IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **40** (2007) 8237–8246 doi:10.1088/1751-8113/40/29/003

The Watts–Strogatz network model developed by including degree distribution: theory and computer simulation

Y W Chen[1,2], L F Zhang[1] and J P Huang[1]

[1] Surface Physics Laboratory and Department of Physics, Fudan University, Shanghai 200433, People's Republic of China
[2] High School affiliated to Fudan University, Shanghai 200433, People's Republic of China

E-mail: jphuang@fudan.edu.cn

Received 19 February 2007, in final form 3 May 2007
Published 3 July 2007
Online at stacks.iop.org/JPhysA/40/8237

Abstract
By using theoretical analysis and computer simulations, we develop the Watts–Strogatz network model by including degree distribution, in an attempt to improve the comparison between characteristic path lengths and clustering coefficients predicted by the original Watts–Strogatz network model and those of the real networks with the small-world property. Good agreement between the predictions of the theoretical analysis and those of the computer simulations has been shown. It is found that the developed Watts–Strogatz network model can fit the real small-world networks more satisfactorily. Some other interesting results are also reported by adjusting the parameters in a model degree-distribution function. The developed Watts–Strogatz network model is expected to help in the future analysis of various social problems as well as financial markets with the small-world property.

PACS numbers: 89.75.Hc, 05.50.+q, 87.23.Ge

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Network analysis is a powerful approach to analyse social problems and financial markets [1–17]. So far, one has discovered two fundamental properties of real complex networks: the small-world property (One can reach a given node from another one, following the path with the smallest number of links between the nodes, in a very small number of steps, which corresponds to the 'six degrees of separation' in social networks [3].) and the scale-free properties. (The probability distribution of the number of links per node, namely, the degree distribution, can be represented by a power law, i.e., scale free [3].) The Watts–Strogatz network model has been introduced by Watts and Strogatz [1, 18], to deal with the networks with the small-world property, namely, the characteristics of both 'high network clustering' and 'short characteristic path length' [2]. The model consists of a regular lattice, typically a one-dimensional lattice with almost periodic boundary conditions (which means that each vertex is connected to almost a fixed number of vertices nearest to it) and a small number of 'shortcut' bonds added between randomly chosen vertices.

In the description of the small-world property, 'high network clustering' means that two individuals are much more likely to be friends with one another if they have one or more other friends in common, while 'short characteristic path length' indicates that the average distance increases with the number of nodes in the networks logarithmically. Throughout the paper, we let l (or l') represent the (average) characteristic path length, C (or C') the clustering coefficient and k (or $k' = k/2$) the number of closest vertices or sub-vertices connected to each vertex or sub-vertex. In the Watts–Strogatz network model, it is proved that one can add a small number of shortcuts to reduce l in a huge amount while C almost remains the same [1]. By using actual living models that have the small-world property, one may get some important information for the characteristics of l and C. For instance, it has been found that the Internet [3] and the dissemination of diseases [3] have the small-world property so that one can improve the safety of the Internet by increasing some links between some irrelevant websites and reduce the range of dissemination of diseases only by reducing the contact between unfamiliar people. So far, there have been some theoretical methods to calculate l [2] and C [3]. After the comparison of l and C between many real networks with the small-world property and the Watts–Strogatz network model [4–8], one [3] reported some deviations of l and C. For overcoming the deviations, a crucial reason might be that in the Watts–Strogatz network model one requires each vertex to connect a fixed number of closest vertices (namely, fixed k). This may lead to the deviations, because in real small-world networks all members (also called vertices) cannot have even importance, thus yielding different number of closest members naturally (i.e., different k). We shall develop the Watts–Strogatz network model by taking into account the influence of degree distribution. Here the Watts–Strogatz network model is developed by introducing degree distribution only for the real networks with the small-world property. In doing so, our focus is on the improvement of the comparison between characteristic path lengths and clustering coefficients predicted by the original Watts–Strogatz network model and those of the real networks with the small-world property. Thus, the developed Watts–Strogatz network model will become more flexible due to the introduction of degree distribution. Based on the theory put forth in [2, 3], we shall develop some theoretical methods to calculate the characteristic path length l' and the clustering coefficient C', which will be shown to agree with our computer simulations. Finally, favourable agreement will also be shown between the Watts–Strogatz network model developed by including degree distribution and some real networks with the small-world property, while comparing their corresponding clustering coefficients.

