of real social network data.

**Keywords:** Social Network, Network Model, BA Model, Small-World, Scale-Free, Clustering Coefficient, Condensed Clustering Coefficient.

## 1 Introduction

Social networks are a description of relationship between people. The relationships include friends, co-authorships, human sexual relations, and other interactions between people. In this study, an individual person is as a node and the relationship between two persons as an edge. If there is any relationship between two persons, one edge is added between the two nodes. Therefore, a social network is constructed. Recent research about social network analysis [1] [2] [3] [4] [5] [6] focused on observing clustering coefficient property, and power law degree distribution phenomena. In this paper, we proposed a social network model based on topology vision and with tunable high hub connectivity.

In last ten years, many social network models were proposed [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16]. The social network model has been used for the simulation of epidemiological transmission [17] [18], rumor spreading [19] [20], opinion formation [21] and the spread of computer viruses in email network [22]. Since social networks can be found in many aspects of our lives, it is worth to investigate how they work precisely.

In this paper, some background knowledge is in Section 2. In Section 3, the data of real social network from Internet is analyzed. In Section 4, a social network model is proposed, which both have scale-free property and high clustering coefficient. In addition, the network model contains tunable high-hub connectivity property. The

simulation results are shown in Section 5. Conclusions and our discussion for future works are presented in Section 6.

## 2 Background Knowledge

In this chapter, some characteristics of network topology which are used in this study are introduced. A network consists of a set of nodes and many edges between the nodes. In this paper, the network is simplified as a connected network with non-weighted and non-directed edge.

### 2.1 Average Shortest Path

The average shortest path is the average of shortest path between all pair nodes in the network. The average shortest path (sometimes called average distance) between network nodes, denoted by  $L$ , is calculated by the formula.

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

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.

### 2.2 Power Law Degree Distribution

The power law degree distribution is a property of a scale-free network. A scale-free network is inspired by empirical studies such as the Internet backbone network [23], web documents link network [24], and email network [25] [26]. The main characteristic property of a scale-free network is that it consists of a few large connected nodes and many nodes with small connectivity.

In 1999, Barabási and Albert [27] proposed the BA model to describe the scale-free network. The network of BA model grows by adding one new node at a time, denoted by  $t$ , which randomly connects to  $m$  nodes already in the network. The connecting probability depends on the node degree it was connected to. For node  $i$  with node degree  $k_i$ , the connecting probability  $\Pi(k_i)$  is denoted by

$$\Pi(k_i) = \frac{k_i}{\sum_{j=1}^N k_j}, \quad (2)$$

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

$$pr(k) = k^{-\gamma}. \quad (3)$$

This phenomenon is sometimes called power law distribution and  $\gamma$  is the tail index. The BA model for scale-free network construction with a preferential attachment strategy creates a scale-free network with tail index  $\gamma=3$ .

### 2.3 Clustering Coefficient

In 1998, Watts and Strogatz (WS) [5] proposed the small-world phenomena. The small-world network is between regular and random network. In the small-world network, the clustering coefficient is large, and the average shortest path is small.

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  $E_i$  is 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.

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

### 2.4 Condensed Clustering Coefficient

In a social network, there are some nodes which only have one degree. We defined the  $O$  is the ratio of how many nodes which only have one degree to the total nodes in the network. If there are more nodes which only have one degree, the  $O$  will be higher. For the nodes which only have one degree, their clustering coefficient will be zero, since the  $E_i$  in (4) will be zero. Because we did not include the self-loop in shortest path  $L$  in (1) which is zero, we consider the condensed clustering coefficient  $C'$ , which is also not to include the  $C_i$  which is zero. For the  $C'$ , the formula is

$$C' = \frac{C}{1 - O}. \quad (6)$$

### 2.5 S-Metric

In 2005, Lun Li et al. proposed the s-metric. In a society, the famous people represent high-degree nodes in a social network. The s-metric measures the level of high-degree nodes connectivity. Let  $g$  be an undirected graph with edge-set  $\varepsilon$ , and let degree of node  $i$  be  $d_i$ , then the metric is defined as

