International Journal of Bifurcation and Chaos, Vol. 18, No. 11 (2008) 3495–3502

GENERATING AN ASSORTATIVE NETWORK WITH A GIVEN DEGREE DISTRIBUTION

JIN ZHOU*, XIAOKE XU, JIE ZHANG,
JUNFENG SUN and MICHAEL SMALL
Department of Electronic and Information Engineering,
Hong Kong Polytechnic University,
Hung Hom, Kowloon, Hong Kong

JUN-AN LU
**College of Mathematics and Statistics, Wuhan University,*
Wuhan 430072, P. R. China

Received January 9, 2008; Revised March 21, 2008

Recently, the assortative mixing of complex networks has received much attention partly because of its significance in various social networks. In this paper, a new scheme to generate an assortative growth network with given degree distribution is presented using a Monte Carlo sampling method. Since the degrees of a great number of real-life networks obey either power-law or Poisson distribution, we employ these two distributions to grow our models. The models generated by this method exhibit interesting characteristics such as high average path length, high clustering coefficient and strong rich-club effects.

Keywords: Complex network; degree distribution; assortativity; clustering coefficient; average path length; rich-club.

1. Introduction

Nowadays, examples of complex networks have been found in every corner of the world, such as the Internet, World Wide Web, communication networks, power grid networks, social networks, genetic regulatory networks, and so on. Over the past two decades, complex networks have been intensively studied in various disciplines. The abundant literature has focused on exploring complex networks from many different angles including modeling networks in real life [Small & Tse, 2005; Small *et al.*, 2007; Catanzaro *et al.*, 2004] analyzing nonlinear time series with a complex network transformation [Zhang & Small, 2006], identifying topology from existing dynamical networks [Zhou & Lu, 2007] and controlling complex networks to reach synchronization [Zhou *et al.*, 2006].

Many observable real-life networks possess an important characteristic of "assortative mixing" or "disassortative mixing" on degree. "Assortative mixing" on degree means that, high-degree vertices are likely to become associated with other high-degree ones, and low-degree vertices would tend to be linked with other low-degree vertices, while "disassortative mixing" means that low-degree vertices are more likely to be connected with high-degree ones, and vice versa.

There have been a few pieces of work on modeling complex networks by assortativity. Newman [2003a] established a scheme to generate a random network having a particular value of the matrix $\mathbf{E}$, whose element e_{jk} is the fraction of edges that connect vertices of degree j and k; Catanzaro *et al.* [2004] presented a network growth model for social networks by preferential attachment and assortative attachment; Xulvi-Brunet and Sokolov [2004] focused on enhancing assortativity of an existing scale-free network through rewiring two links between four endpoints without changing the degree of all the nodes. Our new model presented in this paper, however, is a growing scale-free network generated as assortatively as possible based on a greedy algorithm at a specific degree of distribution.

The paper is organized as follows. Section 2 describes the feature of assortativity and the algorithm for generating the assortative network with a given degree of distribution. The models generated using the given power-law or Poisson degree distributions are introduced in Sec. 3; meanwhile, the key properties, including assortativity, average path length, clustering coefficient, and rich-club characteristics are detailed in this section. Finally, the conclusions are made in Sec. 4.

2. The Algorithm for an Assortative Network with a Given Degree of Distribution

2.1. *Assortativity*

Assortative mixing can have a profound effect on the properties of a network. A distinct pattern of assortative mixing depends on information concerning individual network nodes. The widely used approach is to consider assortative mixing by degree, which allows for connections between nodes that have similar degrees. Counterparts of this scheme can be found in social networks. For example, film actors are likely to be collaborators with those that are equally prolific, or a new student is likely to become friends with others who are similarly solitary.