2. Problems arising from some real small-world networks when elaborated by the Watts–Strogatz network model

The Watts–Strogatz network model describes networks between a regular network and a random one defined as follows. One starts from a ring lattice with n vertices and each connects to its k nearest vertices. We can enumerate every edge and use p as the probability of rewiring the edge which is called a shortcut. If p equals 0, we call the network a regular network while if p equals 1 we call it a random one. Watts and Stogatz found that for intermediate values of p, the network is a small-world network: highly clustered like a regular network, yet with a small characteristic path length, like a random network [9]. In the small-world

Table 1. Comparison between the Watts–Strogatz network model and some real networks. Here n means the number of vertices [3], k takes the average value [3], C the clustering coefficient, $C(w-s)$ the results obtained by theoretical formula (i.e., equation (2) for the Watts–Strogatz network model), C(actual) the results extracted from the real networks and C(random) the results obtained from the random network model. Through the comparison between the clustering coefficients C(actual) of the real networks and those C(random) of the relevant random network model, the real networks were shown to have some small-world network characteristics [3].

Real networks	n	k	$C(w-s)$	C(actual)	C(random)
LANL co-authorship [4, 5]	52 909	9.7	0.58	0.43	0.00 018
Math. co-authorship [6]	70 975	3.9	0.48	0.59	0.000 054
Neurosci. co-authorship [6]	209 293	11.5	0.60	0.76	0.000 055
E. coli, substrate graph [7]	282	7.35	0.54	0.32	0.026
Words, co-occurrence [8]	460 902	70.13	0.69	0.44	0.0001
C. Elegans [1]	282	14	0.62	0.28	0.005

network, the shortest characteristic path length of all vertex–vertex path is denoted as $l(p)$ and the clustering coefficient as $C(p)$.

For the Watts–Strogatz network model, one has established two theoretical formulae for calculating l [2] and C [3]

$$l_{\mathrm{original}}(\xi) = \frac{1}{2\sqrt{\xi^2+2\xi}} \tanh^{-1} \frac{\xi}{\sqrt{\xi^2+2\xi}}, \tag{1}$$

$$C(k) = \frac{3k(k-1)}{2k(2k-1)+8pk^2+4p^2k^2}, \tag{2}$$

where $\xi = npk/2$.

It has been proved that some real networks have the small-world property [3–8], so we are allowed to compare some of these small-world networks with the networks described by the Watts–Strogatz network model [3]. We use the actual data of l (in the real networks) and the formula for l (equation (1) for the Watts–Strogatz network model) to extract the relevant p. Then we use the extracted p to calculate the clustering coefficient C. Therefore, we obtain a theoretical C according to the Watts–Strogatz network model and an actual C according to the data (note the actual C denotes the real clustering coefficient of the real networks with the small-world property).

In table 1, the selected real small-world networks have their specific clustering coefficients C(actual), which, unfortunately, are quite different from those $C(w-s)$ predicted by the Watts–Strogatz network model. Such deviation might be due to the fixed number k of connections used in the Watts–Strogatz network model, because, in reality, no one can guarantee that k always keeps unchanged for a kind of specific networks. In this sense, a very reasonable assumption is that k could be changeable for different vertices in such kind of networks, as implied by the comparison in table 1. Thus, it seems both necessary and instructive to develop the Watts–Strogatz network model by introducing degree distribution.

