Eur. Phys. J. B **53**, 361–366 (2006)
DOI: 10.1140/epjb/e2006-00389-0

THE EUROPEAN PHYSICAL JOURNAL B

Random pseudofractal scale-free networks with small-world effect

L. Wang[a], F. Du, H.P. Dai, and Y.X. Sun

State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027, P.R. China

Received 20 March 2006
Published online 20 October 2006 –

Abstract. A random pseudofractal network (RPN) is generated by a recursive growing rule. The RPN is of the scale-free feature and small-world effect. We obtain the theoretical results of power-law exponent $\gamma = 3$, clustering coefficient $C = 3\pi^2 - 19 \approx 0.74$, and a proof that the mean distance increases no faster than $\ln N$, where N is the network size. These results agree with the numerical simulation very well. In particular, we explain the property of growth and preferential attachment in RPNs. And the properties of a class of general RPNs are discussed in the end.

PACS. 89.75.Hc Networks and genealogical trees – 05.10.-a Computational methods in statistical physics and nonlinear dynamics

1 Introduction

Many social, biological, and communication systems can be described as complex networks with numerous nodes (even can be subsystems) and edges between them [1–7]. Complex networks have attracted much attention across various fields of science and engineering, and researchers have been making their efforts to uncover more generic properties of complex networks. In these endeavors, the small-world effect [8–10] and the scale-free feature [11,12] are two significant discoveries of various complex networks.

The small-world networks are intermediate of the local regular networks and the fully random networks, such small-world networks are usually characterized by two features: a large clustering coefficient, which is a characteristic of regular networks, and a small mean distance, which is typically observed in random networks. In 1998, Watts and Strogatz first introduced a small-world network (WS model) [8], which can be constructed by starting with one-dimensional lattice and randomly rewiring each edge with some probability. Another popular model was proposed by Monasson [9] and by Newman and Watts (NW model) [10] independently. In this model, an edge is added with some probability between each pair nodes without breaking any original clusters. Subsequently, many other models have been suggested and improved [13–17].

The Erdös and Rényi random graph (ER model) [18] and the small-world model have the same feature that the degree distribution peaks at a mean value and decays exponentially. Yet, many networks in reality, such as the Internet, World Wide Web, social networks, metabolic networks and many others, are of the scale-free feature, namely, their degree distribution obeys a power-law. The Barabási and Albert's scale-free network model (BA model) [11,12] suggests that growth and preferential attachment are the two main ingredients of self-organization of a network. The subsequent researchers have focused on the statistical mechanics such as accelerating growth [19], nonlinear attachment [20,21] etc, and on various processes, such as percolation [22,23], cascade processes [24,25] etc.

In fact, the scale-free feature and high clustering are not exclusive for a large number of real networks. Yet, the above models have difficulty capturing these two characteristics simultaneously. For example, in BA model, the mean clustering coefficient is small and decreases with the increasing of network by numerical simulation [1]. Recently, a few authors have demonstrated the use of pure mathematical objects and methods to construct some deterministic graphs, a basic mechanics is recursive operation on graph topology. One is hierarchical organization by generating four replicas repeatedly [26–28], a second is pseudofractal obtained by attachment aiming at each edge [29], a third popular one is Apollonian networks [30], where a certain number of nodes will be added and connected to their three nodes at each step. Yet, these models are all regular graphs in a certain sense without statistical mechanics for consideration. A significant endeavor is the