For convenience, only an undirected network is considered in this paper. As mentioned in the previous section, the matrix $\mathbf{E} = (e_{ij})$ is an important quantity to characterize assortative mixing. Denote the fraction of edges in the entire network that is associated with i-type nodes (typically this will mean precisely that their degrees are i) as a_i. It is obvious that

$$\sum_j e_{ij} = \sum_i e_{ij} = a_i.$$

Newman defined an assortativity coefficient using $\mathbf{E}$ and a_i in [Newman, 2003a] by

$$r = \frac{\sum_i e_{ii} - \sum_i a_i^2}{1 - \sum_i a_i^2} = \frac{\mathrm{Tr}\,\mathbf{E} - \|\mathbf{E}^2\|}{1 - \|\mathbf{E}^2\|}, \qquad (1)$$

where $\mathrm{Tr}\,\mathbf{E}$ means the trajectory of matrix $\mathbf{E}$, $\|\cdot\|$ represents 2-norm of a matrix. It can be found that if there are no assortative features, say in Erdös–Rényi (ER) random network [Erdös & Rényi, 1960], we have $r = 0$ since $e_{ii} = a_i^2$. If $r > 0$, then the network is assortative, while $r < 0$ represents a disassortative mixing network.

If we denote $P(k)$ as the probability that a randomly chosen node whose degree is k, the excess degree of the node at the end of an edge is distributed according to Newman [2001]

$$q(k) = \frac{(k+1)P(k+1)}{\langle k \rangle},$$

where $\langle k \rangle$ is the mean degree of the network. It can be found that $\sum_j e_{jk} = q(k)$. Then we get another form of assortative coefficient as follows [Newman, 2002]

$$r = \frac{\sum_{ij} jk(e_{jk} - q(j)q(k))}{\sigma_q^2}, \qquad (2)$$

where σ_q is the standard deviation of the distribution $q(k)$. For the practical purpose of calculating the assortativity coefficient on an actual network, we can use the following form [Newman, 2002]

$$r = \frac{M^{-1}\sum_i j_i k_i - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}, \qquad (3)$$

where j_i and k_i are the degrees of the two endpoints of ith edge, M is the total number of edges in the network.

2.2. *The algorithm*

In this part, we present the new scheme using Monte Carlo sampling method for modeling an assortative network with a given degree distribution. First we determine the network size N, and establish a degree distribution $P(k)$ for the whole network.

A fully connected network with size n_0 is employed to initiate the model. At each step t ($t = 1, 2, \ldots, N - n_0$), introduce a new node into the network, and choose its degree from the set $\Omega = \{k | 1 \le k \le \min\{n_0 + t - 1, k_{\max}\}\}$ randomly, where $k_{\max}$ is the critical maximum degree satisfying $NP(k) \ge 1$. Before adding a new node to the existing network, judge whether the degree to be assigned to the new node satisfies $N_k \leq NP(k)$, where N_k is the number of nodes whose degrees are k. If it is not the case, generate another degree from the set Ω and repeat the above procedure. Then randomly select k existing nodes to be connected to the new node as assortatively as possible, provided that those vertices whose degrees are about to increase satisfy $N_{k+1} \leq NP(k+1)$. The assortative connecting means that the new node is linked to those nodes that have the same degree. If no such vertices are available, then the new node will attach to those whose degree is larger or smaller by one. If, again, this condition cannot be met, connect it to those with degree larger or smaller by two, etc. Once the new vertex has been successfully added into the existing network, proceed to step $t+1$.

2.3. *Discussion*

In the assortatively growing process, those present nodes to be linked to the new node are chosen randomly. In the real world, however, this is not always the case. Take the social network for example, when a newcomer chooses his friends in a community, some certain rules may often be obeyed such as the "oldest preferential attachment" or the "newest preferential attachment". That is, a newcomer may be more willing to make friends with those senior peers or he may prefer to be affiliated with more junior colleagues. These ideas can be illustrated with a counterpart in social collaboration network, where there is a strict hierarchical social structure and an assortative number of collaborators within different layers. For example, in a certain community, those senior individuals may have priority when obtaining collaborators. While in a vibrant cooperative system, newcomers may be more willing to cooperate with other neophytes.