3. The Watts–Strogatz network model developed by including degree distribution

Now we are in a position to develop the Watts–Strogatz network model by including degree distribution. Let us create a normal Watts–Strogatz network model (with many vertices, each connecting to the two nearest neighbours, so $k = 2$) and then replace each vertex by another (sub-)Watts–Strogatz network with a specific but unique k. Two of these (sub-)Watts–Strogatz networks that correspond to connected vertices are then connected by randomly

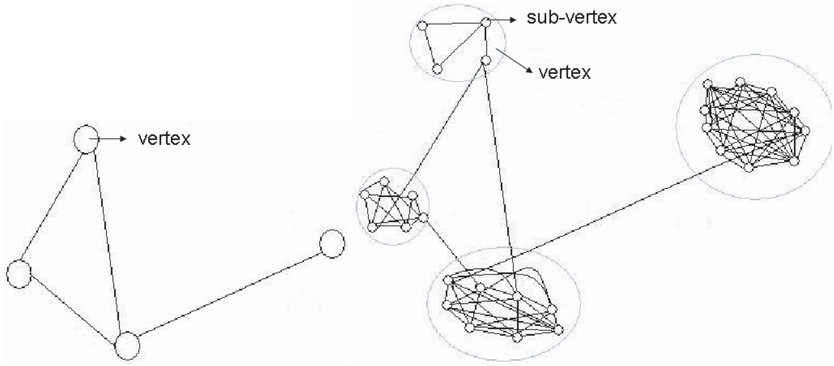

Figure 1. Left panel: schematic graph showing the vertices of the networks with the number of connections $k = 2$ (or $k' = 1$); right panel: schematic graph showing the actual vertices of the network (each vertex contains an aggregation of sub-vertices in which each sub-vertex has the same k (or k')).

choosing and connecting two sub-vertices, one in each (sub-)Watts–Strogatz network. In other words, we make a standard small-world network (left panel of figure 1), but each vertex is actually composed of a small-world sub-network (right panel of figure 1). That is, every 'vertex' only includes all actual 'sub-vertices' that have the same number of connections to closest sub-vertices (right panel of figure 1). It should be noted that the sub-vertices in different vertices have different number of connections to closest sub-vertices (right panel of figure 1). Therefore, in doing so, we can obtain small-world networks with the degree distribution. We determine the fraction of the number of sub-vertices of each k according to a probability distribution. Following the original Watts–Strogatz network model, we also use the same probability to rewire edges for all the networks in different vertices. The Watts–Strogatz network model developed by including degree distribution may fit the phenomenon of hierarchy better because it makes the sub-vertices of different importance (or with different k), so it is expected to have potential important applications in social problems. Hence we may group the nodes that have the same connections (k) together, and then make each group a small-world network. Apparently, here we also require rewiring some edges and making shortcuts between vertices, see the right panel of figure 1. For convenience, we assume these vertices to be linked weakly in order to satisfy the actual situation of the phenomenon of hierarchy. Here the vertices themselves are further seen to be linked like a small-world network, and also have shortcuts. In particular, when a vertex is linked with another vertex, this means that a sub-vertex in the vertex is connected with a sub-vertex of the other (such/sub-vertices are chosen randomly, see figure 1). To make the links between the vertices weak, we set the number of connections of one vertex to others to be 2. In the following, we shall take a distribution function to determine the fraction of the number of sub-vertices with the same k.

For the Watts–Strogatz network model developed by including degree distribution, let us derive the formulae of l' (characteristic path length) and C' (clustering coefficient). First, to be more convenient we make the distribution parameter $k' = k/2$. Hence, we obtain

$$l(k', p) = l_{\text{original}}(nf(k')pk'). \tag{3}$$

Here the distribution function should satisfy the condition

$$\sum_{k'=1}^{\max} f(k') = 1. \tag{4}$$

Let us divide the whole system by max groups (vertices). The minimum k' equals 1. And we denote the maximum k' as max. Then we can obtain the relevant l' and C' for the Watts–Strogatz network model developed by including degree distribution.