[a] e-mail: leiwang@iipc.zju.edu.cn

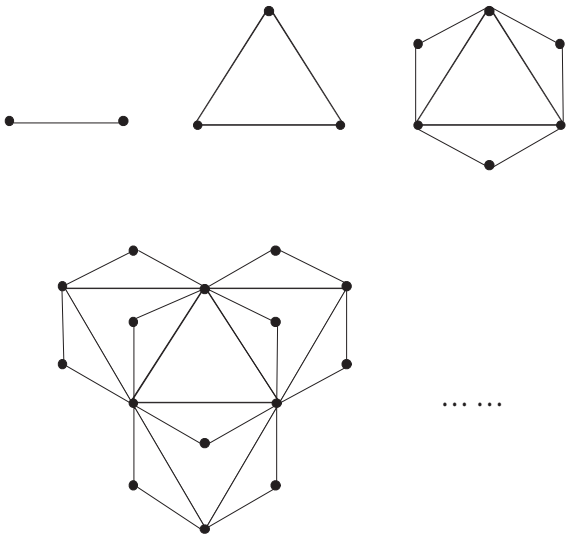

Fig. 1. Evolution of pseudofractal graph. The growth starts from a single edge with two nodes. At each time step, every edge generates an additional node, which is attached to both end nodes of the edge.

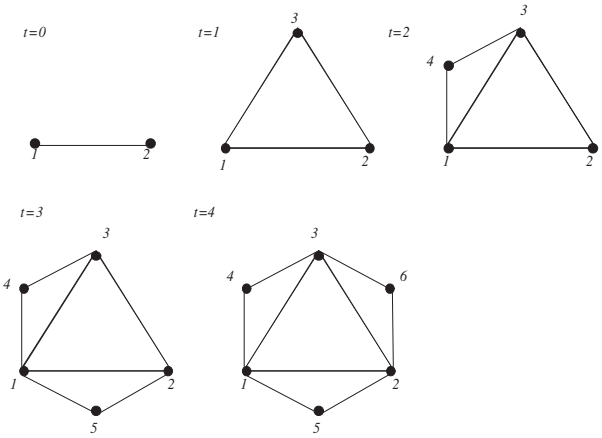

Fig. 2. Evolution of random pseudofractal graph. At the first time step $t = 1$, the fourth node is added to the network and randomly selected an edge labelled as l_{12} and connected to nodes 1 and 2. Then at time step 2, the edge l_{13} is selected and two edges l_{14}, l_{34} are added at the same time. With this rule, one can get a random pseduofractal network. Note that other edges such as l_{12} may be selected at the time step $t = 4$, which is different from pseduofractal network.

random Apollonian networks (RANs) introduced by Zhou etc. [31].

In this paper, we consider a simple rule that generates scale-free networks with high clustering and small mean distance. This network can be called random pseudofractal networks (RPN), which is a variation of pseudofractal graph. In fact, RPN was first proposed by Dorogovsev etc. [32] under the circumstance of self-organization of networks into scale-free features. Our paper focuses on the characteristics of this model. The rest of this paper is organized as follow: in Section 2, the RPN is introduced in detail. In Section 3, we give both theoretical results and numerical simulation of RPN based on rate equation modelling, including the scale-free feature and the small-world effect. In Section 4, some discussions including the homogeneous and different characteristics are given between RPN and BA model. At last, we extend a class of general RPNs by the similar recursive rule and simply discuss their properties.

2 Random pseudofractal networks

Pseudofractal graph. The growth starts from a single edge with two nodes at $t = -1$ (shown in Fig. 1). At each time step, a new node with two edges is added to every node existing of the graph, where the new edges are attached to both ends of the corresponding edge. Thus, this simple rule produces a growing network, where the total number of nodes at time t is $N_t = 3(3^t + 1)/2$, and the total number of edges is $L_t = 3^{t+1}$, so that the average degree is $\overline{k_t} = 2L_t/N_t = 4/(1 + 3^{-t})$.

RPN (shown in Fig. 2) makes a statistical mechanics in the Pseudofractal graph. Namely, at each time step, a new node, which is attached to both the end nodes of the edge, is added to an existing edge randomly chosen. Using this rule, one can get the random pseduofractal networks. It is clear that the total number of nodes at time t is $N_t = t + 2$, and the total number of edges is $L_t = 2t + 1$. Note that the minimum degree is 2.