3. Properties of the Presented Models

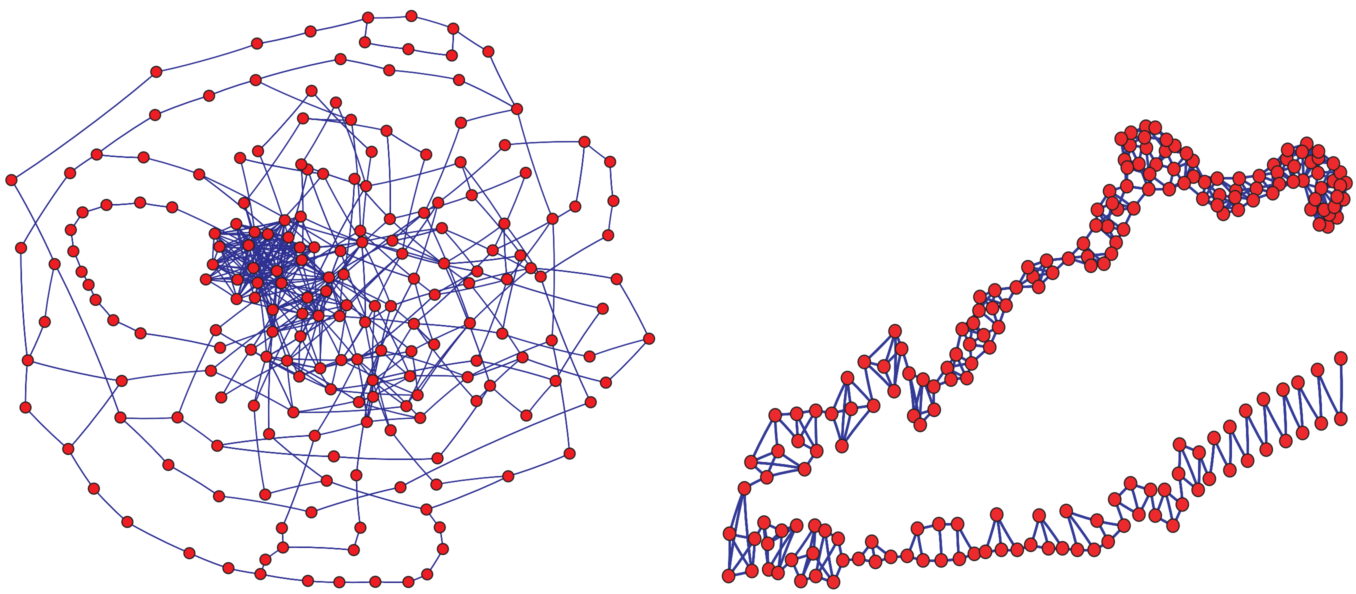
3.1. *Degree distribution*

Barabási and Albert (BA) introduced a growing network model [Barabási & Albert, 1999] of preferential attachment to nodes with already high degree in 1999. This model naturally gives rise to hubs with a degree distribution following a power-law. This kind of network has been called scale-free thereafter. It has been proved that numerous practical networks are scale-free, namely the degrees of many real-world networks obey power-law distribution. On the other hand, small-world network and random network have also been proposed to characterize actual networks. Among them, the degrees of the Watts–Strogtz (WS) small-world network [Watts & Strogatz, 1998], the Newman–Watts (NW) small-world network [Newman & Watts, 1999], and the classic Erdös–Rényi (ER) random network [Erdös & Rényi, 1960] obey Poisson distribution [Barrat & Weigt, 2000; Bollobás, 2001]. Therefore, power-law and Poisson distributions are the two most typical degree distributions in observable networks.