For treating l', we separate the problem into two parts. One is the connections between sub-vertices in each vertex, the other is the vertex–vertex connections. We calculate the whole characteristic path length of both two parts first and then divide it by the whole number of the paths. As a result, we obtain

$$L'_{\text{inner}} = \sum_{k'=1}^{\max} \frac{1}{2}[nf(k')][nf(k') - 1]l(k', p)nf(k'), \tag{5}$$

$$L'_{\text{outer}} = N_{\text{outer}}\left[\max l(1, p) + \sum_{k'=1}^{\max}[l(1, p)nf(k')]/\max[\max + l(1, p) + 1]\right], \tag{6}$$

$$N_{\text{outer}} = n(n-1)/2 - \sum_{k'=1}^{\max} nf(k')[nf(k') - 1]/2, \tag{7}$$

$$l' = (L'_{\text{inner}} + L'_{\text{outer}})/[n(n-1)/2]. \tag{8}$$

After we get l' we start to deal with C'. Since the connections between the vertices are very weak as defined (alternatively, the connections of sub-vertices in different vertices are very few indeed) and k is actually the minimum, we may omit the influence of vertex–vertex connections on C'. Therefore we need just calculate all sub-vertices. As a result, we obtain

$$C' = \sum_{k'=1}^{\max}[C(2k')nf(k')]/n. \tag{9}$$

4. Comparison between theory and computer simulations

To proceed, we assume the distribution function $f(k')$ takes the function for the log-normal distribution [19]:

$$f(k') = \frac{1}{\sqrt{2\pi}\sigma k'} \exp\left[-\frac{\ln^2(k'/\delta)}{2\sigma^2}\right], \tag{10}$$

where δ means the median parameter and σ the standard deviation. In real networks, σ and δ can be varied accordingly. As mentioned above, the minimum of k (or k') is set to be 2 (or 1). Here we should remark that this distribution (equation (10)) has been found to appear in various systems such as financial markets [20], sizes of ferromagnetic nanoparticles fabricated in experiments [21], the number of molecules in recursive states in replicating catalytic networks [22] and so on. (In fact, it is straightforward to extend to other systems with different sorts of distributions, e.g., Poisson distribution, Γ distribution, Gaussian distribution and so on, if necessary.) For this model, we make the average connections δ of the log-normal distribution be $k_{\text{average}}/2$. On the other hand, we perform computer simulations for comparison. In the simulations, we used 1000 sub-vertices, and the simulation data shown in figures 2 and 3 were values averaged from 10 groups of raw data. At smaller p, the characteristic path length predicted by the theory is larger than that obtained from the simulations, and for larger p, such difference becomes very small (figure 2). However, regarding the clustering coefficient, the predictions by the theory are always larger than the simulations for the full range of p, and the difference between the predictions by the theory and by the simulations is smaller at small p than at large p (figure 3). In general, figures 2 and 3 show that the developed theory agrees with