$$s(g) = \sum_{(i,j) \in \varepsilon} d_i d_j, \quad (7)$$

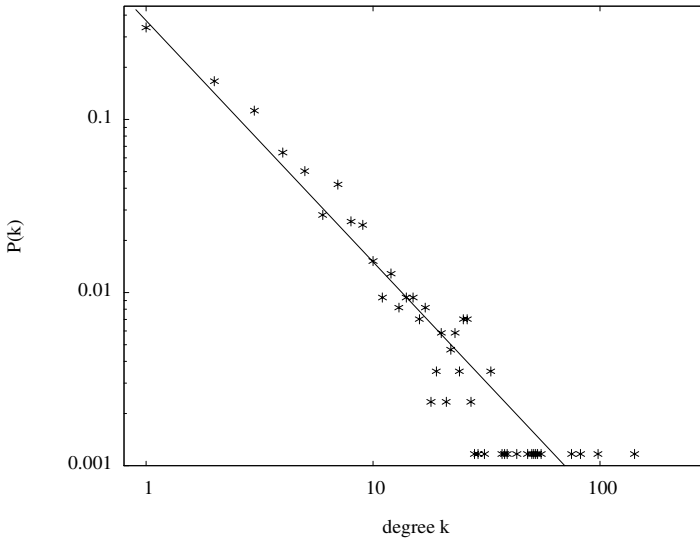
and the  $s_{max}$  is the maximum value of  $s(g)$ . The value of  $s(g)/s_{max}$ , denoted by  $S$ , describes the level of a large hub connected to another large hub, and the more high-hub nodes are connected, the higher value of  $S$  is.

### 3 The Real Network Data

We collected the real network data from the Opel Astra Club (OAC) in Taiwan [28] at June 2008. This is a forum based website. Most of users in the website own and drive an Opel Astra car. They can post articles to forum, upload photos to album or send private messages to other users. We analyzed their private message database. If user A sent a message to user B, and user B also sent a message to user A, in addition, they kept the message in their inbox and not deleted, then we added one edge between node A and node B. In the website, the total number of the private message database was 21,691, and total number of users which had used private messages was 1,378. After deleting duplicated records, the construct connected network was consisted of 835 nodes and 2,223 edges.

**Table 1.** The results of the Opel Astra Club

	OAC
Node $N$	855
Average shortest path $L$	3.574
Average shortest path in random network $L_{random}$ [29]	3.882
Average degree $k$	5.691
One degree in the network $O$	0.339
Tail index $\gamma$	1.439
Clustering coefficient $C$	0.0979
Condensed clustering coefficient $C'$	0.1481
Clustering coefficient in random network $C_{random}$ [30]	0.0067
S-metric $S$	0.645



**Fig. 1.** The degree distribution plot of the OAC

Fig. 1 shows the degree distribution plot of the OAC network. In this figure, we can see clearly the degree distribution following a straight line, which indicates distinguished scale-free property. Table 1 shows the results of the average distance  $L$ , clustering coefficient  $C$ , tail index  $\gamma$  and  $s(g)/s_{max}$  value of the OAC network. The tail index  $\gamma$  is below 3, which indicates scale-free property. The clustering coefficient is much higher than corresponding random network, which indicates that there are many neighbor nodes have the same neighbor nodes.

## 4 New Model

In topology vision, we found that there are many nodes which only have one degree in social networks, which will appear in the position which is 1 of x-axis of degree distribution. In BA model, it cannot generate any nodes which only have one degree. At the same time, we reviewed many social network papers which have presenting degree distribution of real network; most of them have lots of nodes which only have one degree. Therefore, we decide to design a model to fit it.

