Vol 13 No 3, March 2004
1009-1963/2004/13(03)/0273-06

Chinese Physics

© 2004 Chin. Phys. Soc.
and IOP Publishing Ltd

Pair correlations in scale-free networks*

Huang Zhuang-Xiong(黄壮雄)†, Wang Xin-Ran(王欣然), and Zhu Han(朱 涵)†

Department of Physics, Nanjing University, Nanjing 210093, *China*

(Received 15 April 2003; revised manuscript received 16 June 2003)

Correlation between nodes is found to be a common and important property in many complex networks. Here we investigate degree correlations of the Barabasi–Albert (BA) scale-free model with both analytical results and simulations, and find two neighbouring regions, a disassortative one for low degrees and a neutral one for high degrees. The average degree of the neighbours of a randomly picked node is expected to diverge in the limit of infinite network size. As a generalization of the concept of correlation, we also study the correlations of other scalar properties, including age and clustering coefficient. Finally we propose a correlation measurement in bipartite networks.

Keywords: networks, stochastic processes, complex systems, statistical mechanics

PACC: 0250, 0520

Complex networks, described by a set of nodes and links between them, play an important role in the understanding of many natural systems. They take various forms in different fields, including the Internet,[1,2] social networks,[3] ecological networks,[4] etc (see also Refs.[5,6]). Yet, due to technical difficulties of information acquisition and analysis, the rewards of research in this area did not come until several years ago, mainly after the finding that complex networks in reality usually do not conform to the long assumed theory of random network proposed by Erdos and Renyi.[7]

Among the models aimed at explaining the observed nontrivial structures, the Barabasi and Albert (BA) scale-free model[8] is well known for describing systems where geographical distance is not so important. It introduces two mechanisms that are believed to be common in reality: network growth (adding one node at each time step and connecting it to m existing nodes) and preferential attachment (well-connected nodes tend to receive more new links, with the probability proportional to the number of the links they already have). It has explained the power-law degree distribution widely observed in reality (degree of a node is defined as the number of its nearest neighbours). Based on it some new mechanisms have been further introduced to describe other observations such as aging effect,[9] high clustering, and hierarchical structure,[10] etc.

An important property of networks, the topic of the present article, is the pair correlation (or mixing pattern).[1,2,11–14] "Whether gregarious people are more likely to contact with gregarious people?" or "Whether an old web site is more likely to connect to old ones?" These questions often have fundamental importance in reality. The topology and many properties such as resilience and percolation of a network are all susceptible to the answers. Very recently, a particular kind of it, the pair degree correlation, has received much attention. Newman has put forward a direct measurement,[12,13] making it a quantified property, and has applied this measurement to various real networks. The data available reveal an interesting feature: in social networks, high degree nodes often tend to be connected to other high degree nodes (assortative mixing), while in technological and biological networks high degree nodes often tend to connect to low degree nodes (disassortative mixing). At the same time, there are some systems reported to show no significant biases (neutral mixing). This assortative (disassortative, or neutral) pattern has a profound effect on the structure of evolving networks. For example, it may tend to break a network into separate communities and affect its error-and-attack tolerance, etc.

As a leading model in this field of complex networks, the BA model[8] has been developed and stud-

*Project supported by the National Natural Science Foundation for Distinguished Young Scientists of China (Grant No 60225014).
†E-mail: hanzhunju@hotmail.com

ied extensively in the past few years (see Refs.[5,15,16] and references therein). Yet as to correlation, theoretical investigation of the BA model is still inadequate, which we believe is fundamental to understand the correlation properties of real networks, since many of them are shown to be scale free. In this sense, the present work on the pair-correlation properties of the BA model may provide a basis for theoretical explanations of the observations in reality. In particular, this model is likely to be a good starting point for finding the origin of the emergence of mixing patterns.

In this article, we focus on the correlation properties within the framework of the BA networks. We provide an analytical calculation of the average degree of a node's nearest neighbours, indicating that the BA networks are neutral in the limit of large connectivity, otherwise they are disassortative. The comparison of the analytical and simulation results shows that the absolute value of the average connectivity of neighbouring nodes also depends on the size of the network and diverges as this size tends to infinity. We then extend the investigation to the correlation of age and clustering coefficient, i.e. the ratio of edges among the nearest neighbours of a selected node and the maximal number of edges among them. Measurements of these two properties have not been carried out before to our knowledge. In the final part, we propose a way to measure the mixing patterns in bipartite networks.