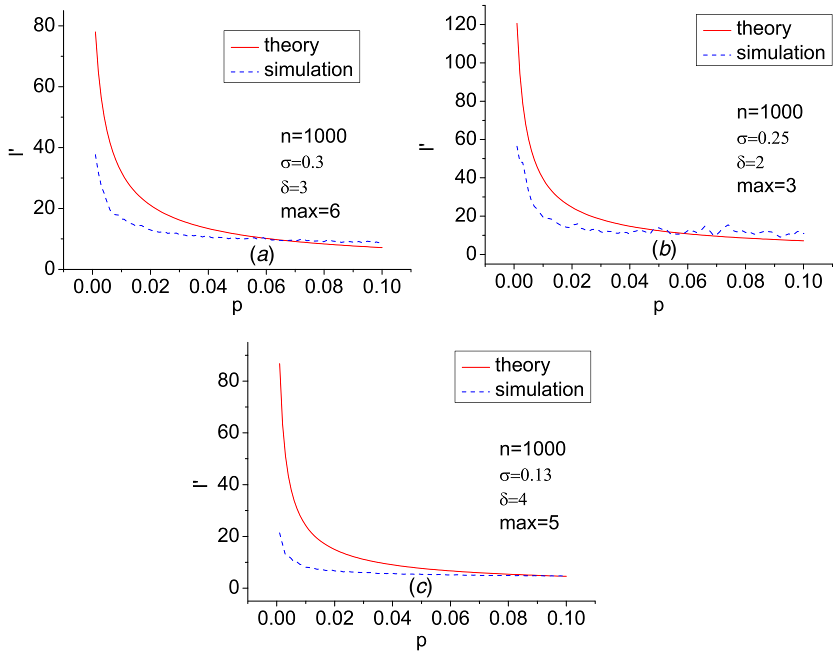

Figure 2. Average characteristic path length l' versus p for various distribution parameters, σ, δ and max, obtained by computer simulations (dashed lines) and theoretical analysis (equation (8), solid lines), respectively.

the computer simulations (at least qualitatively), which shows the validity of the theoretical formulae (8) and (9). Thus, we may convince ourselves to do the following comparisons (table 2) and numerical calculations (figures 4–6). Details will be presented in what follows.

5. Characteristics of the Watts–Strogatz network model developed by including degree distribution

Firstly, we compare the Watts–Strogatz network model developed by including degree distribution with the original Watts–Strogatz network model, see figure 4. We have set $\delta = k/2$. In the case of the distribution of the number of the connections (namely, the existence of $f(k')$), the curve l' of the Watts–Strogatz network model developed by including degree distribution can be changed, in comparison with that of l of the original Watts–Strogatz network model. In the mean time, the difference between the curve of C' and that of C is small enough to be neglected. More results obtained by adjusting the distribution parameters, σ and δ, can be found in figures 5 and 6. We investigate l' and C' of the Watts–Strogatz network model developed by including degree distribution. In detail, for a network with a fixed number of sub-vertices n, we keep max to be constant and change σ and δ in order to make the sum of the fraction of the number of sub-vertices with the same k' nearly equal to 1 (equation (4)). The results are displayed in figures 5 and 6.

Figures 5 and 6 show a framework similar to that of the Watts–Strogatz small-world networks, as already displayed in figure 4. We find that when p is small, the characteristic path length drops quickly while the clustering coefficient almost remains. Thus, the values of the ratios can be adjusted by choosing appropriate distribution parameters, σ and/or δ.

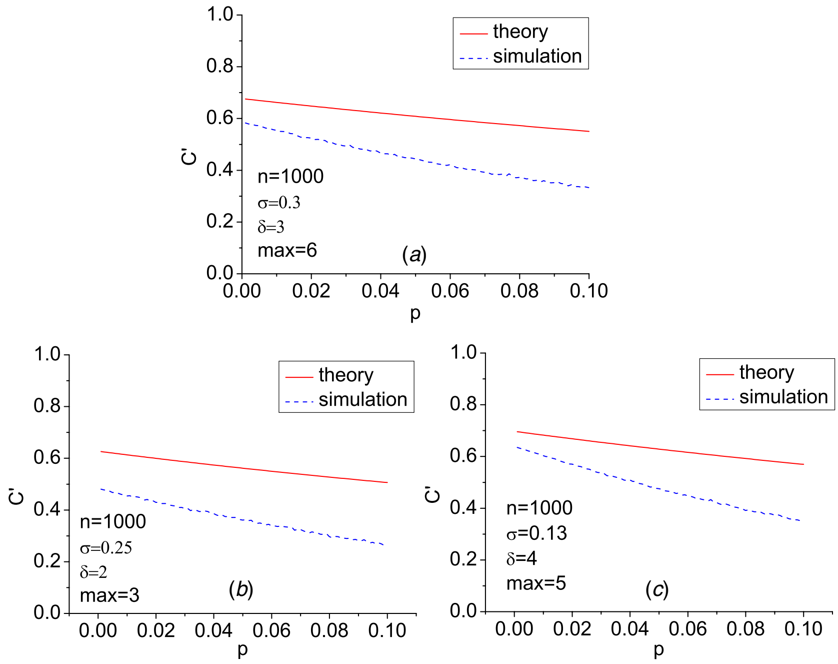

Figure 3. Clustering coefficient C' versus p for various distribution parameters, σ, δ and max, obtained by computer simulations (dashed line) and theoretical analysis (equation (9), solid lines), respectively.