Our model is based on BA model, and then modified for clustering coefficient property and high-hub connectivity. In the real social network data, there are lots of nodes which degree number is one, but if we use the BA model to construct a network, every node has at least  $m$  edges. There are not any nodes which have only one degree unless the  $m$  is 1, but this will cause the average degree  $2m$  is limited to 2. Therefore, we modified the growing stage of BA model. The algorithm of this model is described as follows:

**Step1:** Initialize stage: Add  $m$  nodes into the network, and add one node which connects to the  $m$  nodes by  $m$  edges. There are  $m+1$  nodes and  $m$  edges in the network in this stage.

**Step2:** Growing stage: We defined four functions at this stage: First Attach (FA): The edge is connected between the new adding node and the exist nodes of the network which follow preferential attachment; Friend Function (FF): The edge is connected between the two unlinked neighbor nodes of previous link nodes. The two neighbor nodes have weight by its degree number in connecting condition;  $s_{max}$  Function (SF): The edge is connected between the two unlinked nodes which follow  $s_{max}$  property; Random Function (RF): The edge is connected between any two unlinked nodes which are random chosen.

After initialize stage, we have  $(N-(m+1))$  loops to add nodes and edges. Let total degree number be  $K$ , therefore the average edge, denoted by  $E_{loop}$ , at every loop should be added is

$$E_{loop} = \frac{K - m}{2(N - (m + 1))}. \quad (8)$$

At every loop, it begins with adding one new node and connecting to the network by preferential attachment, which is FA. Other edges connect between two nodes which are FF or SF or RF by parameter  $f$  and  $x$ . The three functions are included in the Attach\_edge function of our algorithm. The parameter  $f$  stands for clustering coefficient part and parameter  $x$  represents high-hub connectivity part. In addition, we

define a variable  $\beta$  to check whether need to do once more edge attaching, since the  $E_{loop}$  may be not integer, but the times number of a for-loop is integer. Our model is demonstrated as the following pseudo codes:

```

Our_model(N, m, K, f, x)
{
  Initialize_stage(m);
  Eloop = (K - m) / 2(N - (m+1));
  For i = (m + 1) to N
    Growing_stage(m, Eloop, i, f, x);
  End_For
}

Growing_stage(m, Eloop, i, f, x)
{
  First_Attach(i);
  Extra_m = Eloop - m;
   $\beta$  = Extra_m;
  For j = 2 to m
    Attach_edge(f, x);
    If ( $\beta \geq 1$ )
      Attach_edge(f, x);
       $\beta$  =  $\beta$  - 1;
    Else
       $\beta$  =  $\beta$  + Extra_m;
  End_For
}

Attach_edge(f, x)
{
  if (!Is_attach(f))
    If (!Is_attach(x)) RF;
    Else SF;
  ElseIf (!FF)
    If (!Is_attach(x)) RF;
    Else SF;
}

```

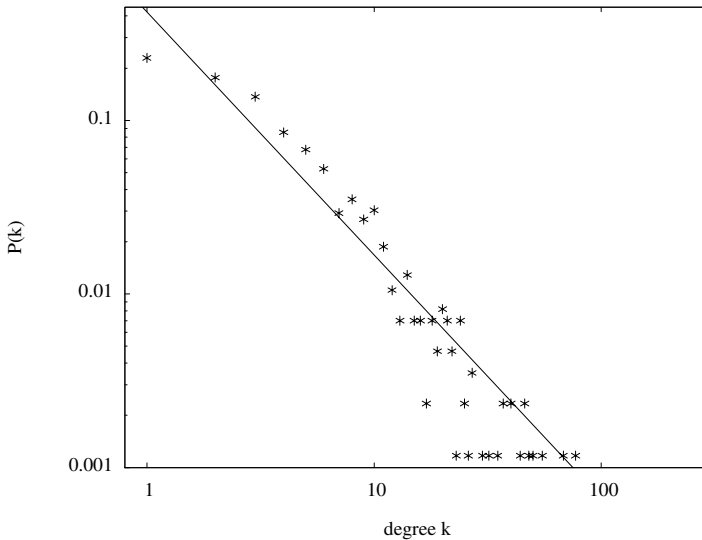
## 5 Simulation

We used our model to simulate the social network for comparison with the OAC network experimental results. The results of simulation are shown in Table 2. In this table, some cells are colored in gray to indicate that their values are 5% smaller or bigger than those in empirical data.