In the following we call a node with k degrees a $D-k$ node, a node with age a an $A-a$ node, a node born at time t a $T-t$ node, and a node with clustering coefficient c a $C-c$ node.

Degree Correlation The best way to completely describe the degree correlation is to obtain the matrix, $P(k,k_{nn})$, i.e. the probability that a nearest neighbour of a $D-k$ node is $D-k_{nn}$. However, due to technical problems this measurement is often unfeasible in practice.[1,2] Instead, Pastor-Satorras *et al*[2] suggested measuring $\langle k_{nn}\rangle_k$, i.e. the average degree of a nearest neighbour of the $D-k$ nodes. Here our work about the BA model is along this line.

In Ref.[15], Krapivsky and Render obtained in the BA model a useful characterization of correlation, $N_{kl}(t)$, the number of $D-k$ nodes that attach to a $D-l$ ancestor. Asymptotically, $N_{kl}(t)\to tn_{kl}$, where

$$n_{kl}=\frac{4(l-1)}{k(k+1)(k+l)(k+l+1)(k+l+2)}+\frac{12(l-1)}{k(k+l-1)(k+l)(k+l+1)(k+l+2)}. \quad (1)$$

With this result Krapivsky and Render [15] concluded that the BA networks have an assortative mixing pattern, i.e. nodes with similar degrees are more likely to be connected. Their calculation has considered the direction of links between nodes, which makes the measurement a rather complex task in practice. Here, as in almost all measurements carried out so far, we consider undirected links. Similarly, we study $N'_{kl}=N_{kl}+N_{lk}$, the numbers of connected pairs of nodes with k degrees and l degrees respectively. Asymptotically

$$N'_{kl}/t\to n'_{lk}=n_{kl}+n_{lk}.$$

The probability that a nearest neighbour of a $D-k$ node is $D-k_{nn}$ is

$$\left(N_{k,k_{nn}}+N_{k_{nn},k}\right)\Big/\sum_{k_{nn}}\left(N_{k,k_{nn}}+N_{k_{nn},k}\right), \quad (2)$$

and the average degree of a nearest neighbour of a $D-k$ node is

$$\langle k_{nn}\rangle_k=\frac{\sum_{k_{nn}}k_{nn}\left(N_{k,k_{nn}}+N_{k_{nn},k}\right)}{\sum_{k_{nn}}\left(N_{k,k_{nn}}+N_{k_{nn},k}\right)}\to\frac{\sum_{k_{nn}}k_{nn}\left(n_{k,k_{nn}}+n_{k_{nn},k}\right)}{\sum_{k_{nn}}\left(n_{k,k_{nn}}+n_{k_{nn},k}\right)}. \quad (3)$$

This is a useful result characterizing the degree correlation of the BA networks. We show it approximately in Fig.1(a) by taking the summation of l to 1.5×10^5 and 2×10^6 (with normalization satisfied) respectively, in comparison with the simulation results at $N=100,1000$ and 10000. (The simulation results reported here are all obtained by taking average of up to 5000 independent runs). This reveals that nodes with large k show no obvious biases in their associations, as has been reported in previous studies. However, when k is relatively small, $\langle k_{nn}\rangle_k$ falls significantly as k increases, and this is a sign of disassortative mixing. (This region has not yet been noticed previously and, actually, nodes falling in this region often take up a significant part in real networks that have been measured so far.) In Fig.1(b) we show two characteristic distributions of the probability that a nearest neighbour of a $D-k$ node is $D-k_{nn}$. Both simulation and theoretical results show that this probability declines as approximately k_{nn}^{-2} for large k_{nn}, which is a sign of neutral mixing since the probability of finding a $D-k_{nn}$ node in the network is approximately k_{nn}^{-3}.[8] It also leads to the conclusion that $\langle k_{nn}\rangle$ will diverge for infinite network size N as $\ln N$ (since

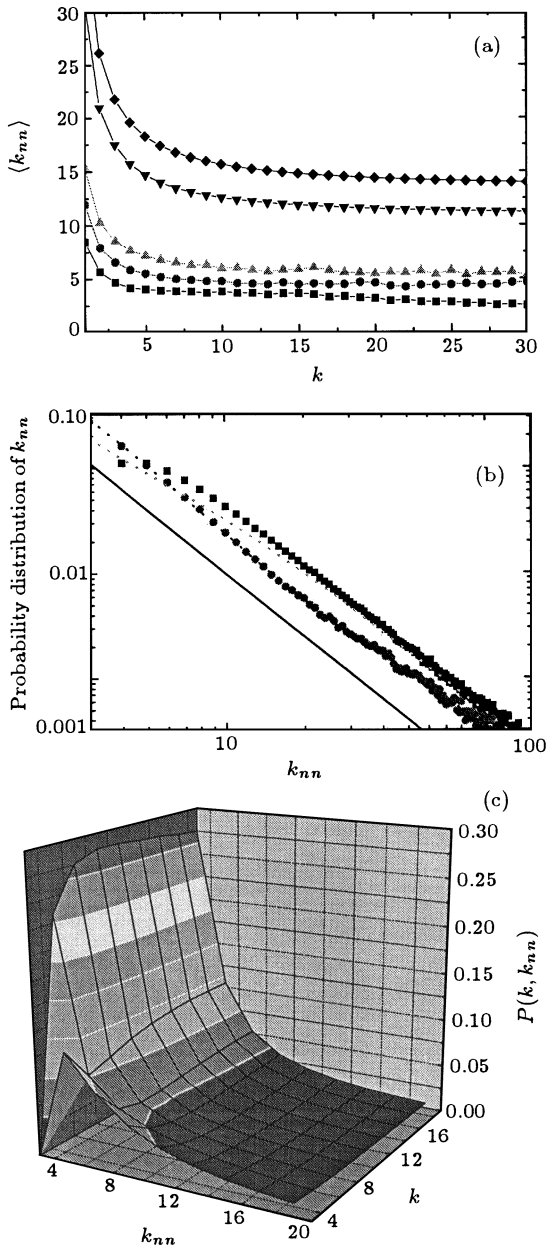


Fig.1. Simulation results of degree correlation of a network with $m = 3$. (a) Average degree $\langle k_{nn} \rangle$ of the neighbouring nodes of the $D-k$ nodes as a function of k. Squares, circles and upward triangles represent the simulation results, with system size $N = 100$, 1000 and 10000 respectively. Downward triangles and diamonds represent the theoretical result with l up to 1.5×10^5 and 2×10^6 respectively. (b) Degree distributions of the nearest neighbours of $D - 3$ nodes (squares) and $D - 20$ nodes (circles) respectively. The dashed lines are the corresponding theoretical results. The solid line with slope -2 serves as a guide to the eye. (c) The probability matrix $P(k, k_{nn})$.

the largest possible value of $k \sim N^{1/2}$, see Eq.(5) below), and this trend is observed in Fig.1(a). The deviation from this power law is evident when k_{nn} is small. On the other hand, as can be noticed, the simulation and theoretical results agree better for larger k. Here the finite size effect mainly results from the fact that, when a network has a finite size, the nodes with degrees out of a certain range will be absent. Finally, Fig.1(c) shows the simulation result of the probability matrix, $P(k, k_{nn})$. While for each value of k the probability has similar distribution for large k_{nn} (k_{nn}^{-2}), the difference is obvious for relatively small values of k_{nn}.

Age Correlation As an extension of the degree correlation, we define age correlation in a similar way, i.e. consider the probability that an $A - a_i$ node is linked with an $A - a_j$ node. (Here age a of a node is defined as $t_{\mathrm{F}} - t$, where t_{F} is the age of the network and t denotes the time when the node is born.) If this conditional probability $P(a, a_{nn})$ is independent of a, we are in a topology without any age correlation. We may also study the quantity $\langle a_{nn} \rangle_a$, i.e. the average age of a nearest neighbour of the $A - a$ nodes. For convenience, in the following we shall calculate $\langle t_{nn} \rangle_t$ instead of $\langle a_{nn} \rangle_a$.

In a BA network, at each time step, a newly born node is connected to m existing nodes, while the rate at which a node receives new links is proportional to its current degree. Using continuous approach,[8] we have

$$\frac{\mathrm{d}k_s(t)}{\mathrm{d}t} = \frac{mk_s(t)}{2mt} = \frac{k_s(t)}{2t}, \tag{4}$$

where $k_s(t)$ denotes the degree of a node bourn at time s. Solving this equation, we can obtain

$$k_s(t) = m\left(\frac{t}{s}\right)^{\frac{1}{2}}. \tag{5}$$

Consider two nodes bourn at time t_i and t_j respectively. Assume $t_i < t_j$, and the probability that they are connected is

$$p(t_i, t_j) = \frac{mk_{t_i}(t_j)}{2mt_j} = \frac{m}{2\sqrt{t_i t_j}}, \tag{6}$$

where $k_{t_i}(t_j)$ is the degree of the $T - t_i$ node at time t_j. This result is also valid if $t_i > t_j$. Then

$$\langle t_{nn} \rangle_{t_i} = \frac{\sum\limits_{t_j=1}^{t_{\mathrm{F}}} t_j p(t_i, t_j)}{\sum\limits_{t_j=1}^{t_{\mathrm{F}}} p(t_i, t_j)} = \frac{\sum\limits_{t_j=1}^{t_{\mathrm{F}}} \sqrt{t_j/t_i}}{\sum\limits_{t_j=1}^{t_{\mathrm{F}}} \left(1/\sqrt{t_i t_j}\right)}. \tag{7}$$

Replacing the summation with integration, we have

$$\langle t_{nn} \rangle_{t_i} \approx \frac{\int_1^{t_{\mathrm{F}}} \sqrt{t_j/t_i}\,\mathrm{d}t_j}{\int_1^{t_{\mathrm{F}}} \left(1/\sqrt{t_i t_j}\right)\mathrm{d}t_j} = \frac{1}{3}\frac{t_{\mathrm{F}}^{3/2} - 1}{t_{\mathrm{F}}^{1/2} - 1} \approx \frac{1}{3}t_{\mathrm{F}}, \tag{8}$$

which is independent of t_i. To support the above analysis, we show the simulation results in Fig.2. $\langle t_{nn} \rangle_t$ as a function of t is shown in Fig.2(a), and is found to be fluctuating around $t_F/3$, as predicted by Eq.(8).

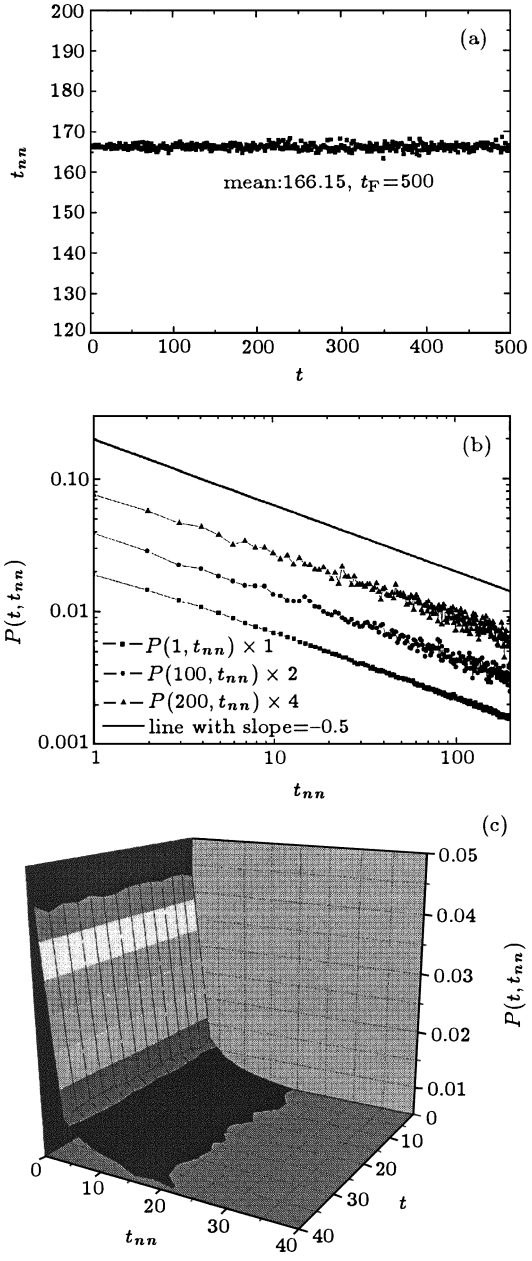


Fig.2. Simulation results of age correlation of a network with size $N = 10^4$, $t_F = 500$ and $m = 3$. (a) Average introduction time of the nearest neighbours of a $T - t$ node as a function of t. (b) Introduction time distributions of the neighbours of the $T - 1$ (squares), $T - 100$ (circles) and $T - 200$ (upward triangles) node. The solid line has the slope -0.5. (c) The matrix $P(t, t_{nn})$.

In Fig.2(a), the nearest-neighbour age distributions of three characteristic nodes all show a power law, $t_{nn}^{-\gamma}$, with the exponent $\gamma \approx -0.5$, in accordance with Eq.(6). Figure 2(c) shows the matrix $P(t, t_{nn})$, i.e. the probability that a nearest neighbour of a $T - t$ node is $T - t_{nn}$. We can see clearly from this graph that the distribution is largely independent of t, which indicates a neutral pattern in this age correlation.

Clustering Correlation As another extension of this concept, we also investigate the pair correlation of the clustering coefficient, using the same method as that in the preceding parts. Different from a node's degree, clustering coefficient is a continuous parameter between 0 and 1. In order to define the correlation, discretization is necessary. Here we evenly divide the region $[0, 1]$ into 250 subregions and each subregion is represented by the median value. In this way, the correlation of clustering can be defined in the way similar as before. Consider the probability that a node with its clustering coefficient in the c-subregion (for convenience we call it a $C - c$ node) is linked with a $C - c_{nn}$ node. If this conditional probability $P(c, c_{nn})$ is independent of c, we are in a topology without any correlation.

In our simulation, the clustering coefficients are mainly limited in a relatively small region $[0, 0.14]$ (we can see from Fig. 3(b)), and we have poor statistics concerning the nodes out of this region. Our simulation result (Fig. 3(a)) shows that the average clustering coefficient of the nearest neighbours of a $C - c$ node, $\langle c_{nn} \rangle_c$, generally shows a trend to ascend as c increases. This is a sign of assortative mixing in the sense of clustering, which may be because clustering is a collective behaviour. If a node with large c belongs to a cluster, it is likely that its neighbours also belong to that cluster and hence also have large c. At the same time, we find that, interestingly, there is a notable peak in the curve of $\langle c_{nn} \rangle_c$. We may call it a central mixing pattern, in order to develop the ideas of Newman.[12,13] The value of this peak agrees with that of the clustering coefficient distribution of the network. This coincidence suggests that the observed peak may result from the fact that, in the network, the nodes that have their clustering coefficients around this value take up a significant part of the system. Figure 3(b) shows the $P(c, c_{nn})$ distributions with three typical values of c. Three curves all show a peak at approximately the above-mentioned position—a coincidence again. Figure 3(c) shows the non-trivial part of the matrix $P(c, c_{nn})$.

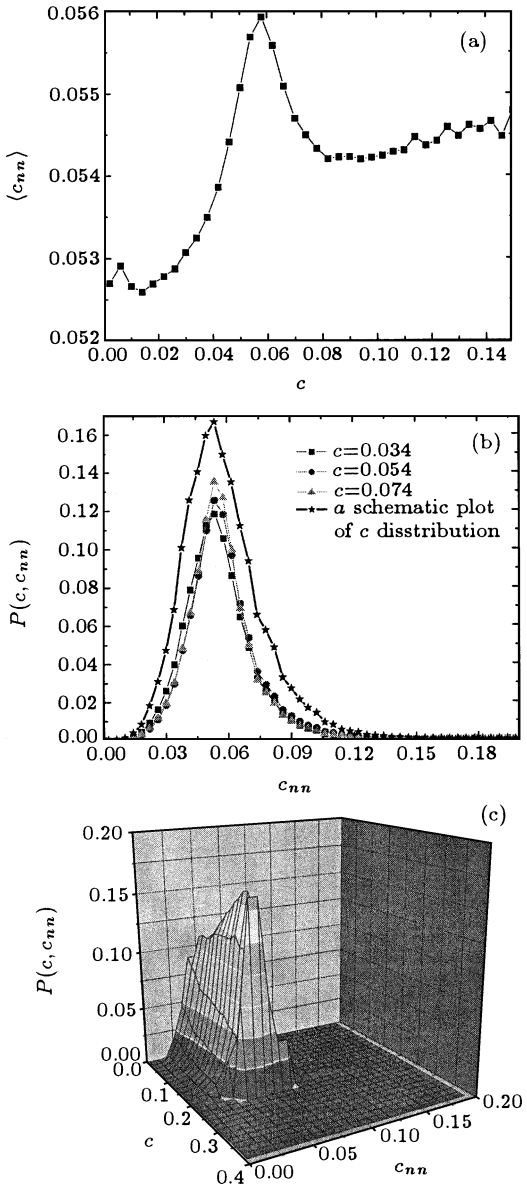


Fig.3. Simulation results of clustering coefficient correlation of a network with size $N = 2500$ and $m = 25$. (a) Average clustering coefficient of the neighbours of $C - c$ nodes as a function of c. (b) Clustering coefficient distributions of the nearest neighbours of the $C - 0.034$ (squares), $C - 0.054$ (circles) and $C - 0.074$ (upward triangles) nodes. The distribution of clustering coefficient of the network (multiplied by 1.68 for convenience) is shown as stars. (c) The matrix $P(c, c_{nn})$.

Before concluding we briefly mention the correlation measurement in bipartite networks,[6,17] which are formed by two (or more) classes A and B of nodes with links running between only different kinds. They may describe various systems, e.g. books and readers in a library, or directors and boards in the business world.[6] Here a new kind of correlation may be of major interest, the correlation between the properties of an A node and the same properties of the B nodes it is linked with. And this may answer such questions as whether a much-reading reader borrows books that are also liked by other people.

To conclude, as is now widely accepted, correlation is an important index of and significantly influences network structure and function, in that it reveals the relationship between a node and its neighbourhood. This article is concerned theoretically with the correlation of the BA model. Most of the previous studies focus on the degree correlation, and here we generalize the concept to various properties. We analytically treat the degree correlation (disassortative plus neutral) and the age correlation (neutral) of the BA network, and provide simulation results about the clustering coefficient correlation (assortative plus central). The mixing patterns of these scalar properties and their origins are presented and explained. Considering the wide applicability and theoretical importance of the BA model, the results reported here may serve as a useful reference for further studies, especially for the investigation in the variants of the BA model that incorporate, for example, nonlinear preferential attachment,[18] aging effect,[9] etc.

Presently the observed assortative (disassortative) patterns in reality have stimulated efforts to study their influence on network structure and function, mainly with models that are created by adding correlation effect to random networks.[11−13] However, it is well known that real networks have robust organization principles (for example, the preferential attachment, as characterized by the BA model). What is the relationship between correlation and these principles? What is the origin of the correlation pattern in the process of network growth? These questions are far from being completely answered. Recent efforts [19] have partly shown how they can be related to some features added to the BA model, such as fitness and some growing restraints.

We appreciate helpful discussions with A.-L. Barabasi and K. Klemm.

References

[1] Vazquez A, Pastor-Satorras R and Vespignani A 2002 *Phys. Rev.* E **65** 066130

[2] Pastor-Satorras R, Vazquez A and Vespignani A 2001 *Phys. Rev. Lett.* **87** 258701

[3] Jin E M, Girvan M and Newman M E J 2001 *Phys. Rev.* E **64** 046132

[4] Lassig M, Bastolla U, Manrubia S C and Valleriani A 2001 *Phys. Rev. Lett.* **86** 4418

[5] Albert R and Barabasi A L 2002 *Rev. Mod. Phys.* **74** 47

[6] Strogatz S H 2001 *Nature* **410** 268

[7] Erdos P and Renyi A 1960 *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17

[8] Barabasi A L and Albert R 1999 *Science* **286** 509

[9] Amaral *et al* 2000 *Proc. Natl Acad. Sci.* USA **97** 11149

[10] Ravasz E and Barabasi A L 2003 *Phys. Rev.* E **67** 026112

[11] Berg J and Lassig M 2002 *Phys. Rev. Lett.* **89** 228701

[12] Newman M E J 2002 *Phys. Rev. Lett.* **89** 208701

[13] Newman M E J 2003 *Phys. Rev.* E **67** 026126

[14] Goh K I, Oh E, Kahng B and Kim D 2003 *Phys. Rev.* E **67** 017101

[15] Krapivsky P L and Render S 2001 *Phys. Rev.* E **63** 066123

[16] Dorogovtsev S N, Mendes J F F and Samukhin A N 2000 *Phys. Rev. Lett.* **85** 4633

[17] Holme P, Liljeros F, Edling C R and Kim B J 2003 cond-mat/0302296

[18] Krapivsky P L, Redner S and Leyvraz F 2000 *Phys. Rev. Lett.* **85** 4629

[19] Park J, Newman M E J 2003 cond-mat/0303327