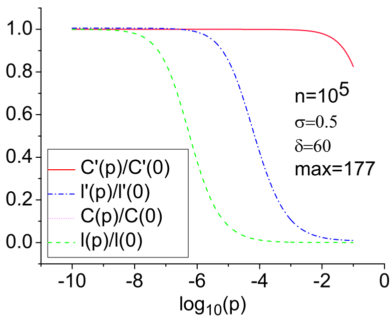

Figure 4. Ratios of characteristic path lengths $l(p)/l(0)$ (equation (1)) and $l'(p)/l'(0)$ (equation (8)) and clustering coefficients $C(p)/C(0)$ (equation (2)) and $C'(p)/C'(0)$ (equation (9)) as a function of $\log_{10}(p)$. The dashed line reflects the ratio of l of the Watts–Strogatz network model and the dash-dotted line reflects that of l' of the Watts–Strogatz network model developed by including degree distribution. The other two lines are respectively for C and C', which, however, are overlapped.

Figure 5 shows that when σ gets smaller, the framework of the curves moves left, especially at low p. Comparing with figure 4, we find that when σ gets closer to 0, the behaviour is closer to the original Watts–Strogatz network model, as expected. What is more, figure 6 displays that when σ becomes lower, the lines of C' rise up but always remain the same trend.

In short, we find that the Watts–Strogatz network model developed by including degree distribution has many characteristics similar to those of the Watts–Strogatz network model. We also find that if the standard deviation σ is very small, the whole characteristics of the

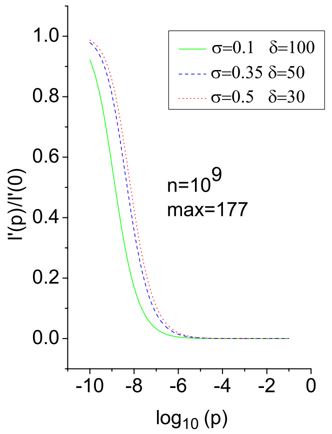

Figure 5. Ratio of characteristic path lengths $l'(p)/l'(0)$ (equation (8)) as a function of $\log_{10}(p)$ for various distribution parameters σ and δ. All the couples of σ and δ are chosen to satisfy the sum rule in equation (4): $\sigma = 0.1$ and $\delta = 100$ (solid line); $\sigma = 0.35$ and $\delta = 50$ (dashed line); $\sigma = 0.5$ and $\delta = 30$ (dotted line).

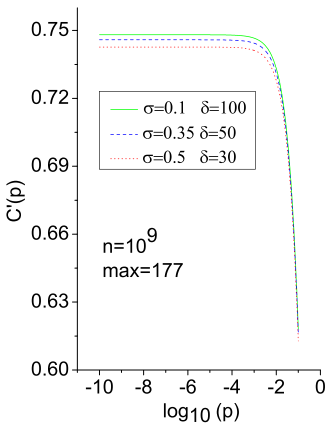

Figure 6. Clustering coefficient $C'(p)$ (equation (9)) versus $\log_{10}(p)$ for various distribution parameters σ and δ. All the couples of σ and δ are chosen to satisfy the sum rule in equation (4): $\sigma = 0.1$ and $\delta = 100$ (solid line); $\sigma = 0.35$ and $\delta = 50$ (dashed line); $\sigma = 0.5$ and $\delta = 30$ (dotted line).