3 Statistical characteristics of RPN

Degree distribution. In RPN model, the mean probability at time t that a node has the randomly chosen edge attached to it is equal to the degree k of the node divided by the total number of edges, $L_t = 2t + 1$. Therefore, let $n(k, t)$ be the number of nodes with degree k when t nodes are present, then we can immediately get a rate equation [20]:

$$n(k+1, t+1) = \frac{k}{L_t} n(k, t) + \left(1 - \frac{k+1}{L_t}\right) n(k+1, t) \quad (1)$$

$n(k, t)$ can be approximated as $tp(k)$, where $p(k)$ is the the probability density function of degree distribution. Then the above equation can be rewritten as

$$(t+1)p(k+1) = \frac{kt}{2t+1} p(k) + \left(1 - \frac{k+1}{2t+1}\right) tp(k+1) \quad (2)$$

furthermore, we have the recursive expression when t is sufficient large,

$$p(k+1) = \frac{k}{k+3} p(k). \quad (3)$$

Obviously, the probability density function is of the form:

$$p(k+1) = \frac{24}{k(k+1)(k+2)} p(2). \quad (4)$$

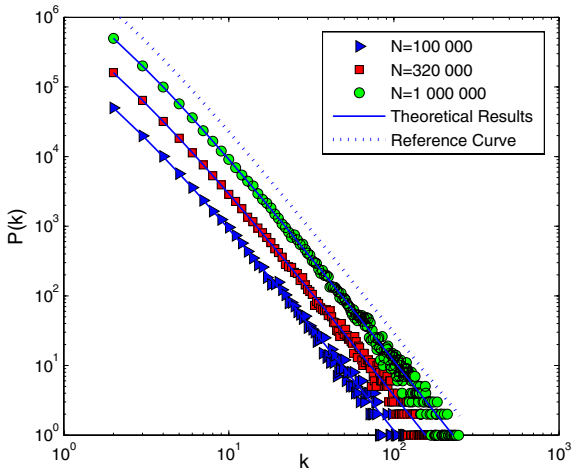

Fig. 3. (Color) Degree distribution of random pseudofractal networks , with $N = 1\,000\,000$ (green), $N = 320\,000$ (red), $N = 100\,000$ (blue). In this figure, $P(k)$ denotes the number of nodes of degree k, the dashed line (black) is the reference value by translation of theoretical results and the curves (blue) is the theoretical results. In particular, the number of the minimal degree has a departure from the power-law, which be easily explained by the density function obtained theoretically.

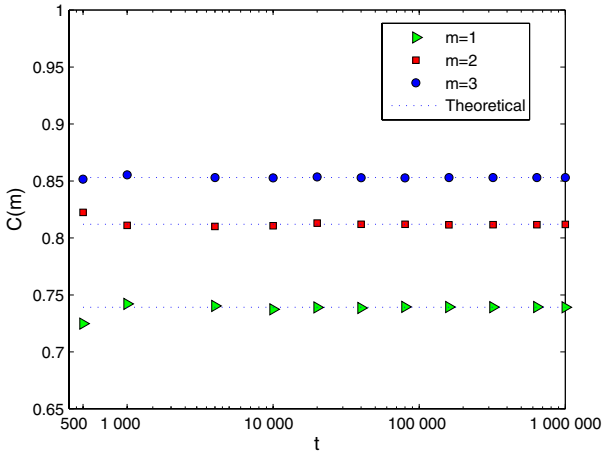

Fig. 4. (Color) Clustering coefficient of random pseudofractal networks, with $m = 1$ (green), $m = 2$ (red), $m = 3$ (blue). In this figure, one can find that the clustering coefficient of RPN is almost a constant, which agrees with the analytic results (dashed line). Furthermore, the clustering coefficients of m-RPNs increase as m increases.

It is clear that equation (4) satisfies the normalization equation:

$$\sum_{i=k_{min}}^{k_{max}} p(i) = 1 \tag{5}$$

and $k_{min} \gg k_{max}$, then $p(2) = \frac{1}{2}$. And the solution of the equation (1) is

$$p(k+1) = \frac{12}{k(k+1)(k+2)}. \tag{6}$$

This solution is the same as in reference [19]. When k is sufficiently big, the degree distribution satisfies $p(k) \propto k^{-\gamma}$ with $\gamma = 3$. In Figure 3, we show that the degree distribution for $t = 100\,000,\ 320\,000, 1\,000\,000$, respectively. The simulation results agree well with the theoretical ones.