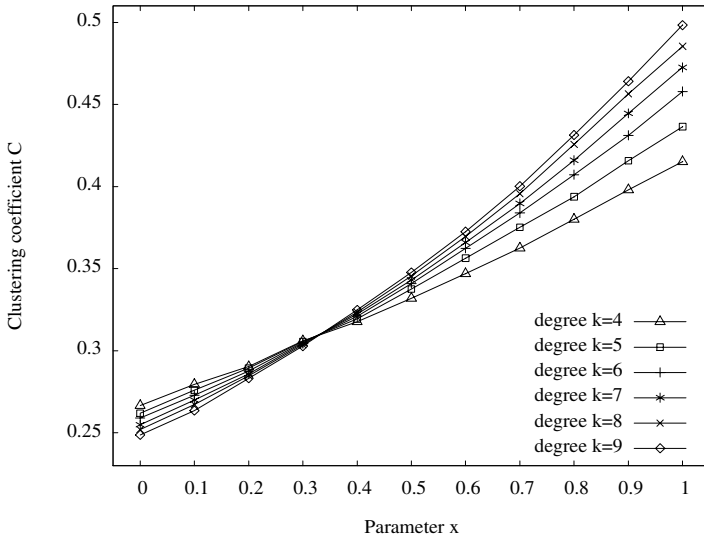
In the simulation of the OAC, we used parameters  $f=0.10$  to  $0.25$  and  $x=0.05$  to  $0.30$ . In clustering coefficient  $C$ , the suitable  $f$  is  $0.15$ , and  $x$  is from  $0.15$  to  $0.30$ . In condensed clustering coefficient  $C'$ , the suitable  $f$  is  $0.20$ , and  $x$  is from  $0.05$  to  $0.20$ . In shortest path  $L$  and s-metric  $S$ , all parameters are suitable. The tail index  $\gamma$  of the OAC is  $1.439$ , but our model only generates  $1.52$  to  $1.68$  under the parameters. The one degree in the network  $O$  of the OAC is  $0.339$ , but our model only generates  $0.19$

**Table 2.** The simulation results with the OAC

	OAC	Simulation	$x=0.05$	$x=0.10$	$x=0.15$	$x=0.20$	$x=0.25$	$x=0.30$
$L$	3.574	$f=0.10$	3.718	3.703	3.685	3.669	3.654	3.641
$O$	0.339		0.204	0.210	0.216	0.222	0.233	0.244
$\gamma$	1.439		1.682	1.666	1.645	1.634	1.622	1.611
$C$	0.0979		0.0671	0.0686	0.0702	0.0721	0.0739	0.0758
$C'$	0.1481		0.0843	0.0868	0.0895	0.0927	0.0963	0.1003
$S$	0.645		0.666	0.667	0.669	0.671	0.673	0.674
$L$	3.574	$f=0.15$	3.694	3.675	3.657	3.641	3.626	3.611
$O$	0.339		0.198	0.208	0.214	0.220	0.230	0.241
$\gamma$	1.439		1.654	1.643	1.629	1.605	1.582	1.570
$C$	0.0979		0.0905	0.0923	0.0944	0.0975	0.100	0.102
$C'$	0.1481		0.1128	0.1165	0.1201	0.1250	0.1299	0.1344
$S$	0.645		0.652	0.653	0.654	0.656	0.658	0.659
$L$	3.574	$f=0.20$	3.669	3.641	3.623	3.604	3.588	3.573
$O$	0.339		0.197	0.204	0.208	0.215	0.222	0.230
$\gamma$	1.439		1.650	1.621	1.593	1.578	1.569	1.561
$C$	0.0979		0.115	0.117	0.120	0.122	0.124	0.126
$C'$	0.1481		0.1432	0.1470	0.1515	0.1554	0.1594	0.1636
$S$	0.645		0.644	0.644	0.645	0.646	0.646	0.647
$L$	3.574	$f=0.25$	3.636	3.614	3.593	3.573	3.554	3.535
$O$	0.339		0.192	0.200	0.207	0.211	0.216	0.222
$\gamma$	1.439		1.621	1.601	1.572	1.556	1.541	1.526
$C$	0.0979		0.143	0.146	0.149	0.151	0.153	0.155
$C'$	0.1481		0.1770	0.1825	0.1879	0.1914	0.1952	0.1992
$S$	0.645		0.630	0.630	0.631	0.634	0.637	0.640