Hence, in our simulations we focus on networks with both power-law and Poisson degree distributions. The denotations SFO, SFN, SFR, PO, PN and PR are the abbreviation of the scale-free oldest-preferential assortative network, the scale-free newest-preferential assortative network, the scale-free assortative network with randomly chosen nodes to be connected to the new node, the oldest-preferential assortative network with Poisson degree distribution, the newest-preferential assortative network with Poisson degree distribution, and the assortative network with Poisson degree distribution as well as randomly chosen nodes to be connected to the new node. The assortative networks to be generated with degree distributions $P(k) = (N/(N-1))k^{-2}$ and $P(k) = \langle k \rangle^k e^{-\langle k \rangle}/k!$ are depicted in Fig. 1, in which the existing nodes to be associated with the new node are chosen randomly, where $m_0 = 6, N = 200$. It should be noted that in all the simulations $\langle k \rangle = (N/(N-1)) \sum_{k=1}^{k_{\max}^{\rm SF}} (1/k)$ in the Poisson degree distribution network is set the same as that in the scale-free network, where $k_{\max}^{\rm SF}$ represents $k_{\max}$ in the scale-free network. It can be seen that their degrees obey strict power-law or Poisson distribution in Fig. 2.

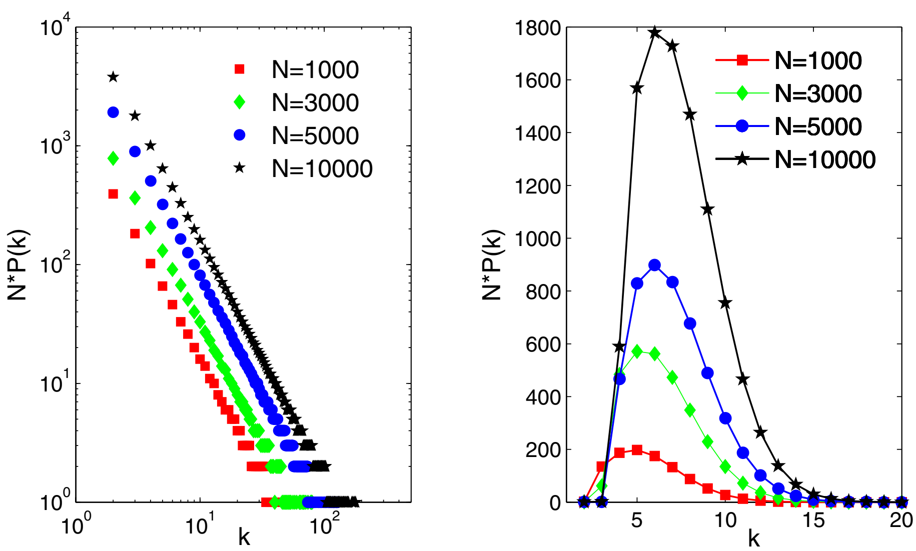
3.2. *Assortativity coefficient*

Using formula (3), the assortativity coefficients of the SFO, SFN, SFR, PO, PN and PR networks are depicted in Fig. 3. We find that all of them are much larger than that of the BA scale-free network and the ER random network, whose assortativity coefficients have the same value $r = 0$ [Newman, 2003b]. All the assortativity coefficients tend to grow with the network size because of the assortative mixing mechanism. The reason that the assortativity coefficients of the Poisson degree distribution models are larger than those of the scale-free models is



(a) (b)