Clustering coefficient. In general, clustering coefficient C_i of a node i is of such form:

$$C_i = \frac{2e_i}{k_i(k_i - 1)} \tag{7}$$

where k_i is the degree of the node i, and e_i is the number of edges among node i's neighbor-set. The clustering coefficient C of the whole network is the mean of all individual C_i's, i.e.,

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{8}$$

where N is the total number of node of the whole network.

In RPN, when node i is added to the network, it is of degree 2 and $e_i = 1$. If the degree of node i increases by 1 at some time, e_i will increase by 1 at the same time. Thus, we can write down the expression of e_i in term of $k(i)$:

$$e(i) = k(i) - 1. \tag{9}$$

Substituting equation (9) into equations (7, 8), we easily derive that

$$C = \frac{2}{t+2}\sum_{i=1}^{t+2} \frac{1}{k(i)}. \tag{10}$$

In order to calculate the expression above, we make use of the results obtained above, namely, $p(k) = 12/[k(k+1)(k+2)]$, thus,

$$C = 24 \sum_{k=k_{min}}^{k_{max}} \frac{1}{k^2(k+1)(k+2)} \tag{11}$$

where $k_{max} \gg k_{min} = 2$, then, $C = 2\pi^2 - 19 \approx 0.739$. Obviously, the clustering coefficient of RPN is constant and very big after a long evolution. Figure 4 reports the simulation results about the clustering coefficient of RPN, which agrees very well with the analytic one.

Mean distance. In small-world effect, most pairs of edges in many networks seem to be connected by a short path through the network, so that the mean distance can be an important parameter that measures a network to be or not be of a small-world effect.

Consider an undirected network with N nodes, let us define $L(N)$ to be the mean distance between nodes pairs in a network:

$$L(N) = \frac{2}{N(N+1)}\sum_{i \le j} d_{i,j} \tag{12}$$

where $\sigma(N) = \sum_{i \le j} d_{i,j}$ is the total distance of a network with N nodes, and $d_{i,j}$ denotes the distance between i and j.

According to the routine relation, the addition of a new nodes will not affect the distance between those already existing, then in our RPN at time t,

$$\sigma(t+3) = \sigma(t+2) + \sum_{i=1}^{t+2} d_{i,t+3}. \tag{13}$$

Assume that the node $t+3$ is added to the edge labelled as l_{mn}, where m and n are the labelled numbers of nodes shown in Figure 2. Note that the new node added at time t is labelled as $t+2$ since there are two initial nodes. Then,

$$\sum_{i=1}^{t+2} d_{i,t+3} = t+2+\sum_{i\in\Gamma} D_{i,x} \tag{14}$$

where $D_{i,x} = \min\{d_{i,m}, d_{i,n}\}$, and $\Gamma = \{1,2,\cdots,t+2\} - \{m,n\}$ is a node set with cardinality t. The sum $\sum_{i\in\Gamma} D_{i,x}$ can be considered as the total distance from one node x to all the other nodes in RPN with t nodes, which can be approximated in terms of $L(t+1)$:

$$\sum_{i\in\Gamma} D_{i,x} = tL(t+1). \tag{15}$$

Note that the mean distance $L(t)$ increases monotonously with t, then

$$tL(t+1) = \frac{2\sigma(t+1)}{t+1} < \frac{2\sigma(t+2)}{t+2}. \tag{16}$$

Combining equations (12–14), we have

$$\sigma(t+1) < \sigma(t) + t + \frac{2\sigma(t)}{t}. \tag{17}$$

Now consider the equation obtained from equation (17)

$$\frac{d\sigma(t)}{dt} = \frac{2\sigma(t)}{t} + t. \tag{18}$$

The solution is $\sigma(t) = t^2(\ln t + c)$, where c is a constant. Thus, we have $L(t) \sim \ln t$. Note that expression (17) is an inequality, then $L(t)$ increases at most as $\ln t$ with t. We show the simulation in Figure 5, which agrees well with the theoretical results.