Watts–Strogatz network model developed by including degree distribution are very close to those of the original Watts–Strogatz network model. Furthermore, the features of the figures can be changed accordingly by adjusting the distribution parameters. From this point of view, the Watts–Strogatz network model developed by including degree distribution seems to be more flexible than the original Watts–Strogatz network model, as expected.

Because of the flexibility of the Watts–Strogatz network model developed by including degree distribution, it is expected to have more applications. In the developed Watts–Strogatz network model, the sub-vertices are assigned into different groups (vertices) with different k', which seems more suitable for social situations since the phenomenon of a hierarchical society is common in life. In other words, such hierarchical phenomena may be expected to be treated by using the developed Watts–Strogatz network model.

Table 2. Comparison between the Watts–Strogatz network model developed by including degree distribution and real networks with the small-world property. Here $C(w-s)$ means the results obtained from the original Watts–Strogatz network model (equation (2)) and C'(theoretical) the results obtained from the formula (equation (9)) of the developed Watts–Strogatz network model.

Real networks	Max	σ	C'(theoretical)	$C(w-s)$	C(actual)
LANL co-authorship [4, 5]	13	0.4	0.48	0.58	0.43
Math. co-authorship [6]	44	0.54	0.54	0.48	0.59
Neurosci. co-authorship [6]	50	0.6	0.63	0.60	0.76
E. coli, substrate graph [7]	15	0.7	0.42	0.54	0.32
Words, co-occurrence [8]	290	1.1	0.47	0.69	0.44
C. Elegans [1]	50	1.3	0.21	0.62	0.28

In view of the above analysis of the characteristics, we are now in a position to compare the Watts–Strogatz network model developed by including degree distribution with real networks with the small-world property, as shown in table 2. For this comparison, we set $\delta = k(\text{actual})/2$. And we also use $l(\text{actual})$ to calculate p and then compare the theoretical $C'(\text{theoretical})$ and the actual $C(\text{actual})$. Table 2 shows that some model real networks with the small-world property, which do not fit well in the Watts–Strogatz network model (as represented by $C(w-s)$ herein), can fit satisfactorily in the Watts–Strogatz network model developed by including degree distribution (i.e., $C'(\text{theoretical})$). In detail, $C'(\text{theoretical})$ obtained for the developed Watts–Strogatz network model is compared favourably with $C(\text{actual})$ of the real small-world networks. Such good agreement shown in table 2 shows that the developed network model has its reasonable and physical factors, namely, the introduction of degree distribution.

6. Discussion and conclusion

Here some comments are in order. It is known that a large number of real networks are referred to as scale free because they show a power-law degree distribution. For various degree distributions including power law, it is no doubt that degree distribution functions $P(k)$ should always satisfy the sum rule $\sum_k P(k) = 1$ (equation (4)). For scale-free networks, $P(k)$ is given by the known expression $P(k) = k^{-\gamma}$, with degree exponent γ satisfying $2 < \gamma < 3$ [23]. In our work, while the detailed expression for $P(k)$ is a variant depending on various types of real networks with the small-world property (which corresponds to the fact that $P(k)$ can be varied in different networks, as also implied by the scale-free network model), it is given by a certain distribution function, which can be any reasonable distribution functions (that satisfy the sum rule) corresponding to real networks with the small-world property. Thus, the Watts–Strogatz network model developed by including degree distribution seems more flexible to explain real networks with the small-world property whereas other existing network models like the scale-free network model have not yet been clearly shown to work for this purpose. It is worth noting that the developed Watts–Strogatz network model is presented only for the real networks with the small-world property, to improve the comparison between characteristic path lengths and clustering coefficients predicted by the original Watts–Strogatz network model and those of the real networks with the small-world property. For model calculations, we take the distribution function to be the well-known log-normal distribution, simply because this distribution function is common in nature. However, whichever distribution function should be used for a degree distribution depends on specific networks concerned. In this direction, the developed Watts–Strogatz network model can help, as expected.