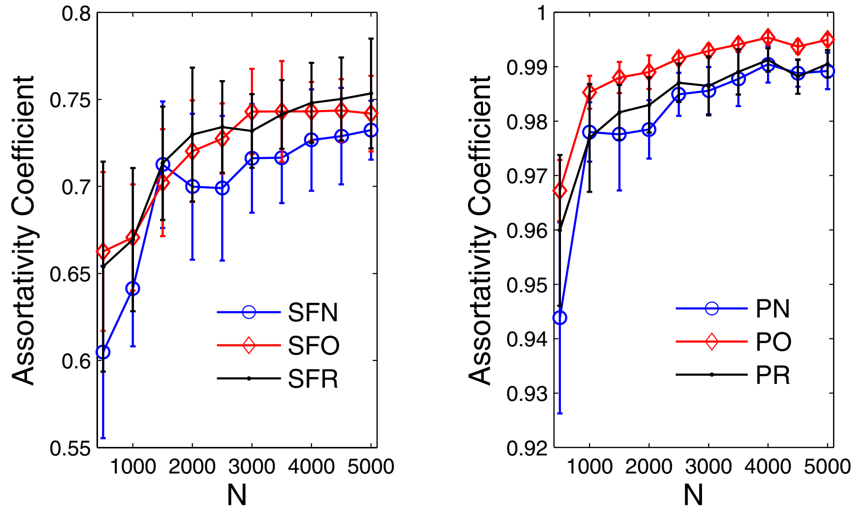
Fig. 1. The network models to be generated with degree distributions (a) $P(k) = (N/(N-1))k^{-2}$ and (b) $P(k) = \langle k \rangle^k e^{-\langle k \rangle}/k!$, in which the existing nodes to be associated with the new node are chosen randomly. Here, $m_0 = 6$, $N = 200$, and $\langle k \rangle = 4.79$ in the Poisson degree distribution network is set the same as that in the scale-free network.



(a) (b)

Fig. 2. Degree distributions of the new models to be grown with power-law (a) $P(k) = (N/(N-1))k^{-2}$ and (b) Poisson distribution $P(k) = \langle k \rangle^k e^{-\langle k \rangle}/k!$.

explained below. Note that in the scale-free models, there are a great number of vertices whose degrees are very low, say, a small integer k_0. Each of these vertices possesses just k_0 links, so under the circumstance of connectivity, many of them have to be connected to those whose degree is $k_0 + 1$, or $k_0 + 2$, etc. Then they have less chance to obtain perfectly assortative attachment (the attachment to those who have exactly the same degree). While in the Poisson degree distribution models, the degrees of most vertices are closer to the mean degree, which is much higher than k_0, hence they have more links available to get perfectly assortative mixing. This results in the much higher values of the assortativity coefficients in the Poisson degree distribution models than that in the scale-free models.



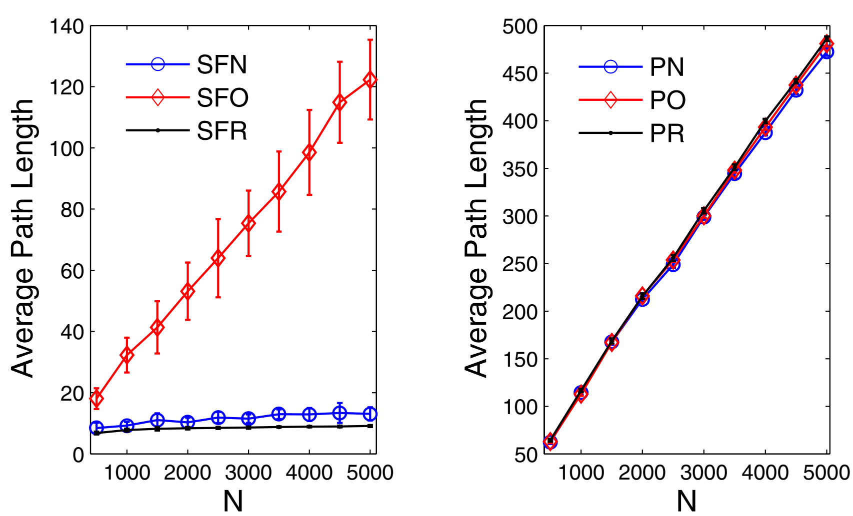
(a) (b)

Fig. 3. Assortativity coefficients of (a) the scale-free models (SFN, SFO, SFR) and (b) the Poisson degree distribution models (PN, PO, PR). All of them are much higher than that of the BA scale-free network and the ER random network ($r = 0$) [Newman, 2003b], and tend to increase as the size of the network model grows.

3.3. *Average path length*

Average path length represents the average number of steps along the shortest paths for all possible pairs of network nodes [Albert & Barabási, 2002]. It is a measure of the efficiency of information transfer or mass transport on a network and is usually denoted as L. This concept can be explained as the average number of friends you have to communicate through when you want to contact a complete stranger, or the average number of clicks which will lead you from one website to another.