4 Discussion and summary

The power-law exponent of RPN, $\gamma = 3$, is the same as the BA mdoel's. Moreover, the evolving mechanism is also similar to the BA's, and we can adopt continuous model to explain the growth of RPN. For simplicity, there are two nodes $s = 1, 2$, and each with degree 1 initially ($t = 2$), thus each node is labelled by the time of its birth, $0 < s \leq t$. Let $\overline{k}(s,t)$ be the mean degree of the node s at time t, then the continuous approximation [12,33] is of the form $p(k,s,t) = \delta(k - \overline{k}(s,t))$, consequently,

$$\frac{\partial \overline{k}(s,t)}{\partial t} = \frac{\overline{k}(s,t)}{L_t}. \tag{19}$$

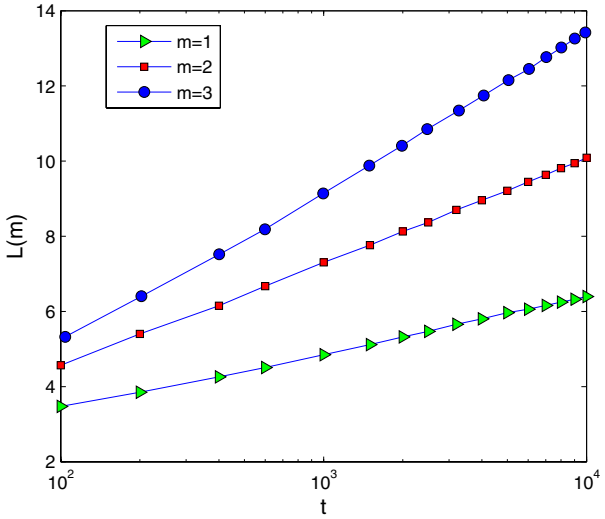

Fig. 5. (Color) The dependence between the mean distance L and $\ln t$ of random pseudofractal networks, with $m = 1$ (green), $m = 2$ (red), $m = 3$ (blue). In this figure, the dependence between L and $\ln t$ is almost a line, which accords with the analytic results. For m-RPNs, the slopes of corresponding lines increase as m increases.

Furthermore, equation (19) can be rewritten as

$$\frac{\partial \overline{k}(s,t)}{\partial t} = 2\frac{\overline{k}(s,t)}{\int_0^t \overline{k}(u,t)du} \tag{20}$$

where $\int_0^t \overline{k}(u,t)du = 2L_t$ is the degree of the total network at time t. And the number 2 on the right of equation (20) means the two added edges at each time step. Therefore, equation (20) consists of two ingredients of BA model: growth (at each time step, a new node is added to the network) and preferential attachment (the probability Π that a new node will be connected to node s depends on the average degree $\overline{k}(s,t)$ of node s, such that $\Pi = \overline{k}(s,t)/\int_0^t \overline{k}(u,t)du$ at time t). Though the true meaning of equations (19) and (20) are different, the two mathematical expressions are the same. We can study the RPN's evolution from the point of view of the BA's.

On the other hand, the mean clustering coefficient of BA depends on the network size as $C(N) \sim N^{-0.75}$ [1], which is much smaller than the RPN's. The degree distribution of a growing network may mainly determined by the two ingredients as reported in BA model, but the other significant properties such as the clustering coefficients, topology structures, are likely to be of great distinguish.