**Fig. 2.** The log-log degree distribution plot of our model with node  $N=855$ , degree  $k=5.691$ , and parameter  $f=0.15$ ,  $x=0.20$ . The slope of the line is  $-1.605$ .



**Fig. 3.** The clustering coefficient  $C$  with node  $N=2000$ , degree  $k=4$  to  $9$ , parameter  $f=0.5$ , and parameter  $x=0$  to  $1$

to  $0.24$  under the parameters. The reason is that we use only two parameters to tune our social network, and that is a little difficult to control six characteristics. In this simulation, we choice  $f=0.15$  and  $x=0.20$  to plot the degree distribution as Fig. 2, and it demonstrates scale-free property.

Furthermore, we used other parameters to simulate the social network. The parameters in Fig. 3, 4, 5, 6 and 7 are node  $N=2000$ , degree  $k=4$  to  $9$ , and parameter  $f=0.5$ . Fig. 3 shows the clustering coefficient relation with parameter  $x=0$  to  $1$ , and when  $x$  is large,  $C$  is large. The reason is that, when  $x$  is large, there will generate more high-degree nodes. The more high-degree nodes cause higher clustering coefficient since the neighbor nodes of high-degree nodes maybe become their neighbors' neighbor nodes.

Fig. 4 demonstrates the tail index relationship with the same parameters of Fig.3. The more  $x$  makes, the less  $\gamma$  is. The reason is that, the large  $x$  is, the larger hub-nodes become. The more large hub-nodes make, the less tail index  $\gamma$  is. Fig. 5 shows the s-metric relationship with the same parameters of Fig. 3, and this is absolutely for large  $x$  made large  $S$ . Fig. 6 displays the one degree in the network relationship with the same parameters of Fig.3. The more  $x$  makes, the more  $O$  is. Since the more  $x$  makes more hub-nodes, and cause the other nodes have less degree, and the more  $O$  is. Fig. 7 demonstrates the condensed clustering coefficient relationship with the same parameters of Fig. 3. This figure is like Fig. 3, and then divides by  $(1-O)$ . That presents the clustering coefficient eliminates the nodes which only have one degree.



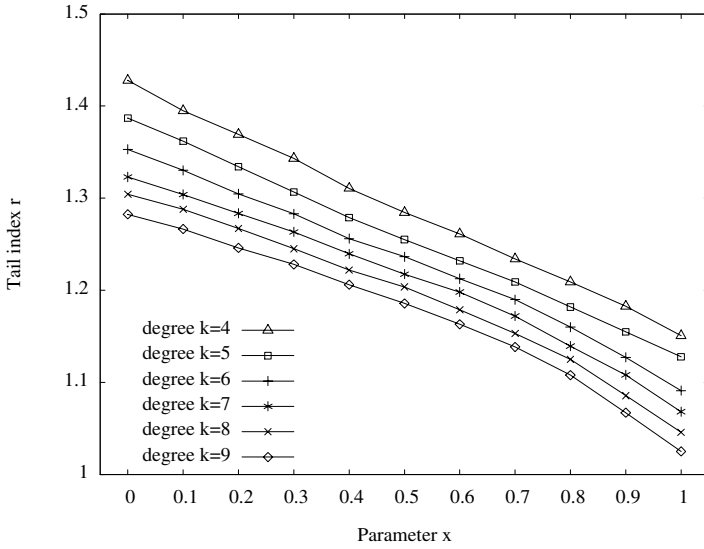


Fig. 4. The tail index  $\gamma$  with node  $N=2000$ , degree  $k=4$  to 9, parameter  $f=0.5$ , and parameter  $x=0$  to 1