As is shown in Fig. 4, the average path lengths of the SFO, PN, PO and PR networks have nearly the same tendency that they scale linearly with the network size, while the other two are much lower. A small number of long-range links guaranteed by the "oldest preferential attachment" in the scale-free models, as well as the great number of vertices whose degrees are very low, lead to the long



(a) (b)

Fig. 4. Average path lengths of (a) the scale-free models (SFN, SFO, SFR) and (b) the Poisson degree distribution models (PN, PO, PR). Clearly, the average path lengths of SFN, PN, PO, PR have nearly the same tendency that they scale linearly with the network size.

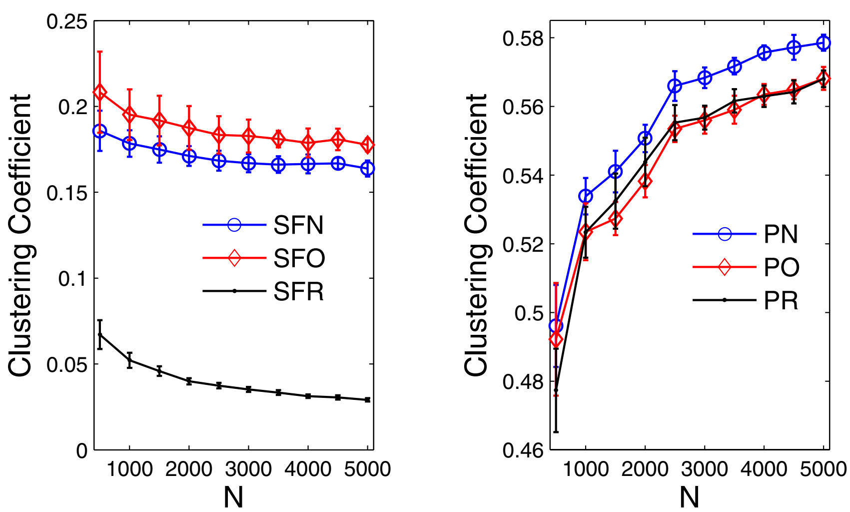
loops in the SFO network. We find that it is these long loops that bring about a large average path length in the SFO network. On the other hand, with the high assortativity coefficient, the Poisson degree distribution models possess very few long-range links as is described in Fig. 1(b), and this results in "huge" average path lengths. Since a network can be regarded as a small-world one if the average path length scales no more than logarithmically with its size N [Newman, 2002], we conclude that the SFN, PN, PO and PR networks show a "large-world" and under some circumstances even a "huge-world" effect.

3.4. *Clustering coefficient*

Clustering coefficient is one of the three most robust measures in network topology, along with the average path length and the degree distribution. Clustering coefficient for a vertex is the proportion of links between the vertices within its neighborhood divided by the number of links that could possibly exist between them. Clustering coefficient for the whole network is defined by Watts and Strogatz as the average value of the clustering coefficient for each vertex [Watts & Strogatz, 1998].