In fact, a general construction of RPNs can be introduced by the similar rule. This model can be described as: there is an edge with two nodes initially. At each time step, m nodes orderly connected by $m+1$ edges are attached to both ends of the corresponding edge randomly chosen. With this rule, a class of general RPN networks (m-RPN) will be produced, where the total number of nodes at time t is $N_{t,m} = mt + 2$, the total number of edges is $L_{t,m} = (m+1)t + 1$, and the mean degree is $\overline{k}_{t,m} \approx 2(m+1)/m$ as t is sufficient large. In particular, $m = 1$, the network will degenerate the RPN. Let

$n(k,t,m)$ be the number of nodes with degree k at the time t when m nodes are added at each time step, then

$$n(k+1,t+1,m) = \frac{k}{L_{t,m}} n(k,t,m) + \left(1 - \frac{k+1}{L_{t,m}}\right) n(k+1,t,m) + m\delta_{k2}. \quad (21)$$

The solution of equation (21) is given by following formula:

$$p(k,m) = \frac{(m+1)(m+2)!(k-1)!}{(k+m+1)!}. \quad (22)$$

When k is big enough, the degree distribution satisfies $p(k) \propto k^{-\gamma}$ with $\gamma = m+2$. A numerical simulation for $m = 1, 2, 3$ is shown in Figure 6, which accords with the analytic results. The definition of clustering coefficients in equation (10) for m-RPNs ($m \geq 2$) does not hold any more, we define a *clustering coefficient of order* x as the probability that there is a distance of length x between two neighbors of a node i [34]. Putting the number of such x-distance equal to $e_i(x)$, the higher order clustering coefficient follows:

$$C(x) = \frac{1}{N}\sum_{i=1}^{N} C_i(x) = \frac{1}{N}\sum_{i=1}^{N} \frac{2e_i(x)}{k_i(k_i-1)}. \quad (23)$$

And we are interested in the shortest distance among two neighbors for general RPNs, namely, $C(m)$. According to the growing way of general RPNs, it is easily derived that

$$e_i(m) = k_i - 1. \quad (24)$$

Obviously, $C(m)$ is determined by m such that

$$C(m) = \sum_{j=0}^{m} \left((\sum_{i=0}^{j} \alpha_i) \frac{1}{2+j} \right) + 1.29(m^2+3m+2) \quad (25)$$

where α_i, $(i = 0, 1, 2, \cdots, m)$ is the corresponding residue of function $C_i(m)$ such that

$$\frac{(k-1)!}{k(k+m+1)!} = \sum_{i=0}^{m} \frac{\alpha_i}{k+i} + \frac{\beta}{k^2}. \quad (26)$$

The numerical simulation is reported in Figure 4. In this figure, the clustering coefficients of the whole network are constant, which agree well with the analytic results, and increase as m increases for each m. The laws of mean distance of the m-RPN $L(t,m)$ are almost the same as the 1-RPN, i.e., $L(t,m) \propto \ln t$ shown in Figure 5 by simulation.

According to the analyses and simulation, the m-RPNs, which are constructed by a simple recursive rule, are not only scale-free, but also small-world. In particular, statistical mechanism is introduced in the growth process of m-RPN. Moreover, the m-RPN's growth is similar to the BA's, but there is a quite different clustering coefficient, which leads to a small-world effect for RPN. Since many real networks are both scale-free and small-world, this class (including RAN, m-RPN and hierarchical organization etc.) may perform perfectly in description of real networks properties.