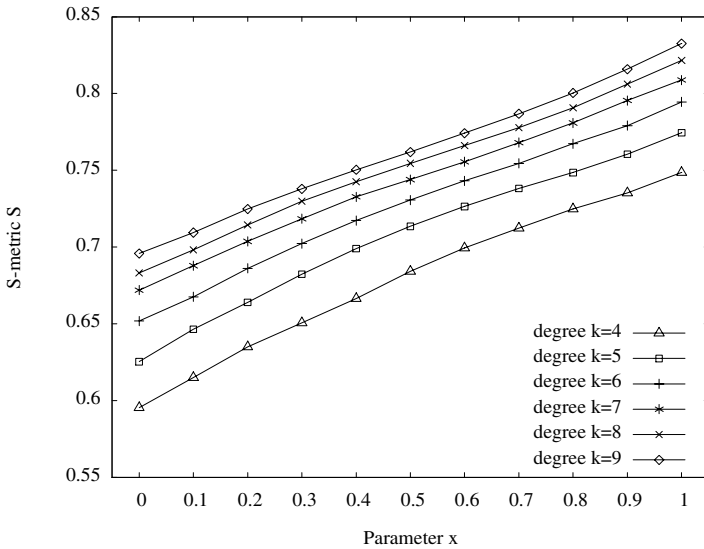
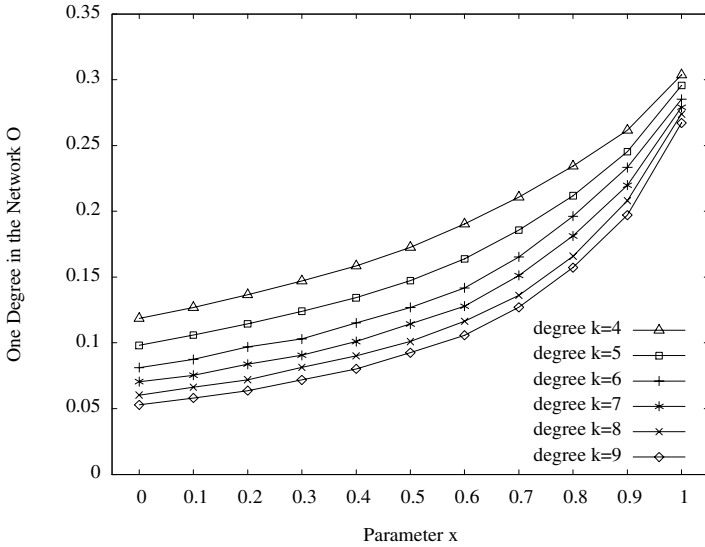
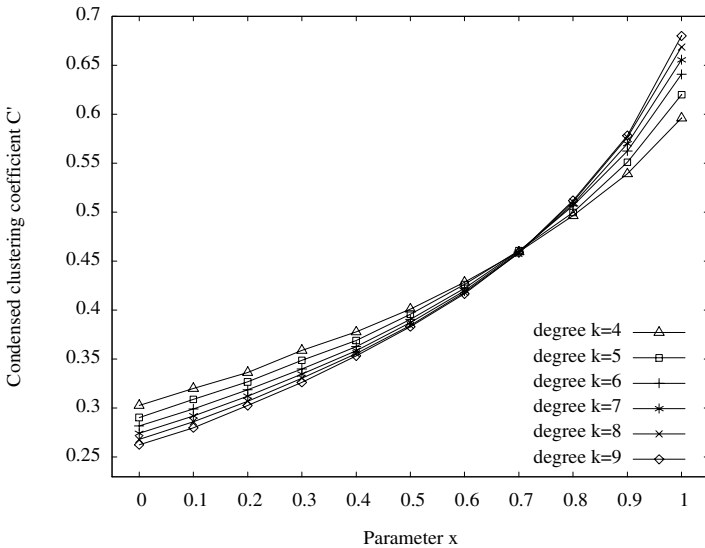


Fig. 5. The s-metric  $S$  with node  $N=2000$ , degree  $k=4$  to 9, parameter  $f=0.5$ , and parameter  $x=0$  to 1



**Fig. 6.** The one degree in the network  $O$  with node  $N=2000$ , degree  $k=4$  to  $9$ , parameter  $f=0.5$ , and parameter  $x=0$  to  $1$



**Fig. 7.** The condensed clustering coefficient  $C'$  with node  $N=2000$ , degree  $k=4$  to  $9$ , parameter  $f=0.5$ , and parameter  $x=0$  to  $1$

## 6 Conclusion and Future Work

In this paper we proposed a social network model based on topology vision, and try to define a new characteristic which is condensed clustering coefficient. The condensed clustering coefficient did not include the nodes which only have one degree, since their value of clustering coefficient is zero.

The social network analysis can be used in business for improving knowledge creation and sharing, and city planning for designing spaces to support human interaction [31] [32]. Besides, it can also be used for the simulation of epidemiological transmission [17] [18], the rumor spreading [19] [20], opinion formation [21], and computer virus dissemination [22]. Our model is compared with the OAC network which is formed by forum, and fit it approximately. Therefore, we think that our model is suitable in simulation of rumor spreading and opinion formation.

Using social network model to simulate epidemiological disease spreading allows us to predict its damage under different measures. We believe that this study constitutes a valuable contribution to understanding the phenomena of social network.

## References

1. Amaral, L.A.N., Scala, A., Barthélemy, M., Stanley, H.E.: Classes of Small-World Networks. *Proceedings of the National Academy of Sciences* 97(21), 11149–11152 (2000)
2. Newman, M.E.J.: Clustering and Preferential Attachment in Growing Networks. *Physical Review E* 64, 025102 (2001)
3. Liljeros, F., Edling, C.R., Amaral, L.A.N., Stanley, H.E., Åberg, Y.: The Web of Human Sexual Contacts. *Nature* 411, 907–908 (2001)
4. Newman, M.E.J.: The Structure of Scientific Collaboration Networks. *Proceedings of the National Academy of Sciences* 98(2), 404–409 (2001)
5. Watts, D.J., Strogatz, S.H.: Collective Dynamics of ‘Small-World’ Networks. *Nature* 393, 440–442 (1998)
6. Barabási, A.-L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., Vicsek, T.: Evolution of the Social Network of Scientific Collaborations. *Physica A* 311, 590–614 (2002)
7. Klemm, K., Eguiluz, V.M.: Highly Clustered Scale-Free Networks. *Physical Review E* 65, 036123 (2002)
8. Holme, P., Kim, B.J.: Growing Scale-Free Networks with Tunable Clustering. *Physical Review E* 65, 026107 (2002)
9. Klemm, K., Eguiluz, V.M.: Growing Scale-Free Networks with Small-World Behavior. *Physical Review E* 65, 057102 (2002)
10. Warren, C.P., Sander, L.M., Sokolov, I.M.: Geography in a Scale-Free Network Model. *Physical Review E* 66, 056105 (2002)
11. Csanyi, G., Szendrői, B.: Structure of a Large Social Network. *Physical Review E* 69, 036131 (2004)
12. Boguñá, M., Pastor-Satorras, R., Díaz-Guilera, A., Arenas, A.: Models of Social Networks Based on Social Distance Attachment. *Physical Review E* 70, 56122 (2004)
13. Wang, B., Tang, H., Zhang, Z., Xiu, Z.: Evolving Scale-Free Network Model with Tunable Clustering. *International Journal of Modern Physics B* 19(26), 3951–3959 (2005)

14. Wong, L.H., Pattison, P., Robins, G.: A Spatial Model for Social Networks. *Physica A* 360, 99–120 (2006)
15. Toivonen, R., Onnela, J.-P., Saramäki, J., Hyvönen, J., Kaski, K.: A Model for Social Networks. *Physica A* 371, 851–860 (2006)
16. Tsai, Y., Lin, C.-C., Hsiao, P.-N.: Modeling Email Communications. *IEICE Transactions on Information and Systems* E87-D(6), 1438–1445 (2004)
17. Anderson, R.M., Fraser, C., Ghani, A.C., Donnelly, C.A., Riley, S., Ferguson, N.M., Leung, G.M., Lam, T.H., Hedley, A.J.: Epidemiology, Transmission Dynamics and Control of SARS: the 2002–2003 Epidemic. *Philosophical Transactions: Biological Sciences* 359, 1091–1105 (2004)
18. Meyers, L.A., Pourbohloul, B.P., Newman, M.E.J., Skowronski, D.M., Brunham, R.C.: Network Theory and SARS: Predicting Outbreak Diversity. *Journal of Theoretical Biology* 232, 71–81 (2005)
19. Moreno, Y., Nekovee, M., Pacheco, A.F.: Dynamics of Rumor Spreading in Complex Networks. *Physical Review E* 69, 066130 (2004)
20. Lind, P.G., da Silva, L.R., Andrade, J.S., Herrmann, H.J.: The Spread of Gossip in American Schools. *Europhysics Letters* 78, 68005 (2007)
21. Kacperski, K., Hoyst, J.A.: Phase Transitions as a Persistent Feature of Groups with Leaders in Models of Opinion Formation. *Physica A* 287, 631–643 (2000)
22. Newman, M.E.J., Forrest, S., Balthrop, J.: Email Networks and the Spread of Computer Viruses. *Physical Review E* 66, 035101 (2002)
23. Crovella, M.E., Taqqu, M.S., Bestavros, A.: Heavy-Tailed Probability Distributions in the World Wide Web. In: *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, pp. 3–25 (1998)
24. Albert, R., Jeong, H., Barabási, A.-L.: Internet: Diameter of the World-Wide Web. *Nature* 401, 130–131 (1999)
25. Tsai, Y., Lin, C.-C., Hsiao, P.-N.: Heavy Tail Distribution in Email Network. In: *Proceedings of the 2002 SoftCOM*, pp. 167–170 (2002)
26. Ebel, H., Mielsch, L.-I., Bornholdt, S.: Scale-Free Topology of E-mail Networks. *Physical Review E* 66, 035103 (2002)
27. Albert, R., Barabási, A.-L.: Statistical Mechanics of Complex Networks. *Reviews of Modern Physics* 74, 48–97 (2002)
28. The Opel Astra Club (2008), <http://oac.idv.tw>
29. Cassar, A.: Coordination and Cooperation in Local, Random and Small World Networks: Experimental Evidence. *Games and Economic Behavior* 58(2), 209–230 (2007)
30. Zhou, H.: Scaling Exponents and Clustering Coefficients of a Growing Random Network. *Physical Review E* 66, 016125 (2002)
31. Huysman, M., Wulf, V.: IT to Support Knowledge Sharing in Communities, Towards a Social Capital Analysis. *Journal of Information Technology* 21, 40–51 (2006)
32. Cross, R.L., Parker, A., Borgatti, S.P.: A Bird's-Eye View: Using Social Network Analysis to Improve Knowledge Creation and Sharing. *Knowledge Directions* 2(1), 48–61 (2000)