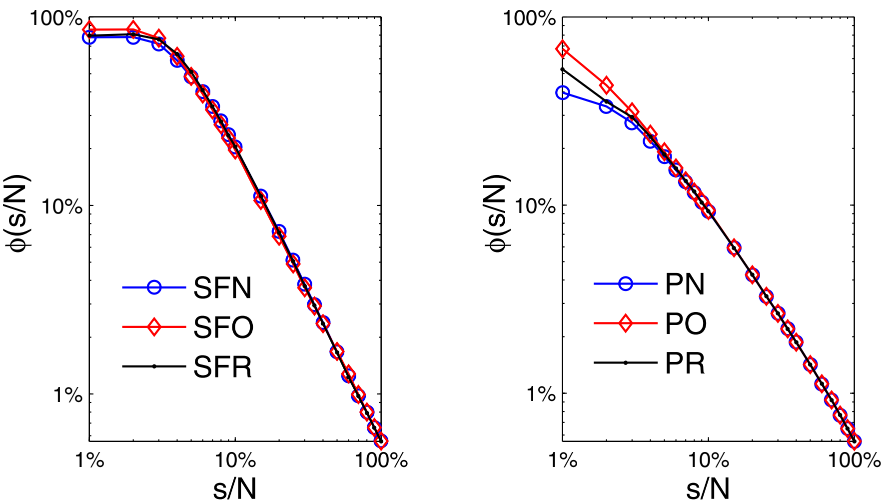
Figure 5 reveals the similarly decreasing trends of the clustering coefficients in the three scale-free models (SFN, SFO, SFR), while the values of the SFN and SFO networks are remarkably higher than that of the SFR network. This is because the "oldest" or "newest" preferential attachment guarantees that those nodes which are chosen to link a new node have a high probability to have been joined together. Furthermore, due to the fact that the set of oldest nodes which are selected to be attached to the new node with a particular degree is roughly the same and that the set of the newest nodes is dynamically changing, the clustering coefficient of the SFO network is a little higher than that of the SFN network. Besides, since the lengths of loops increase as network size expands, the clustering coefficients of the three models exhibit decreasing tendency.

On the other hand, the clustering coefficients of the three Poisson degree distribution models (PN, PO, PR) rise rapidly with the network size. This is because, for the Poisson degree distribution, the maximum degree is much smaller than that in the scale-free models, and then the degrees of most of the new nodes generated are widely distributed in the degree interval (the interval from the minimum degree to the maximum degree). As the network grows, the new node can be easily grouped into existing clusters, because its degree can easily be matched to those of existing nodes. Thus the clustering coefficient curves in Fig. 5(b) increase greatly. In addition, we find that the "oldest preferential attachment" makes the oldest nodes possess the highest degrees and connect to each other. This leads to a higher level of clustering in the high degree nodes than that in the PN network because



(a) (b)

Fig. 5. Clustering coefficients of (a) the scale-free models (SFN, SFO, SFR) and (b) the Poisson degree distribution models (PN, PO, PR). There are similar trends of the clustering coefficients in the three scale-free models and the three Poisson degree distribution models, respectively.



(a) (b)

Fig. 6. (a) Rich-club coefficients of the scale-free models (SFN, SFO, SFR) and (b) the Poisson degree distribution models (PN, PO, PR), where $N = 1000$. It is shown that the rich-club coefficients in the Poisson degree distribution models with the same degree distribution are slightly different at first and tend to be equal gradually.

the highest degree nodes are less clustered in the latter one. However, the clustering coefficients of the nodes whose degrees are not so high in the PN network are larger than that in the PO network for the "newest preferential attachment". Since there are a great number of nodes whose degrees are near the mean degree according to the Poisson degree distribution, the average value of the clustering coefficient for each node in the PN network exhibits a higher value.

3.5. *Rich-club coefficient*

Rich-club connectivity describes the characteristics that high-degree nodes tend to be connected with other high-degree nodes. When nodes in the network are sorted by decreasing number of links that each node possess, there are instances where groups of nodes contain an identical number of links. Once this occurs, they are arbitrarily assigned a position within that group. The node rank s denotes the position of a node on this ordered list, where s is normalized by the total number of nodes N. The rich-club connectivity coefficient of the s richest nodes is defined by [Zhou & Mondragón, 2004] to be

$$\Phi(s/N) = \frac{R}{s(s-1)/2}$$

where R is the actual edge between the richest most s nodes.

From Fig. 6, which is obtained by choosing the network sizes $N = 1000$, it can be seen that the assortative attachment mechanism and power-law degree distribution lead to about 3% richest vertices almost fully connected with each other and with significantly less links to other poorer vertices. While in the Poisson degree distribution models, the "oldest preferential attachment" results in a more significant rich-club property due to the fact that the highest degree nodes, namely the oldest ones, are connected to each other. The "newest preferential attachment", however, renders the highest nodes to be less connected to each other. As a result, the rich-club coefficients in the models with the same degree distribution are slightly different at small s/N and coincide at large s/N.