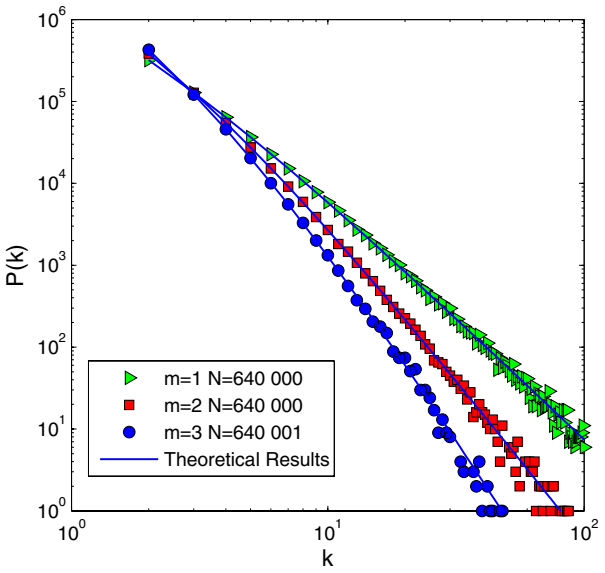

Fig. 6. (Color) Degree distribution of m-RPNs, with $m = 1$ (green), $m = 2$ (red), $m = 3$ (blue) and $N = 640\,000$. In this figure, $P(k)$ denotes the number of nodes of degree k, the dashed line (black) is the reference value by translation of theoretical results and the curves (blue) is the theoretical results. In particular, the number of the minimal degree has a departure from the power-law, which can be easily explained by the density function obtained theoretically.

This work is supported by the Chinese National Natural Science Foundation under grant 60304018.

References

1. R. Albert, A.L. Barabási, Rev. Mod. Phys. **74**, 47 (2002)
2. S.H. Strogatz, Nature **410**, 268 (2001)
3. D.J. Watts, *Small Worlds* (Princeton University Press, Princeton, 1999)
4. M.E.J. Newman, J. Stat. Phys. **101**, 819 (2000)
5. A.L. Barabási, *Linked: The New Science of Networks* (Perseus, Cambridge, 2002)
6. X.F. Wang, G.R. Chen, IEEE Circuits Syst. Mag. **3**, 6 (2003)
7. L.A.N. Amaral, J.M. Ottino, Eur. Phys. J. B **38**, 147 (2004)
8. D.J. Watts, S.H. Strogatz, Nature **393**, 440 (1998)
9. R. Monasson, Eur. Phy. J. B **12**, 555 (1999)
10. M.E.J. Newman, D.J. Watts, Phys. Lett. A **263**, 341 (1999)
11. A.L. Barabási, R. Albert, Science **286**, 509 (1999)
12. A.L. Barabási, R. Albert, H. Jeong, Physica A **272**, 173 (1999)
13. C.F. Moukarzel, Phys. Rev. E **60**, R6263 (1999)
14. M.A. Menezes, C. Moukarzel, T.J.P. Penna, Europhys. Lett. **50**, 574 (2000)
15. M. Ozana, Europhys. Lett. **55**, 762 (2001)

16. P. Sen, B.K. Chakrabarti, J. Phys. A **34**, 7749 (2001)
17. C.P. Zhu, S.J. Xiong, Y.J. Tian et al., Phys. Rev. Lett. **92**, 218702 (2004)
18. P. Erdös, A. Rényi, Publ. Math. **6**, 290 (1959)
19. S.N. Dorogovtsev, J.F.F. Mendes, Phys. Rev. E **63**, 025101 (2001)
20. P.L. Krapivsky, S. Redner, F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000)
21. P.L. Krapivsky, S. Redner, Phys. Rev. E **63**, 066123 (2001)
22. R. Albert, H. Jeong, A.L. Barabási, Nature **406**, 378 (2000)
23. S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, Eur. Phys. J. B **38**, 177 (2004)
24. Y. Moreno, R. Gómez, A.F. Pacheco, Europhys. Lett. **58**, 630 (2002)
25. T. Zhou, B.H. Wang, Chin. Phys. Lett. **22**, 1072 (2005)
26. A. L. Barabási, E. Ravasz, T. Vicsek, Physica A **299**, 559 (2001)
27. E. Ravasz, A.L. Somera, D.A. Mongru et al., Science **297**, 1551 (2002)
28. E. Ravasz, A.L. Barabási, Phys. Rev. E **67**, 026112 (2003)
29. S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, Phys. Rev. E **65**, 066122 (2002)
30. J.S. Andrade, Jr., H.J. Herrmann, R.F.S. Andrade et al., Phys. Rev. Lett. **94**, 018702 (2005)
31. T. Zhou, G. Yan, B.H. Wang, Phys. Rev. E. **71**, 046141 (2005)
32. S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Phys. Rev. E **63**, 062101 (2001)
33. S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000)
34. A. Fronczak, J.A. Holyst, M. Jedynak et al., `arXiv: cond-mat/0212237`