In the developed Watts–Strogatz network model, we have used a predefined degree distribution to demonstrate the virtues of the developed model in reproducing features of real-world networks with the small-world property, namely both high network clustering and short characteristic path length. The inclusion of the degree distribution is both reasonable and physical because there should be a certain degree distribution for any networks with the small-world property (however, the detailed function for the distribution is unknown, or alternatively, dependent on different small-world networks).

Owing to the interesting flexible features of the Watts–Strogatz network model developed by including degree distribution, it is expected to help in the future analysis of economics and other social fields with the small-world property.

Acknowledgments

We acknowledge the financial supports by RSI-Fudan 2006 (YWC), by the National Natural Science Foundation of China under grant no 10604014, by Chinese National Key Basic Research Special Fund under grant no 2006CB921706, by the Shanghai Education Committee and the Shanghai Education Development Foundation ('Shu Guang' project) and by the Pujiang Talent Project (no 06PJ14006) of the Shanghai Science and Technology Committee. YWC thanks Ms Y Y Zhao, Ms C Z Fan, Ms L L Ma, Ms K Aull, Mr D Farris and Mr D Wang for their useful help. JPH thanks Professor K Yakubo of Hokkaido University for fruitful discussion.

References

[1] Watts D J and Strogatz S H 1998 *Nature* **393** 440
[2] Newman M E, Moore C and Watts D J 2000 *Phys. Rev. Lett.* **84** 3201
[3] Albert R and Barabási A-L 2002 *Rev. Mod. Phys.* **74** 47
[4] Newman M E 2001 *Proc. Natl Acad. Sci. USA* **98** 404
[5] Newman M E J 2001 *Phys. Rev.* E **64** 016131
[6] Barabási A-L, Jeong H, Néda Z, Ravasz E, Schubert A and Vicsek T 2002 *Physica* A **311** 590
[7] Wagner A and Fell D 2000 *Technical Report* No 00-07-041 (Santa Fe Institute) http://www.santafe.edu/research/publications/workingpapers/00-07-041.pdf
[8] Ferrer i Cancho R and Solé R V 2000 Santa Fe Institute working paper 01-03-016 (Santa Fe Institute) http://www.santafe.edu/sfi/publications/Working-Papers/01-03-016.pdf
[9] Watts D J 2003 *Small Worlds: The Dynamics of Networks Between Order and Randomness* (Princeton, NJ: Princeton University Press)
[10] Watts D J 2003 *Six Degrees: The Science of a Connected Age* (London: Heinammann)
[11] Song C, Havlin C and Makse H A 2005 *Nature* **433** 392
[12] Dallástas L, Baronchelli A, Barrat A and Loreto V 2006 *Europhys. Lett.* **73** 969
[13] Estrada E 2006 *Eur. Phys. J.* B **52** 563
[14] Huang Z-G, Xu X-J, Wu Z-X and Wang Y-H 2006 *Eur. Phys. J.* B **51** 549
[15] Zhang Z Z, Rong L L and Comellas F 2006 *J. Phys. A: Math. Gen.* **39** 3253
[16] Shao Z G *et al* 2006 *J. Phys. A: Math. Gen.* **39** 2035
[17] Kuperman M N, Ballard M and Laguna F 2006 *Eur. Phys. J.* B **50** 513
[18] Milgram S 1967 *Psychol. Today* **2** 60
[19] For example, see Huang J P and Holm C 2004 *Phys. Rev.* E **70** 061404
[20] Ghashghaie S, Breymann W, Peinke J, Talkner P and Dodge Y 1996 *Nature* **381** 767
[21] Odenbach S 2002 *Magnetoviscous Effects in Ferrofluids* (Berlin: Springer)
[22] Kaneko K 2003 *Phys. Rev.* E **68** 031909
[23] Albert R, Jeong H and Barabasi A-L 1999 *Nature* **401** 130