4. Conclusions

In summary, we have proposed a new scheme to generate an assortative growth network with a given degree distribution. Particularly, the key features of our new models with given power-law and Poisson degree distribution have been discussed in detail. In view of the difference between the two degree distributions, we have found that our models exhibit interesting characteristics. Despite possessing the same assortative mixing mechanism, the assortativity coefficients of the Poisson degree distribution models are much higher than that of the scale-free models; the average path lengths of the former are very large or even huge compared to that of the latter; the clustering coefficients show an opposite tendency in the two different degree distribution models; the rich-club coefficients of the former are different at small s/N and coincide at large s/N, while that of the latter are nearly the same.

Acknowledgments

This research was funded by Hong Kong University Grants Council Grant Competitive Earmarked Research Grant (CERG) No. PolyU 5269/06E. Ms. Zhou Jin is supported by the National Natural Science Foundation of China under Grant No. 70771084, 60574045, and the National Basic Research Program of China under Grant No. 2007CB310805.

References

Albert, R. & Barabási, A.-L. [2002] "Statistical mechanics of complex networks," *Rev. Mod. Phys.* **74**, 47–97.

Barabási, A.-L. & Albert, R. [1999] "Emergence of scaling in random networks," *Science* **286**, 509–512.

Barrat, A. & Weigt, M. [2000] "On the properties of small-world network models," *Eur. Phys. J. B* **13**, 547–560.

Bollobás, B. [2001] *Random Graphs*, 2nd edition (Academic Press, NY).

Catanzaro, M., Caldarelli, G. & Pietronero, L. [2004] "Assortative model for social networks," *Phys. Rev. E* **70**, 037101.

Erdös, P. & Rényi, A. [1960] "On the evolution of random graphs," *Science* **5**, 17–61.

Newman, M. E. J. & Watts, D. J. [1999] "Renormalization group analysis of the small-world network model," *Phys. Lett. A* **263**, 341–346.

Newman, M. E. J., Strogatz, S. H. & Watts, D. J. [2001] "Random graphs with arbitrary degree distributions and their applications," *Phys. Rev. E* **64**, 026118.

Newman, M. E. J. [2002] "Assortative mixing in networks," *Phys. Rev. Lett.* **89**, 208701.

Newman, M. E. J. [2003a] "Mixing pattern in networks," *Phys. Rev. E* **67**, 026126.

Newman, M. E. J. [2003b] "The structure and function of complex networks," *SIAM Rev.* **45**, 167–256.

Small, M. & Tse, C. K. [2005] "Small world and scale free model of transmission of SARS," *Int. J. Bifurcation and Chaos* **15**, 1745–1755.

Small, M., Walker, D. M. & Tse, C. K. [2007] "Scale-free distribution of avian influenza outbreaks," *Phys. Rev. Lett.* **99**, 188702.

Watts, D. J. & Strogatz, S. H. [1998] "Collective dynamics of 'Small-World' networks," *Nature* **393**, 440–442.

Xulvi-Brunet, R. & Sokolov, I. M. [2004] "Reshuffling scale-free networks: From random to assortative," *Phys. Rev. E* **70**, 066102.

Zhang, J. & Small, M. [2006] "Complex network from pseudoperiodic time series: Topology vs dynamics," *Phys. Rev. Lett.* **96**, 238701.

Zhou, J., Lu, J. & Lü, J. [2006] "Adaptive synchronization of an uncertain complex dynamical network," *IEEE Trans. Autom. Contr.* **51**, 652–656.

Zhou, J. & Lu, J. A. [2007] "Topology indentification of weighted complex dynamical network," *Physica A* **386**, 481–491.

Zhou, S. & Mondragón, R. J. [2004] "The rich-club phenomenon in the internet topology," *IEEE Commun. Lett.* **8**, 180–182.