Eur. Phys. J. B **58**, 337–344 (2007)
DOI: 10.1140/epjb/e2007-00229-9

THE EUROPEAN PHYSICAL JOURNAL B

Evolving pseudofractal networks

Zhongzhi Zhang[1,2,a], Shuigeng Zhou[1,2,b], and Lichao Chen[1,2]

[1] Department of Computer Science and Engineering, Fudan University, Shanghai 200433, P.R. China
[2] Shanghai Key Lab of Intelligent Information Processing, Fudan University, Shanghai 200433, P.R. China

Received 17 January 2007 / Received in final form 6 July 2007
Published online 8 September 2007 –

Abstract. We present a family of scale-free network model consisting of cliques, which is established by a simple recursive algorithm. We investigate the networks both analytically and numerically. The obtained analytical solutions show that the networks follow a power-law degree distribution, with degree exponent continuously tuned between 2 and 3. The exact expression of clustering coefficient is also provided for the networks. Furthermore, the investigation of the average path length reveals that the networks possess small-world feature. Interestingly, we find that a special case of our model can be mapped into the Yule process.

PACS. 89.75.-k Complex systems – 89.75.Fb Structures and organization in complex systems – 05.10.-a Computational methods in statistical physics and nonlinear dynamics

1 Introduction

Over the last few years, it has been suggested that a lot of social, technological, biological, and information networks share the following three striking statistical characteristics [1–4]: power-law degree distribution [5], high clustering coefficient [6], and small average path length (APL). Power-law degree distribution indicates that the majority of nodes (vertices) in such networks have only a few connections to other nodes, whereas some nodes are connected to many other nodes in the network. Large clustering coefficient implies that nodes having a common neighbor are far more likely to be linked to each other than are two nodes selected randomly. Short APL shows that the expected number of links (edges) needed to pass from one arbitrarily selected node to another one is low, that is, APL grows logarithmically with the number of nodes or slower.

In order to mimic such complex real-life systems, a wide variety of models have been proposed [1–4], among which the most well-known successful attempts are the Watts and Strogatz's (WS) small-world network model [6] and Barabási and Albert's (BA) scale-free network model [5], which have attracted an exceptional amount of attention from a wide circle of researchers and started an avalanche of research on the models of systems within the physics community. After that, a considerable number of other models and mechanisms, which may represent processes more realistically taking place in real-life systems, have been developed. These mainly include nonlinear preferential attachment [7], initial attractiveness [8], edge rewiring [9] and removal [10], aging and cost [11], competitive dynamics [12], duplication [13], weight [14,15], geographical constraint [16,17], Apollonian packing [18–23] and so forth. Today, modeling complex systems with small-world and scale-free characteristics is still an important issue.

Recently, Dorogovtsev, Goltsev, and Mendes have demonstrated that scale-free behavior and small-world effect can be excellently modeled by using pure mathematical objects and methods to construct a deterministic graph [24], called pseudofractal scale-free network (PSFN) which was extended by Comellas et al. in [25]. PSFN has drawn much attention from the scientific community, many dynamical processes taking place in PSFN have been intensively studied, including synchronization [27], diffusion [28], and opinion formation [29]. The PSFN is a regular deterministic network in a certain sense without statistical mechanics for consideration. In [30], Dorogovtsev, Mendes and Samukhin proposed a random growing network, which we call random pseudofractal scale-free network (RPSFN). The PSFN and RPSFN may provide valuable insight into some particular real-life networks.

In this paper, we propose a general scenario for constructing evolving pseudofractal networks (EPNs) governed by three parameters m, p, and q, which control the relevant network characteristics. The EPN unifies the PSFN and RPSFN to the same framework, i.e. the PSFN and RPSFN are special cases of EPN. In addition to PSFN and RPSFN, the EPN also includes many other models as its particular cases. More interestingly, one particular

[a] e-mail: zhangzz@fudan.edu.cn
[b] e-mail: sgzhou@fudan.edu.cn

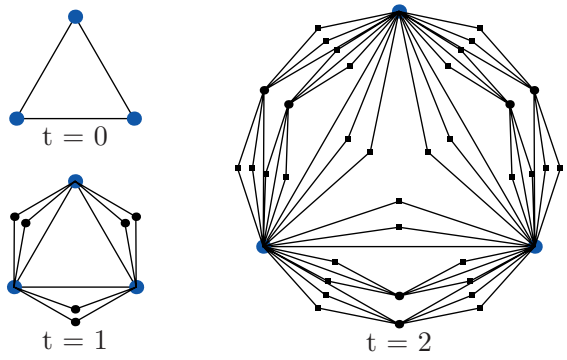

Fig. 1. Illustration of a deterministically growing network in the case of $m = 2$, $p = 1$, and $q = 2$, showing the first three steps of growing process.

case of EPNs can be mapped into the Yule process. The growing EPNs are composed of cliques, and result in a power-law degree distribution with degree exponent changeable between 2 and 3, a very large clustering coefficient, and a small-world feature.

2 Network construction

We construct the networks in a recursive manner and denote the networks after t generations by $Q(q,t)$, $q \geq 2, t \geq 0$. Then the network construction process is as follows: For $t = 0$, $Q(q,0)$ is a complete graph K_{q+1} (or $(q+1)$-clique). For $t \geq 1$, $Q(q,t)$ is obtained from $Q(q,t-1)$. For each of the existing subgraphs of $Q(q,t-1)$ that is isomorphic to q-clique, with probability p $(0 < p \leq 1)$, m (m is a positive integer) new vertices are created, and each is connected to all the vertices of this subgraph. The growing process is repeated until the network reaches a desired size. Figure 1 shows the network growing process for a particular case of $m = 2$, $p = 1$, and $q = 2$.

There are at least five limiting cases of our model listed below. (i) When $m = 1$, $p = 1$, and $q = 2$, the networks are exactly the same as the pseudofractal scale-free network (PSFN) [24]. (ii) When $m = 1$, $p \to 0$ (but $p \neq 0$), and $q = 2$, our model is reduced to the random pseudofractal scale-free network (RPSFN) [30]. (iii) When $m = 1$, $0 < p \leq 1$, and $q = 2$, our networks coincide with the stochastically growing scale-free network described in [31]. (iv) When $m = 1$, $p = 1$, and $q \geq 2$, our networks reduce to the recursive graphs discussed in [25]. (v) When $p = 1$ and $q \geq 2$, our networks turn out to be the graphs introduced in [26]. Thus, varying parameters m, p, and q, we can study many crossovers between these limiting cases.

Next we compute the numbers of nodes and links in $Q(q,t)$. Let $L_v(t)$, $L_e(t)$ and $K_{q,t}$ be the numbers of vertices, edges and q-cliques created at step t, respectively. Note that the addition of each new node leads to q new q-cliques and q new edges. So, we have $L_e(t) = K_{q,t} = qL_v(t)$. Then, at step 1, we add expected $L_v(1) = mp(q+1)$ new nodes and $L_e(1) = mpq(q+1)$ new edges to $Q(q,0)$. After simple calculations, one can obtain that at $t_i (t_i > 1)$ the numbers of newly born nodes and edges are $L_v(t_i) = mp(q+1)(1+mpq)^{t_i-1}$ and $L_e(t_i) = mpq(q+1)(1+mpq)^{t_i-1}$, respectively. Thus the average number of total nodes N_t and edges E_t present at step t is

$$N_t = \sum_{t_i=0}^{t} L_v(t_i) = \frac{(q+1)[(mpq+1)^t + q - 1]}{q} \quad (1)$$

and

$$E_t = \sum_{t_i=0}^{t} L_e(t_i) = \frac{(q+1)[2(mpq+1)^t + (q-2)]}{2}, \quad (2)$$

respectively. So for large t, The average degree $\overline{k}_t = \frac{2E_t}{N_t}$ is approximately $2q$.

3 Topological properties

Topology properties are of fundamental significance to understand the complex dynamics of real-life systems. Here we focus on three important characteristics: degree distribution, clustering coefficient, and average path length, which are determined by the tunable model parameters m, p, and q.

3.1 Degree distribution

Degree distribution is one of the most important statistical characteristics of a network. Firstly, we follow the method that was introduced in [20,23] for the calculation of degree distribution for the general case; then, we use the rate-equation approach [32] to get the degree distribution for some limiting cases.

3.1.1 General case

When a new node i is added to the networks at step t_i, it has degree q and forms q q-cliques. Let $L_q(i,t)$ be the number of q-cliques at step t that will possibly create new nodes connected to the node i at step $t+1$. At step t_i, $L_q(i,t_i) = q$. By construction, we can see that in the subsequent steps each new neighbor of i generates $q-1$ new q-cliques with i as one node of them. Then at step $t_i + 1$, there are mpq new nodes which forms $mpq(q-1)$ new q-cliques containing i. Let $k_i(t)$ be the degree of i at step t. We can easily find following relation for $t > t_i + 1$:

$$\Delta k_i(t) = k_i(t) - k_i(t-1) = mpL_q(i,t-1) \quad (3)$$

and

$$L_q(i,t) = L_q(i,t-1) + (q-1)\Delta k_i(t). \quad (4)$$

From the above two equations, we can derive: $L_q(i,t+1) = L_q(i,t)[1 + mp(q-1)]$. Since $L_q(i,t_i) = q$, we have

$L_q(i,t) = q[1+mp(q-1)]^{t-t_i}$ and $\Delta k_i(t) = mpq[1+mp(q-1)]^{t-t_i-1}$. Then the degree $k_i(t)$ of node i at time t is

$$k_i(t) = k_i(t_i) + \sum_{t_h=t_i+1}^{t} \Delta k_i(t_h) = q\left(\frac{[1+mp(q-1)]^{t-t_i}+q-2}{q-1}\right). \quad (5)$$

Since the degree of each node has been obtained explicitly as in equation (5), we can get the degree distribution via its cumulative distribution [3], i.e., $P_{cum}(k) \equiv \frac{1}{N_t}\sum_{k'\geq k} N(k',t) \sim k^{1-\gamma}$, where $N(k',t)$ denotes the number of nodes with degree k'. The detailed analysis is given as follows. For a degree k

$$k = q\left(\frac{[1+mp(q-1)]^{t-s}+q-2}{q-1}\right), \quad (6)$$

there are $L_v(s) = mp(q+1)(1+mpq)^{s-1}$ nodes with this exact degree, all of which were born at step s. All nodes born at time s or earlier have this or a higher degree. So we have

$$\sum_{k'\geq k} N(k',t) = \sum_{a=0}^{s} L_v(a) = \frac{(q+1)[(mpq+1)^s+q-1]}{q}.$$

As the total number of nodes at step t is given in equation (1) we have

$$\left(\frac{[1+mp(q-1)]^{t-s}+q-2}{1-1/q}\right)^{1-\gamma} = \frac{\frac{(q+1)[(mpq+1)^s+q-1]}{q}}{\frac{(q+1)[(mpq+1)^t+q-1]}{q}}.$$

Therefore, for large t we obtain

$$\left[[1+mp(q-1)]^{t-s}\right]^{1-\gamma} = (1+mpq)^{s-t} \quad (7)$$

and

$$\gamma \approx 1 + \frac{\ln(1+mpq)}{\ln[1+mp(q-1)]}. \quad (8)$$

Thus, the degree exponent γ is a continuous function of m, p and q, and belongs to the interval [2,3], coinciding with the empirically found results. In some limiting cases, equation (8) recovers the results previously obtained in references [24–26,30,31]. Figure 2 shows, on a logarithmic scale, the scaling behavior of the cumulative degree distribution $P_{cum}(k)$ for different values of p in the case of $m = 1$ and $q = 2$. Simulation results agree very well with the analytical ones.

3.1.2 Rate equation for some limiting cases

When $m = 1$ and $p \to 0$ (but $p \neq 0$), our model turns out to be the graph, which we call expanded random pseudofractal scale-free network (ERPSFN) that evolves as follows (see [31] for interpretation): starting with a $(q+1)$-clique ($t = 0$), at each time step, we choose an existing

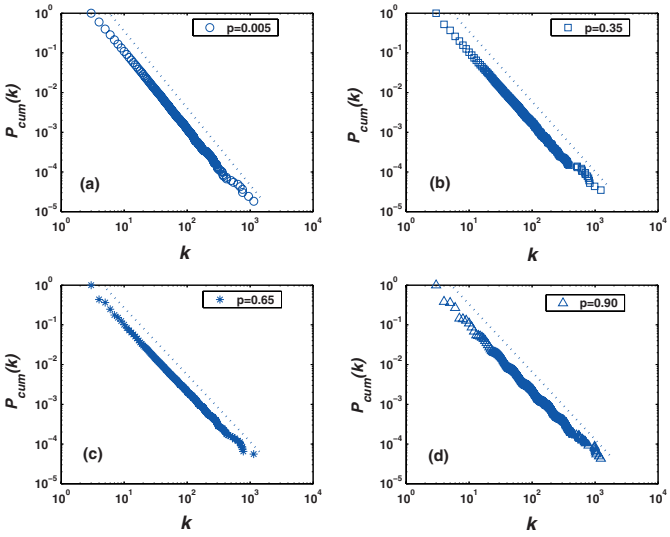

Fig. 2. The cumulative degree distribution $P_{cum}(k)$ at various p values for the case of $m = 1$ and $q = 2$. The circles (a), squares (b), stars (c), and triangles (d) denote the simulation results for networks with different evolutionary steps $t = 1350$, $t = 25$, $t = 16$, and $t = 13$, respectively. The four straight lines are the theoretical results of $\gamma(m,p,q)$ as provided by equation (8). All data are from the average of 50 independent runs.

q-clique, then we add a new node and join it to all the nodes of the selected q-clique. When $q = 2$, the particular model is exactly the random pseudofractal scale-free network (RPSFN) [30].

In fact, the expanded random pseudofractal scale-free network can be easily mapped into the *Yule process* [33,34], which was inspired by observations of the statistics of biological taxa. The Yule process can be prescribed mathematically as follows: we measure the passage of time by the number of genera. At each time step one new species founds a new genus, thereby increasing the number of genera by 1, and q other species are added to various pre-existing genera which are selected in proportion to the number of species they already have. Let the nodes and q-cliques in ERPSFN correspond to genera and species, respectively, then the mapping from ERPSFN to the Yule process is established. From this perspective, our model may find some applications in biological systems. Next, we show that the degree distribution of ERPSFN is power-law with the same degree exponent as the Yule distribution.

Since the size of ERPSFN is incremented by one with each step, here we use the step value t to represent a node created at this step. Furthermore, after a new node is added to the network, the number of q-cliques increases by q. We can see easily that at step t, the network consists of $N = t+q+1$ nodes and $N_q = qN - q^2 + 1$ q-cliques.

One can analyze the degree distribution mathematically as follows. Given a node, when it is born, it has degree q, and the number of q-cliques containing this node is also q. After that, when its degree increases by one, the number of q-cliques with this node as one of

its components increases by $q-1$, so the number of q-cliques for selection containing a node with degree k is $(q-1)k-q^2+2q$. We denote by $P_{k,N}$ the fraction of nodes with degree k when the network size is N. Thus the number of such nodes is $NP_{k,N}$. Then the probability that the new node happens to be connected to a particular node i having degree k_i is proportional to $(q-1)k_i-q^2+2q$, and so when properly normalized is just $[(q-1)k_i-q^2+2q]/(qN-q^2+1)$. Hence, between the appearance of the Nth and the $(N+1)$th node, the total expected number of nodes with degree k that gain a new link during this interval is

$$\frac{(q-1)k-q^2+2q}{qN-q^2+1}NP_{k,N} \simeq \frac{q-1}{q}kP_{k,N}, \tag{9}$$

which holds for large N. Observe that the number of nodes with degree k will decrease at each time step by exactly this number. At the same time the number increases because of nodes that previously had $k-1$ degrees and now have an extra one. Thus we can write a rate equation [32] for the new number $(N+1)P_{k,N+1}$ of nodes with degree k as:

$$(N+1)P_{k,N+1} = NP_{k,N} + \frac{q-1}{q}\left[(k-1)P_{k-1,N} - kP_{k,N}\right]. \tag{10}$$

The only exception to equation (10) is for nodes having degree q, which instead obey the equation

$$(N+1)P_{q,N+1} = NP_{q,N} + 1 - \frac{q-1}{q}qP_{q,N}, \tag{11}$$

since by construction exactly one new such node appears at each time step. When N approaches infinity, we assume that the degree distribution tends to some fixed value $P_k = \lim_{N\to\infty} P_{N,k}$. Then from equation (11), we have

$$P_q = 1/q. \tag{12}$$

And equation (10) becomes

$$P_k = \frac{q-1}{q}\left[(k-1)P_{k-1} - kP_k\right], \tag{13}$$

from which we can easily obtain the recursive equation

$$P_k = \frac{k-1}{k+1+\frac{1}{q-1}}P_{k-1}, \tag{14}$$

which can be iterated to get

$$\begin{aligned} P_k &= \frac{(k-1)(k-2)\dots q}{(k+1+\frac{1}{q-1})(k+\frac{1}{q-1})\dots(q+2+\frac{1}{q-1})}P_q \\ &= \frac{(k-1)(k-2)\dots(q+1)}{(k+1+\frac{1}{q-1})(k+\frac{1}{q-1})\dots(q+2+\frac{1}{q-1})}, \end{aligned} \tag{15}$$

where equation (12) has been used. This can be simplified further by making use of a handy property of the Γ-function, $\Gamma(a) = (a-1)\Gamma(a-1)$ with $\Gamma(a)$ defined by:

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx. \tag{16}$$

By this property and $\Gamma(1)=1$, we get

$$\begin{aligned} P_k &= \frac{(q+1+\frac{1}{q-1})(q+\frac{1}{q-1})\dots(2+\frac{1}{q-1})}{q(q-1)\dots 1}\frac{\Gamma(k)\Gamma(2+\frac{1}{q-1})}{\Gamma(k+2+\frac{1}{q-1})} \\ &= \frac{(q+1+\frac{1}{q-1})(q+\frac{1}{q-1})\dots(2+\frac{1}{q-1})}{q(q-1)\dots 1}\mathbf{B}(k,2+\frac{1}{q-1}), \end{aligned} \tag{17}$$

where $\mathbf{B}(a,b)$ is the Legendre beta-function, which is defined as

$$\mathbf{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \tag{18}$$

Note that the beta-function has the interesting property that for large values of either of its arguments it itself follows a power law. For instance, for large a and fixed b, $\mathbf{B}(a,b) \sim a^{-b}$. Then we can immediately see that for large k, P_k also has a power-law tail with a degree exponent

$$\gamma = 2 + \frac{1}{q-1}. \tag{19}$$

For $q=2$, $\gamma=3$, which has been obtained previously in [31]. Note that equation (17) is similar to the Yule distribution [35] called by Simon [36].

3.2 Clustering coefficient

In a network if a given node is connected to k nodes, defined as the neighbors of the given node, then the ratio between the number of links among its neighbors and the maximum possible value of such links $k(k-1)/2$ is the clustering coefficient of the given node [6]. The clustering coefficient of the whole network is the average of this coefficient over all nodes in the network.

For our network, the analytical expression of clustering coefficient $C(k)$ for a single node with degree k can be derived exactly. When a node is created it is connected to all the nodes of a q-clique, in which nodes are completely interconnected. So its degree and clustering coefficient are q and 1, respectively. In the following steps, if its degree increases one by a newly created node connecting to it, then there must be $q-1$ existing neighbors of it attaching to the new node at the same time. Thus for a node of degree k, we have

$$C(k) = \frac{\frac{q(q-1)}{2} + (q-1)(k-q)}{\frac{k(k-1)}{2}} = \frac{2(q-1)(k-\frac{q}{2})}{k(k-1)}, \tag{20}$$

which depends on both k and q. For $k \gg q$, the $C(k)$ is inversely proportional to degree k. The scaling $C(k) \sim k^{-1}$ has been found for some network models [18–26,30,31,37], and has also been observed in several real-life networks [37].

Using equation (20), we can obtain the clustering $\overline{C}_t$ of the networks at step t:

$$\overline{C}_t = \frac{1}{N_t}\sum_{r=0}^{t}\frac{2(q-1)(D_r-\frac{q}{2})L_v(r)}{D_r(D_r-1)}, \tag{21}$$

where the sum runs over all the nodes and D_r is the degree of the nodes created at step r, which is given by equation (5).

In the infinite network order limit ($N_t \to \infty$), equation (21) converges to a nonzero value C. Obviously, network clustering coefficient $\overline{C}_t$ is a function of parameters m, p and q. If we fixed

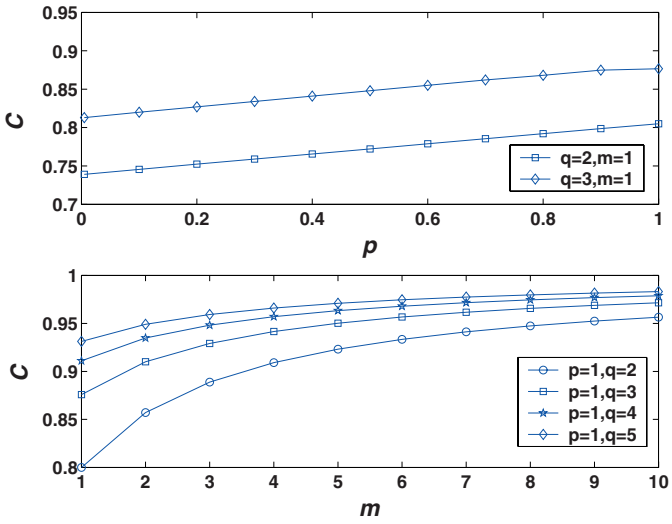

Fig. 3. The dependence of network clustering coefficient C on m, p, and q. Results are averaged over ten network realizations for each datum.

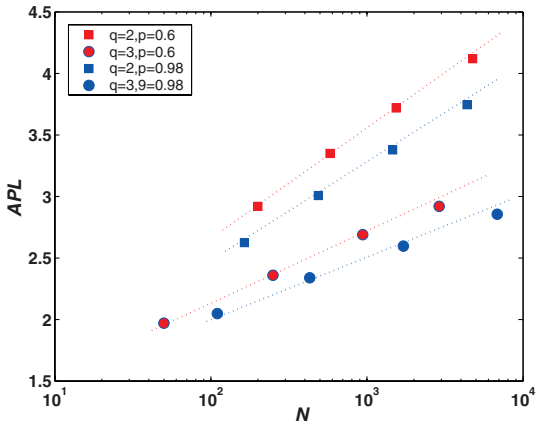

Fig. 4. Semilogarithmic graph of the APL vs the network size N in the special case of $m = 1$. Each datum point is obtained as an average of 50 independent network realizations. The lines are linear functions of $\ln N$.

any two of them, $\overline{C}_t$ increases with the rest. Exactly analytical computations show: in the case $m = 1$ and $q = 2$, when p increases from 0 to 1, C grows from 0.739 [38] to 0.8 [24]; In the case $p = 1$ and $q = 2$, when m increases from 1 to infinite, C grows from 0.8 [24] to 1; Likewise, in the case $m = 1$ and $p = 1$, C increases from 0.8 to 1 when q increases from 2 to infinite, with special values $C = 0.8571$ and $C = 0.8889$ for $q = 3$ and $q = 4$, respectively. Therefore, the average clustering coefficient is very large, which shows the evolving networks are highly clustered. Figure 3 exhibits the dependence of the clustering coefficient C on m, p and q, which agree well with our above conclusions.

From Figures 2 and 3 and equations (8) and (21), one can see that both degree exponent γ and clustering coefficient C depend on the parameter m, p, and q. The mechanism resulting in this should be paid further effort. The fact that there is a biased choice of the cliques at each evolving step may be a possible explanation, see reference [39].

3.3 Average path length

The most important property for a small-world network is a logarithmic average path length (APL) (with the number of nodes). It has obvious implications for the dynamics of processes taking place on networks. Therefore, its study has attracted much attention. Here APL means the minimum number of edges connecting a pair of nodes, averaged over all pairs of nodes. In this subsection, first we give an upper bound of APL for the general case; then, we compute exactly the APL for a particular deterministic network. Both of the obtained values grow logarithmically with the network size.

3.3.1 An upper bound of APL for general case

We denote the network nodes by the time step of their generations, $v = 1, 2, 3, \cdots, N-1, N$. Using $L(N)$ to represent the APL of the our model with system size N, then we have following relation: $L(N) = \frac{2\sigma(N)}{N(N-1)}$, where $\sigma(N) = \sum_{1\le i<j\le N} d_{i,j}$ is the total distance, in which $d_{i,j}$ is the shortest distance between node i and node j. By using the approach similar to that in [17, 21–23], we can evaluate the APL of the present model.

Obviously, when a new node enters the networks, the smallest distances between existing node pairs will not change. Hence we have

$$\sigma(N+1) = \sigma(N) + \sum_{i=1}^{N} d_{i,N+1}. \tag{22}$$

Equation (22) can be approximately represented as:

$$\sigma(N+1) = \sigma(N) + N + (N-q)L(N-q+1), \tag{23}$$

where

$$(N-q)L(N-q+1) = \frac{2\sigma(N-q+1)}{N-q+1} < \frac{2\sigma(N)}{N}. \tag{24}$$

Equations (23) and (24) provide an upper bound for the variation of $\sigma(N)$ as

$$\frac{d\sigma(N)}{dN} = N + \frac{2\sigma(N)}{N}, \tag{25}$$

which yields

$$\sigma(N) = N^2(\ln N + \omega), \tag{26}$$

where ω is a constant. As $\sigma(N) \sim N^2 \ln N$, we have $L(N) \sim \ln N$.

Note that equation (25) was deduced from an inequality, which implies that the increasing tendency of $L(N)$ is at most as $\ln N$ with N. Thus, our model exhibits the presence of small-world property. In Figure 4, we show the dependence of the APL on system size N for different p and q in the case of $m = 1$. From Figure 4, one can see that for fixed q, APL decreases with increasing p; and for fixed p, APL is a decreasing function of q. When network size N is small, APL is a linear function of $\ln N$; while N becomes large, APL increases slightly slower than $\ln N$. So the simulation results are in agreement with the analytical prediction. It should be noted that in our model, if we fix p and q, considering other values of m greater than 1, then the APL will increase more slowly than in the case $m = 1$ as in those cases the larger m is, the denser the network becomes.

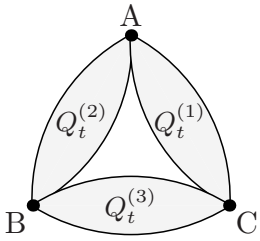

Fig. 5. The network after $t+1$ generations, Q_{t+1}, is obtained by joining three copies of generation t (i.e. $Q_t^{(1)}, Q_t^{(2)}, Q_t^{(3)}$) at the three nodes of highest degree, denoted by A, B, C.

3.3.2 Exact result of APL for a special case

For $p = 1$, the networks are deterministic, which allows one to calculate the APL analytically. Here we only consider a particular case of $m = 1$, $p = 1$, and $q = 2$, which we denote after t generations by Q_t. Then the average path length of Q_t is defined to be:

$$\bar{d}_t = \frac{D_t}{N_t(N_t-1)/2}\,, \tag{27}$$

where

$$D_t = \sum_{i,j\in Q_t} d_{i,j}\,. \tag{28}$$

The deterministic recursive construction of this particular network has a self-similar structure that allows us to exactly calculate $\bar{d}_t$ by following a similar approach introduced in [40]. As shown in Figure 5, the network Q_{t+1} may be obtained by joining at the hubs (the most connected nodes) three copies of Q_t, which we label $Q_t^{(\alpha)}$, $\alpha = 1, 2, 3$ [28]. Then one can write the sum over all shortest paths D_{t+1} as

$$D_{t+1} = 3D_t + \Delta_t, \tag{29}$$

where Δ_t is the sum over all shortest paths whose endpoints are not in the same Q_t branch. The solution of equation (29) is

$$D_t = 3^{t-1}D_1 + \sum_{\tau=1}^{t-1} 3^{t-\tau-1}\Delta_\tau\,. \tag{30}$$

The paths that contribute to Δ_τ must all go through at least one of the three hubs (A, B, C) where the three different Q_t branches are joined. Below we give the analytical expression for Δ_t named the crossing paths, which is given by

$$\Delta_t = \Delta_t^{1,2} + \Delta_t^{2,3} + \Delta_t^{1,3}\,, \tag{31}$$

where $\Delta_t^{\alpha,\beta}$ denotes the sum of all shortest paths with endpoints in $Q_t^{(\alpha)}$ and $Q_t^{(\beta)}$. It should be noted that $\Delta_t^{\alpha,\beta}$ excludes the paths where either endpoint is the hub they have in common, and includes only one of the paths from the unshared hub in one Q_t branch (e.g. $Q_t^{(\alpha)}$) to all non-hub nodes in the other Q_t branch (e.g. $Q_t^{(\beta)}$).

By symmetry, $\Delta_t^{1,2} = \Delta_t^{2,3} = \Delta_t^{1,3}$, we have

$$\Delta_t = 3\Delta_t^{1,2}\,, \tag{32}$$

where $\Delta_t^{1,2}$ is given by the sum

$$\Delta_t^{1,2} = \sum_{\substack{i\in Q_t^{(1)},\ j\in Q_t^{(2)}\\ i\neq A,C,\ j\neq A}} d_{i,j}. \tag{33}$$

In order to find $\Delta_t^{1,2}$, we define

$$\begin{aligned}
d_t^{\rm tot} &\equiv \sum_{Z\in Q_t^{(1)}} d_{Z,A},\\
d_t^{\rm near} &\equiv \sum_{\substack{Z\in Q_t^{(1)}\\ d_{Z,A}<d_{Z,C}}} d_{Z,A}\,, \qquad N_t^{\rm near} \equiv \sum_{\substack{Z\in Q_t^{(1)}\\ d_{Z,A}<d_{Z,C}}} 1\,,\\
d_t^{\rm mid} &\equiv \sum_{\substack{Z\in Q_t^{(1)}\\ d_{Z,A}=d_{Z,C}}} d_{Z,A}\,, \qquad N_t^{\rm mid} \equiv \sum_{\substack{Z\in Q_t^{(1)}\\ d_{Z,A}=d_{Z,C}}} 1\,,\\
d_t^{\rm far} &\equiv \sum_{\substack{Z\in Q_t^{(1)}\\ d_{Z,A}>d_{Z,C}}} d_{Z,A}\,, \qquad N_t^{\rm far} \equiv \sum_{\substack{Z\in Q_t^{(1)}\\ d_{Z,A}>d_{Z,C}}} 1\,,
\end{aligned} \tag{34}$$

where $Z \neq A$ and $Z \neq C$. Then we can easily have $d_t^{\rm tot} = d_t^{\rm near} + d_t^{\rm mid} + d_t^{\rm far}$ and $N_t = N_t^{\rm near} + N_t^{\rm mid} + N_t^{\rm far} + 2$. By symmetry $N_t^{\rm near} = N_t^{\rm far}$. Thus, by construction, we obtain

$$\begin{cases} N_t = 2\,N_t^{\rm near} + N_t^{\rm mid} + 2,\\ N_{t+1}^{\rm mid} = N_t^{\rm mid} + 2\,N_t^{\rm near} + 1. \end{cases} \tag{35}$$

Using these two relation and considering the initial values, we obtain partial quantities in equation (34) as

$$\begin{cases} N_t^{\rm far} = N_t^{\rm near} = \frac{1}{6}\left(-3 + 3^{t+1}\right),\\ N_t^{\rm mid} = \frac{1}{6}\left(3 + 3^{t+1}\right). \end{cases} \tag{36}$$

Now we return to the quantity $\Delta_t^{1,2}$ which can be further decomposed into the sum of four terms as

$$\begin{aligned}
\Delta_t^{1,2} &= \sum_{\substack{i\in Q_t^{(1)},\ j\in Q_t^{(2)}\\ i\neq A,C,\ j\neq A}} d_{i,j}\\
&= \sum_{\substack{i\in Q_t^{(1)},\ j\in Q_t^{(2)},\ i\neq A,C,\ j\neq A,B\\ d_{i,A}>d_{i,C},\ d_{j,A}>d_{j,B}}} (d_{i,C} + d_{j,B} + 1)\\
&\quad + \sum_{\substack{i\in Q_t^{(1)},\ j\in Q_t^{(2)}\\ i\neq A,C,\ j\neq A,B,\ d_{i,A}\le d_{i,C}}} (d_{i,A} + d_{A,j})\\
&\quad + \sum_{\substack{i\in Q_t^{(1)},\ j\in Q_t^{(2)},\ i\neq A,C,\ j\neq A,B\\ d_{i,A}>d_{i,C},\ d_{j,A}\le d_{j,B}}} (d_{i,A} + d_{A,j})\\
&\quad + \sum_{i\in Q_t^{(1)},\ i\neq A,C,\ j=B} (d_{i,A} + 1)\\
&= 2N_t^{\rm near} d_t^{\rm near} + (N_t^{\rm near})^2\\
&\quad + (N_t^{\rm near} + N_t^{\rm mid})(d_t^{\rm near} + d_t^{\rm mid} + d_t^{\rm far})\\
&\quad + (N_t - 2)(d_t^{\rm near} + d_t^{\rm mid})\\
&\quad + (N_t^{\rm near} + N_t^{\rm mid})d_t^{\rm far}\\
&\quad + N_t^{\rm near}(d_t^{\rm near} + d_t^{\rm mid})\\
&\quad + d_t^{\rm near} + d_t^{\rm mid} + d_t^{\rm far} + N_t - 2\,.
\end{aligned} \tag{37}$$

Having $\Delta_n^{1,2}$ in terms of the quantities in equation (34), the next step is to explicitly determine these quantities unresolved.

Since A and C are linked by one edge, for any node i in the network, $d_{i,A}$ and $d_{i,C}$ can differ by at most 1. In addition, considering the self-similar structure of the graph, we can easily

know that at time $t+1$, the quantities $d_{t+1}^{\rm mid}$, $d_{t+1}^{\rm near}$ and $d_{t+1}^{\rm far}$ evolve as

$$\begin{cases} d_{t+1}^{\rm mid} = d_t^{\rm mid} + 2\, d_t^{\rm far} + 1\,, \\ d_{t+1}^{\rm near} = d_t^{\rm mid} + 2\, d_t^{\rm near}\,, \\ d_{t+1}^{\rm far} = d_t^{\rm mid} + 2\, d_t^{\rm far} + N_t^{\rm mid}\,. \end{cases} \tag{38}$$

From these recursive equations we can obtain

$$\begin{cases} d_t^{\rm mid} = 3^{t-2}\,(t+5)\,, \\ d_t^{\rm near} = 3^{t-2}\,(t+2)\,, \\ d_t^{\rm far} = \dfrac{1}{54}\left(2\,(t+1)\cdot 3^{t+1} + 11\cdot 3^{t+1} - 27\right)\,. \end{cases} \tag{39}$$

Substituting the obtained expressions in equations (36) and (39) into equations (37) and (32), the crossing paths Δ_t is obtained to be

$$\Delta_t = \frac{1}{12}\left[(4\,t+13)\,9^t - 9\right]. \tag{40}$$

Inserting equation (40) into equation (30) and using $D_1 = 21$, we have

$$D_t = \frac{1}{8}\left(4\,t\cdot 9^t + 10\cdot 3^t + 11\cdot 9^t + 3\right). \tag{41}$$

Substituting equations (1) and (41) into (27), the exact expression for the average path length is obtained to be

$$\bar{d}_t = \frac{4\,t\cdot 9^t + 10\cdot 3^t + 11\cdot 9^t + 3}{3 + 4\cdot 3^{t+1} + 9^{t+1}}. \tag{42}$$

In the infinite network size limit ($t \to \infty$),

$$\bar{d}_t \simeq \frac{4}{9}t + \frac{11}{9} \sim \ln N_t, \tag{43}$$

which means that the average path length shows a logarithmic scaling with the size of the network.

It should be mentioned that the final expressions contained in equations (42) and (43) were quoted earlier in reference [24] [Eqs. (6) and (7) of that work]. However reference [24] did not provide any of the details of the derivation, so the explicit calculation presented here is pedagogically useful. Moreover, the analytical method may guide and shed light on related studies for other deterministic network models.

4 Conclusions and discussions

In summary, we have proposed and studied a class of evolving networks consisting of cliques. We have obtained the analytical and numerical results for degree distribution, clustering coefficient, as well as the average path length, which are determined by the model parameters and in accordance with large amount of real observations. The networks are scale-free, with degree exponent adjusted continuously between 2 and 3. The clustering coefficient of single nodes has a power-law spectra, the network clustering coefficient is very large and independent of network size. The intervertex separation is small, which increases at most logarithmically as the network size.

In real-life world, many networks consist of cliques. For example, in movie actor collaboration network [6] and science collaborating graph [41], actors acting in the same film or authors signing in the same paper form a clique, respectively. In corporate director network [42], directors as members in the same board constitute a clique. Analogously, in public transport networks [43], bus (tramway, or underground) stops shape a clique if they are consecutive stops on a route, and in the network of concepts in written texts [44], words in each sentence in the text are added to the network as a clique. All these pose a very interesting and important question of how to build evolution models based on this particularity of network component—cliques. Interestingly, our networks, although different from real world, are formed by cliques, this particularity of the composing units may provide a comprehensive aspect to understand some real-life systems. In future, it would be more interesting to establish a model describing real systems consisting of cliques such as actor collaboration network where cliques arise from mutual cooperation [45,46].

Future work should also include studying in detail dynamical processes taking place on our networks, which may provide some original and interesting results to the field. For example, it should be possible to adapt the renormalization-group techniques used in references [40,47,48] for studying the Ising model to examine cooperative behavior on the general case of our model. There could be a variety of interesting phase transition behaviors as the network structure is modified through the parameters m, p, and q.

The authors would like to thank Zhen Shen for his assistance in preparing the manuscript. This research was supported by the National Basic Research Program of China under grant No. 2007CB310806, the National Natural Science Foundation of China under Grant Nos. 60496327, 60573183, and 90612007, the Postdoctoral Science Foundation of China under Grant No. 20060400162, the Program for New Century Excellent Talents in University of China (NCET-06-0376), and the Huawei Foundation of Science and Technology (YJCB2007031IN).

References

1. R. Albert, A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002)
2. S.N. Dorogvtsev, J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002)
3. M.E.J. Newman, SIAM Rev. **45**, 167 (2003)
4. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwanga, Phys. Rep. **424**, 175 (2006)
5. A.-L. Barabási, R. Albert, Science **286**, 509 (1999)
6. D.J. Watts, H. Strogatz, