Scaling exponents and clustering coefficients of a growing random network

Haijun Zhou
Max-Planck-Institute of Colloids and Interfaces, Potsdam 14424, Germany
(Received 7 March 2002; published 25 July 2002)

The statistical property of a growing scale-free network is studied based on an earlier model proposed by Krapivsky, Rodgers, and Redner [Phys. Rev. Lett. **86**, 5401 (2001)], with the additional constraints of forbidding self-connection and multiple links of the same direction between any two nodes. Scaling exponents in the range of 1–2 are obtained through Monte Carlo simulations and various clustering coefficients are calculated, one of which, C_{out}, is of the order of 10^{-1}, indicating that the network resembles a small world. The out-degree distribution has an exponential cutoff for large out degree.

DOI: 10.1103/PhysRevE.66.016125 PACS number(s): 89.75.Hc, 05.40.−a, 89.75.Da, 87.10.+e

I. INTRODUCTION

To study the statistical property of a complex system composed of many interacting individual components, it is often helpful to map the system into a network of nodes and links (edges). Each node in this network represents one component of the real system and the interaction, if there is any, between two components is denoted by an edge, either directed or undirected, between the two corresponding nodes in the network. One quantity of interest is the node-degree profile of the formed network: How does $n(k)$, the total number of nodes with a given number k of links attached (the node degree), scales with k? Empirical observations revealed that many social and biological networks have the scale-free property [1,2], that is,

$$n(k) \sim k^{-\nu} \tag{1}$$

as k becomes large enough. The scaling exponent ν is typically in the range of $2<\nu<3$; but there are evidences that some networks have scaling exponents in the range of 1–2 [3] while a few other networks have scaling exponents larger than 3 (for a collection of experimental data, please refer to Table II of Ref. [2]).

To explain the scale-free characteristics, one appealing mechanism is to assume that the network (1) keeps growing and, (2) during this growth process, new edges are generated and are attached preferentially to those nodes that have already been attached by a large number of edges [1]. Based on this mechanism several models have been suggested [1,4–6], but they predicted the scaling exponent ν to be greater than 2, thus failed to explain the behavior of those networks with smaller ν. In Ref. [7] this "preferential attachment" mechanism was questioned partly because of this apparent discrepancy between theory and empirical data. In Ref. [8] it was shown that if one assumes that a network is growing accelerately, it is possible to generate scaling exponent in the range between 3/2 and 2. However, it is still not clear whether or not this condition is absolutely necessary to explain experimental observations.

An emerging property of almost all the so-far studied scale-free networks is that they can at the same time be classified as small-world networks [9]. That is, (i) the diameter of the network scales as ln(N), where N is the total number of nodes in the system, and (ii) the clustering coefficient C is independent of N and is thus much greater than that of a random network ($\simeq\langle k\rangle/N$, where $\langle k\rangle$ denotes the average node degree of the network). On the theoretical side, it was confirmed that growing networks generated by the mechanism of preferential attachment will typically have diameter scales as ln(N) (see, for example, Ref. [10]). However, the original Barabási-Albert model [1,2] predicted a very small clustering coefficient $C\sim N^{-0.75}$. The clustering coefficients for other models [4–6] were not reported.

To improve our understanding on scale-free networks, in this work two questions are addressed: Will it be possible to generate a scale-free network with scaling exponent $\nu<2$ based on the preferential attachment mechanism? Will it be possible for a scale-free network generated in this way to have relatively constant clustering coefficients? We answer these questions confirmatively by studying a revised Krapivsky-Rodgers-Redner model [4] with Monte Carlo simulation approach.

II. THE GROWING NETWORK MODEL

The Krapivsky-Rodgers-Redner model [4] is a generalized version of the original Barabási-Albert model on scale-free networks [1]. It has the following key elements: (i) edges are directed; (ii) new nodes are added into the network and are attached preferentially to existing nodes with larger in degrees; and (iii) creation of edges between "old" nodes are possible and a newly created edge also prefers to attach to nodes with larger degrees. This model is general in the sense that it takes into account directional interactions of the real network systems, and that the growth of the network is not solely caused by the inclusion of new nodes but also as a result of the increased interactions among the existing nodes of the system. It can be corresponded to the real systems including the World Wide Web (WWW), the Internet, the food web, the transportation network, the email network, etc. Because of its generality, we reexamine this model in the present work to illustrate the property of growing scale-free network.

We noticed in the original Krapivsky-Rodgers-Redner model [4] that the growth process permits the following two possibilities illustrated in Fig. 1: (1) a directed edge can originate from and end into the same node (self-connection)

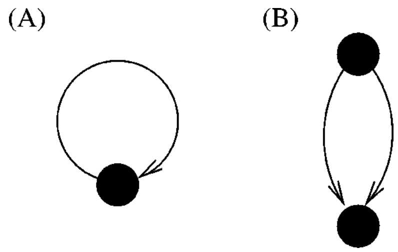


FIG. 1. Self-connection (a) and multiconnection of the same direction (b) are forbidden in the present simulation.

and (2) there can be more than one edge of the same direction between two nodes (multiconnection). Allowing these two possibilities makes analytical calculations possible and these authors found that in the growing network, both the in-degree and the out-degree distributions follow the power law [4]:

$$n_{\mathrm{in}}(k) \sim k^{-\nu_{\mathrm{in}}}, \quad 2 < \nu_{\mathrm{in}} = 2 + p\lambda < \infty, \tag{2}$$

$$n_{\mathrm{out}}(k) \sim k^{-\nu_{\mathrm{out}}}, \quad 2 < \nu_{\mathrm{out}} = 1 + 1/(1-p) + \mu p/(1-p) < \infty, \tag{3}$$

(p, λ, and μ are the three adjustable model parameters). Although the permission of self-connections and multiconnections might be reasonable for some networks (such as WWW pages, in which a page can have several links to another page and it can have a link to bring the reader from one portion to another portion of the same page), it may fail for other kinds of networks (in a co-authorship network [11], it is meaningless for an author to be the co-author of himself/herself; and in a food web [12], there is at most one edge of the same direction between two species). Because of the preferential attachment mechanism, if a node already has a large number of incoming and outgoing edges, it has a good possibility to form self-connections and by doing so its dominance is further amplified; the inclusion of multiconnection could lead to similar effects. As a result, edges may concentrate on just few nodes, making the scaling behavior steeper. In the present work, the two kinds of edges listed in Fig. 1 are discounted. Because of the reasoning outlined above, scaling exponents $1 < \nu < 2$ might occur in a growing network without self-connection and multiconnection. This point will be checked by Monte Carlo simulation.

III. MONTE CARLO SETUP

The revised Krapivsky-Rodgers-Redner model is studied by Monte Carlo (MC) simulation. Started with a single node, at each step:

(i) With probability p, a new node is created and a directed edge from it to an existing target node β is set up. Of all the N existing nodes, β will be selected with probability [4]

$$P_{\mathrm{attach}}(\beta) = \frac{k_{\mathrm{in}}(\beta) + \lambda}{E + \lambda N}. \tag{4}$$

In the above equation, $k_{\mathrm{in}}(\beta)$, the in degree, is the total number of incoming edges of node β; λ is a constant signifying the "initial attractiveness" of a node [6]; and E is the total number of edges in the system before this new edge is created.

(ii) With probability $q = 1 - p$, a new edge pointing from one node α to another node β is created, provided that (1) $E < N(N-1)$, (2) α and β are not identical, and (3) there is no preexisting directed edge from α to β. The probability that α and β will be selected is governed by the probability

$$P_{\mathrm{connect}}(\alpha,\beta) = \frac{[k_{\mathrm{out}}(\alpha) + \mu][k_{\mathrm{in}}(\beta) + \lambda]}{\sum_{\gamma} \sum_{\delta}{}' \, [k_{\mathrm{out}}(\gamma) + \mu][k_{\mathrm{in}}(\delta) + \lambda]}, \tag{5}$$

where $k_{\mathrm{out}}(\alpha)$ denotes the out degree of node α; μ is another constant with similar physical meaning as λ; $\Sigma_{\delta}{}'$ denotes the summation over all the nodes δ, which is not yet approached by a directed edge from node γ.

For large system size N it turns out to be quite inefficient and complicated when performing procedure (ii) based on a direct application of Eq. (5). This is partly because of the fact that, after a new edge has been created between nodes α and β, one must update the value of the summation in Eq. (5) by $O(N)$ iterations. To speed up procedure (ii), the selection of two nodes and the connection of an edge between them is finished actually through the following way:

(1) Select an outgoing node α with probability $[k_{\mathrm{out}}(\alpha) + \mu]/(E + \mu N)$.

(2) Select an incoming node β with probability $[k_{\mathrm{in}}(\beta) + \lambda]/(E + \lambda N)$.

(3) If α and β are identical, or if there is already a directed edge from α to β, repeat steps (1) and (2); if else, accept α and β and update the system.

It is not difficult to prove that by this method the probability for nodes α and β to be chosen is identical to Eq. (5).

The algorithm code is written in C++ language [13], with some of its standard containers (including *map* and *set*) being exploited.

To estimate the scaling exponents from the simulated data, we use two methods. One can directly fit the data with Eq. (1). Alternatively, one can define the cumulative degree distribution by

$$P(k) = \sum_{k' \geq k} n(k') \sim k^{-(\nu-1)}, \tag{6}$$

and from the cumulative distribution data an estimation of the value of ν could be obtained.

IV. DEGREE DISTRIBUTION OF THE GROWING NETWORK

In the growing network model, there are three adjustable parameters, namely, p, λ, and μ. Figure 2 shows the relations between the average number of nodes and the in degree and out degree for both the original and the revised Krapivsky-Rodgers-Redner models. Figure 3 shows the corresponding cumulative degree distributions for the two models. The net-

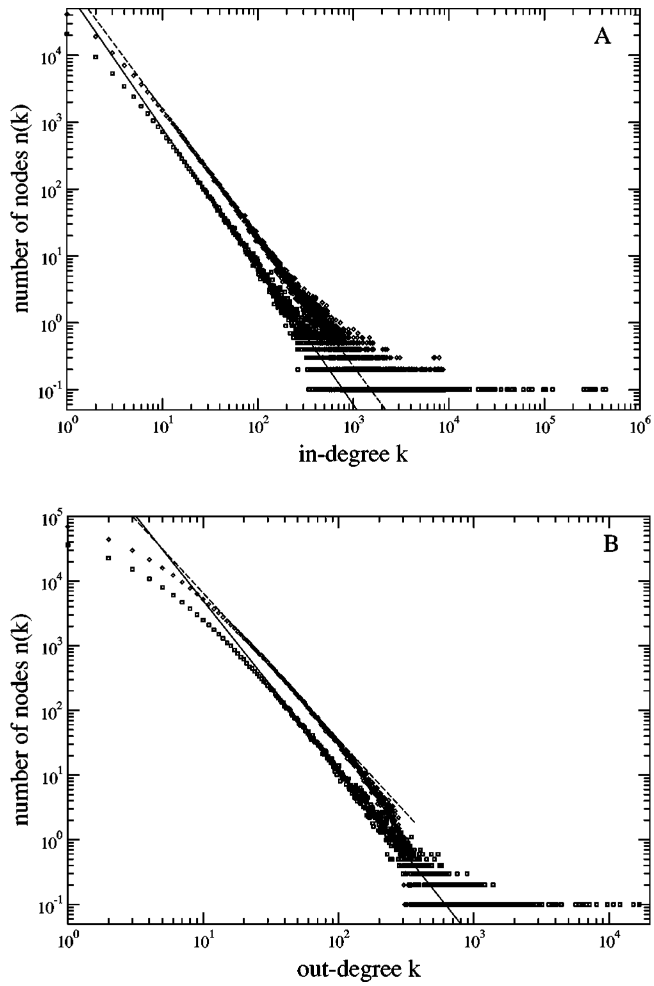


FIG. 2. The profiles of in-degree (a) and out-degree (b) distribution at $p=0.133\,334$, $\lambda=0.75$, and $\mu=3.55$, after a growing process of 10^6 steps. Square symbols are the data for the original Krapivsky-Rodgers-Redner model and diamonds are the data for the revised model. Each data point is the average over 20 (diamonds) or 10 (squares) realizations of the network. The thin solid line has a slope of -2.066 in (a) and -2.626 in (b). The thin dashed line has a slope of -1.925 in (a) and -2.269 in (b). The average number of nodes in the revised network is 133 271, and the average number of edges is 999 984.

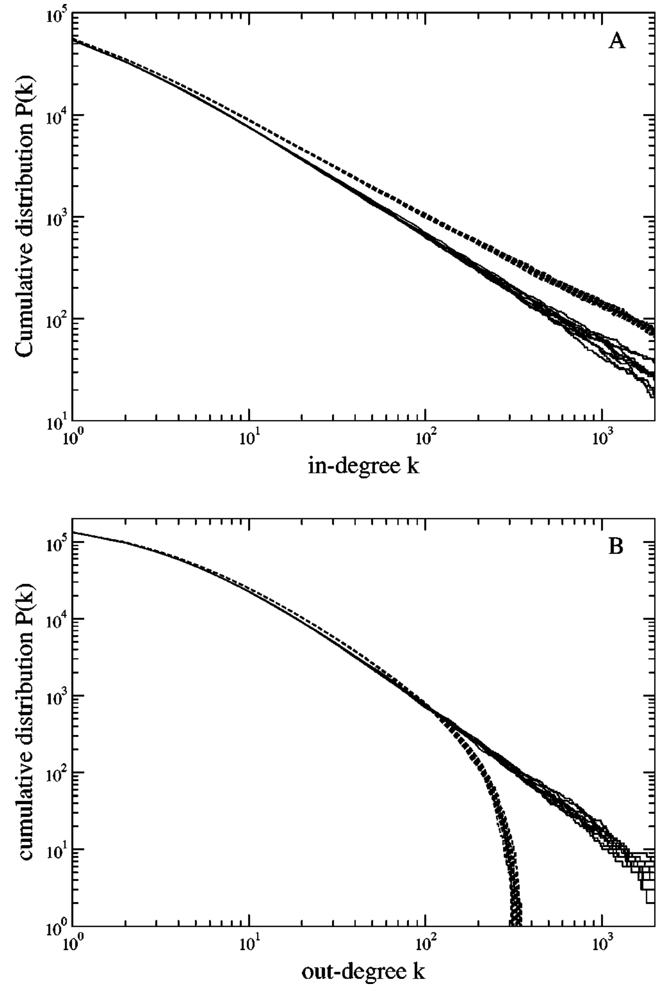


FIG. 3. In-degree (a) and out-degree (b) cumulative distribution for the data sets in Fig. 2. The solid lines correspond to the original model, and the dashed lines to the revised model. From these curves, we estimate $\nu_{\rm in}=2.066\pm0.014$ (the original model) and $\nu_{\rm in}=1.925\pm0.007$ (the revised model). $\nu_{\rm out}=2.626\pm0.036$ for the original model. The out-degree cumulative distribution of the revised model does not fit well to the power law. Therefore, the out-degree scaling exponent is not estimated by this method but by a direct fit to the distribution in Fig. 2(b), resulting in $\nu_{\rm out}=2.269$.

work is the result of $N=10^6$ growing steps.

At $p=0.133\,334$ (a new node will be included on average after every 7.5 steps), $\lambda=0.75$, and $\mu=3.55$, the original model [4] predicts $\nu_{\rm in}=2.1$ and $\nu_{\rm out}=2.7$. From the MC data we obtain that $\nu_{\rm in}=2.066\pm0.014$ and $\nu_{\rm out}=2.626\pm0.036$, in close agreement with the analytical values. At these same parameters, the revised model has $\nu_{\rm in}=1.925\pm0.007$ and $\nu_{\rm out}=2.269$. Thus, exclusion of self-connection and multiconnection leads to decreased values for the scaling exponents. Other quantitative differences are:

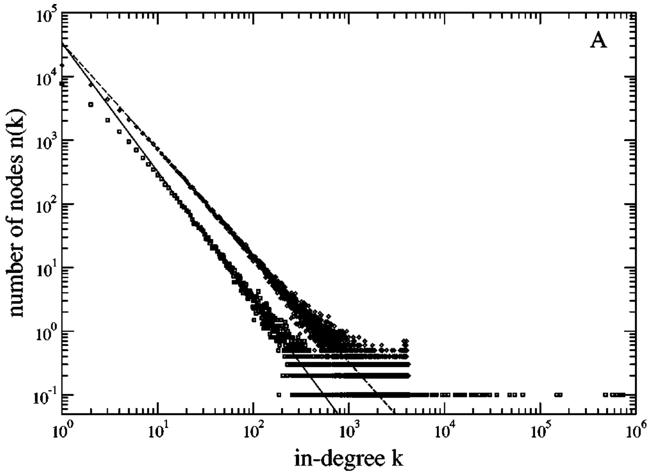
(1) In the revised model there is a cutoff in the in-degree distribution; no node has in-degree $k>9\times10^3$. While in the original model, there are nodes with in degrees as large as $k=4\times10^5$.

(2) In the revised model the out-degree distribution has an exponential cutoff around $k=250$; while in the original model, there are nodes with out degrees as large as $k=1.5\times10^4$.

(3) The value of $n(k)$ is much larger in the revised model than in the original model for a given k (less than the cutoff value). This holds both for the in-degree distribution and for the out-degree distribution.

These observations lead to the following picture: By prohibiting self-connection and multiconnection, edges that originally belonged to just few ‘‘supernodes’’ are now redistributed to the nodes of small or intermediate in and out degrees. Consequently, the number of nodes with small and intermediate node degrees increases considerably, resulting in a smaller scaling exponent in the power-law decrease of the distribution and a cutoff in the tail of this distribution.

In Fig. 4 we demonstrate the simulation result when the probability of node addition is changed to $p=0.05$ (a new node will be included on average after every 20 steps) while



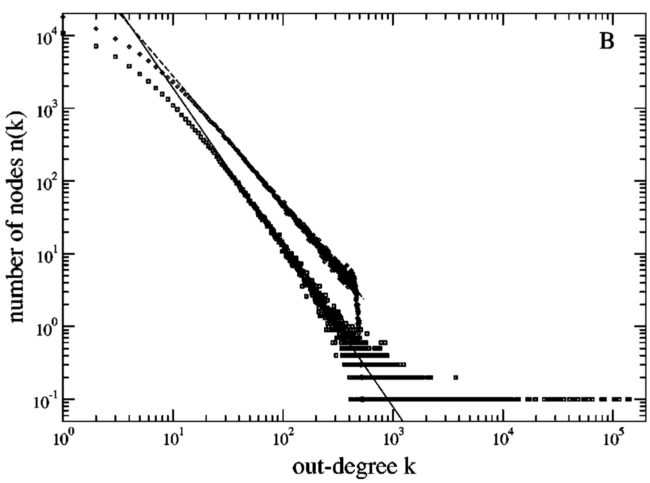


FIG. 4. The profiles of in-degree (a) and out-degree (b) distributions at $p=0.05$, $\lambda=0.75$, $\mu=3.55$, and 10^6 growing steps. Squares (averaged over 10 realizations) correspond to the original model and diamonds (averaged over 20 realizations) to the revised model. The thin solid line has slope -2.018 in (a) and -2.190 in (b). The thin dashed line has slope -1.672 in (a) and -1.764 in (b). In the revised network, the average number of nodes is 49 831 and the average number of edges is 999 860.

for the other two parameters the same values are kept as in Fig. 2. In these parameters, the original model predicts $\nu_{\mathrm{in}}=2.04$ (theory) and 2.018 ± 0.015 (MC) and $\nu_{\mathrm{out}}=2.24$ (theory) and 2.190 ± 0.014 (MC); while the revised model has an in-degree exponent $\nu_{\mathrm{in}}=1.672\pm0.003$ and an out-degree exponent $\nu_{\mathrm{out}}=1.764$, both of which are markedly smaller than 2.

Therefore, the exclusion of self-connection and multiconnection can change the scaling exponent of the scale-free network dramatically when each node has a relatively large average node degree. It can be anticipated that similar behavior will be observed when the initial attractiveness parameters, λ and μ, of each node are varied. It is therefore possible for the present model to explain networks with scaling exponent $\nu<2$.

However, could the observation of scaling exponents $1<\nu<2$ be an artifact caused by finite-size effects? In Fig. 5 the calculated in-degree scaling exponent is plotted as a function of the total growth steps. This figure strongly indicates that even for an infinite system the scaling exponents will still be less than 2.

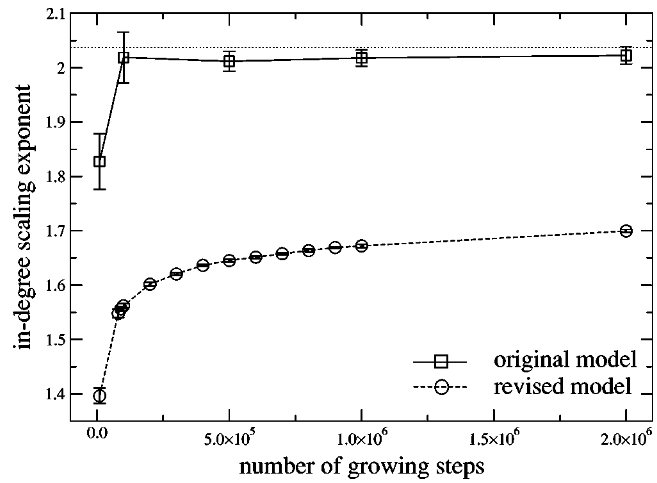


FIG. 5. The relation between the in-degree scaling exponent and growing steps for the original (squares) and the revised (circles) model b. The parameters are set equal to that of Fig. 4, namely, $p=0.05$, $\lambda=0.75$, and $\mu=3.55$. The thin dotted line indicates the theoretical prediction of $\nu_{\mathrm{in}}=2.075$ for the original model.

Adamic and Huberman [7] studied the WWW by mapping each web domain (rather than each web page) as a single node, and they reported an in-degree scaling exponent of $\nu_{\mathrm{in}}=1.94$. Mossa and co-workers, upon their reinterpretation of the WWW data of Barabási and Albert [1], reported an exponent of $\nu=1.25$ [14]. The email network studied in Ref. [15] has a scaling exponent in the range $1.47<\nu<1.82$. And even smaller scaling exponents are reported in several other networks [2]. To attain quantitative agreement to these empirical data, one needs nevertheless more information to help fixing the values of the adjustable parameters.

A persistent property of the revised network model is that there is a rapid decay in the out-degree distribution [which occurs at $k_{\mathrm{out}}\simeq400$ in Fig. 4(b)]. Such a rapid decay was not observed in the original model. This feature is also absent in the in-degree distribution of both the revised and the original models, although the in-degree distribution of the revised model does have a cutoff for large k. Experimentally, it was reported that both the WWW network [14,16] and the email network [15] show exponential cutoff in the node-degree distribution.

V. CLUSTER COEFFICIENTS OF THE GROWING NETWORK

As was mentioned in the Introduction, many real scale-free networks at the same time show small-world behavior [2], having small diameters and being highly clustered. For the original Krapivsky-Rodgers-Redner model, it has been reported in Ref. [10] that the average minimum path scales as $\ln(N)$. We anticipate this to hold also for the revised model. Here we focus on the clustering characteristics of the revised model system, with the parameters setting equal to those of Fig. 2, i.e., $p=0.133\,334$, $\lambda=0.75$, and $\mu=3.55$, and a total of 10^6 growing steps.

A. Mutual-connection coefficient C_{mutual}

Denote $G_{\mathrm{down}}(\alpha)$ as the *complete* set of nodes that are the "downstream" neighbors of node α, namely, there exists a directed edge from α to each node in $G_{\mathrm{down}}(\alpha)$; similarly we define $G_{\mathrm{up}}(\alpha)$ as the complete set of nodes that are the "upstream" neighbors of node α. $|G_{\mathrm{down}}(\alpha)|$ means the size of set $G_{\mathrm{down}}(\alpha)$.

The mutual-connection coefficient is defined as

$$C_{\mathrm{mutual}}=\frac{\sum_{\alpha} |G_{\mathrm{down}}(\alpha)\cap G_{\mathrm{up}}(\alpha)|/|G_{\mathrm{down}}(\alpha)|}{N}, \qquad (7)$$

which signifies to what extent the downstream neighbors of one node intersect with their upstream neighbors. The value of C_{mutual} averaged over 20 realizations of the growing network is 0.0010. This indicates that the interaction between one node and its "neighbors" in the network is usually not bidirectional. However, this value is still much larger than the value for a random network of the same size ($N=133\,271$) and the same average degree of out-going edges ($\langle k\rangle=7.50$), for which $C_{\mathrm{mutual}}=5.63\times10^{-5}$.

B. Incoming clustering coefficient C_{in}

Suppose a given node α has in-degree $k_{\mathrm{in}}(\alpha)$. The maximal number of edges existing between the nodes in $G_{\mathrm{up}}(\alpha)$ is $k_{\mathrm{in}}(\alpha)[k_{\mathrm{in}}(\alpha)-1]$. Denote $i_{\mathrm{actual}}(\alpha)$ as the actual number of edges existing between these edges. We define the incoming clustering coefficient as

$$C_{\mathrm{in}}=\frac{\sum_{\alpha}' \, i_{\mathrm{actual}}(\alpha)/[k_{\mathrm{in}}(\alpha)(k_{\mathrm{in}}(\alpha)-1)]}{N'}, \qquad (8)$$

where Σ_{α}' indicates summation over all the nodes whose incoming edges are larger than 1, and N' is the total number of nodes with this property. We find that $C_{\mathrm{in}}=0.0044$. This value indicates that the degree of cliqueness of the upstream neighbors of a given node is usually very small. For a completely random graph, $C_{\mathrm{in}}=5.63\times10^{-5}$.

C. Outgoing clustering coefficient C_{out}

The definition of the outgoing clustering coefficient C_{out} is similar to that of C_{in}. Suppose a particular node α has $k_{\mathrm{out}}(\alpha)$ outgoing edges, and $i_{\mathrm{actual}}(\alpha)$ is the total number of edges between the nodes in $G_{\mathrm{down}}(\alpha)$, then

$$C_{\mathrm{out}}=\frac{\sum_{\alpha}' \, i_{\mathrm{actual}}(\alpha)/[k_{\mathrm{out}}(\alpha)(k_{\mathrm{out}}(\alpha)-1)]}{N''}, \qquad (9)$$

where N'' is the total number of nodes whose out degree is larger than 1. We find $C_{\mathrm{out}}=0.229$ for the present growing network. Compared with the small values of C_{mutual} and C_{in}, such a large value of C_{out} is surprising. It suggests that the average interaction between two nodes belonging to the same downstream group of a given node is considerably strong. How to understand this kind of asymmetry, namely, $1\sim C_{\mathrm{out}}\gg C_{\mathrm{in}}$? We suggest the following possibility: In the network, there are some nodes that are so popular that a large population of the whole nodes will have an edge pointing to them [see Fig. 2(a)]. Consequently, these nodes will have great possibility to belong to the downstream group of any particular node, and they will also have great possibility to be pointed to by other members of this group, making C_{out} to be proportional to unity. However, the number of nodes decays quickly when the out degree increases to about 250 [see Fig. 2(b)]. Therefore, in the network there is no node that is so "generous" that it points to a large population of the whole network. This may make the value of C_{in} small. In other words, it might be the existence of a steep cutoff in the out-degree profile that accounts for the difference in the clustering coefficients C_{out} and C_{in}.

D. Triangle coefficient C_{triangle}

For a particular node α, suppose node $\beta\in G_{\mathrm{down}}(\alpha)$. Then $i_{\triangle}(\alpha,\beta)=|G_{\mathrm{down}}(\alpha)\cap G_{\mathrm{down}}(\beta)|$ is the total number of nodes that are pointed to by both α and β. We define

$$C_{\mathrm{triangle}}=\frac{\sum_{\alpha} \sum_{\beta\in G_{\mathrm{down}}(\alpha)} i_{\triangle}(\alpha,\beta)/k_{\mathrm{out}}(\beta)}{N}. \qquad (10)$$

The triangle coefficient C_{triangle} signifies the extent, if there is a directed edge from node α to node β and there is a directed edge from β to node γ, there will also be a directed node from α to γ. Calculation revealed that $C_{\mathrm{triangle}}\simeq0.011$.

VI. CONCLUSION AND DISCUSSION

In this work we have used a revised Krapivsky-Rodgers-Redner model to investigate the degree distribution of growing random network and to investigate whether such a kind of growing network could be regarded as a small-world network. After excluding the possibility of self-connection [Fig. 1(a)] and requiring that there is at most one directed edge from one node to any another node [Fig. 1(b)], the Monte Carlo simulation demonstrated that scale-free network with degree distribution coefficient ν less than 2 can be generated. And it is also revealed that the average interactions between the nodes that belong to the downstream group of a particular node is very strong, suggesting that the growing network at the same time forms a "small world." The strong interaction in the downstream group of a particular node was suggested to be closely related to the existence of several "popular" nodes that are pointed to by a large fraction of the total population in the node system. Previous efforts often predicted that the scaling exponent ν should be greater than 2, and there are still not many efforts to understand why many scale-free networks are at the same time small-world networks. It is hoped that the present work will help to improve our understanding of the occurrence of scale-free networks with $\nu < 2$ and to improve our understanding of the close relationship between scale-free and small-world networks.

The present work suggests that, by excluding the possibility of self-connection and multiconnection, many of those edges that were associated with several nodes of extremely large in or out degrees in the original Krapivsky-Rodgers-Redner network, are now redistributed to nodes of small or intermediate degrees. This may explain why a dramatic decrease in scaling exponents could be observed.

ACKNOWLEDGMENTS

This work was financially supported by the Alexander von Humboldt foundation. The author is grateful for the hospitality of Professor Reinhard Lipowsky.

[1] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] If the node-degree distribution looks like a power law with the exponent lesser than 2, its tail must have a different exponent or have some sort of cutoff, since the average number of edges connected to one node should be finite. A cutoff in the distribution is observed in the present work.
[4] P.L. Krapivsky, G.J. Rodgers, and S. Redner, Phys. Rev. Lett. **86**, 5401 (2001).
[5] P.L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[6] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[7] L.A. Adamic and B.A. Huberman, Science **287**, 2115 (2000).
[8] S.N. Dorogovtsev and J.F.F. Mendes, Phys. Rev. E **63**, 025101(R) (2001).
[9] D.J. Watts and S.H. Strogatz, Nature (London) **393**, 440 (1998).
[10] G.J. Rodgers and K. Darby-Dowman, Eur. Phys. J. B **23**, 267 (2001).
[11] M.E.J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
[12] J.M. Montoya and R.V. Solé, J. Theor. Biol. **214**, 405 (2002).
[13] B. Stroustrup, *C++ Programming Language*, 3rd ed. (Addison-Wesley, Reading, MA, 1997).
[14] S. Mossa, M. Barthélémy, H.E. Stanley, and L.A.N. Amaral, Phys. Rev. Lett. **88**, 138701 (2002).
[15] H. Ebel, L.-I. Mielsch, and S. Bornholdt, e-print cond-mat/0201476.
[16] B.A. Huberman and L.A. Adamic, Nature (London) **401**, 131 (1999).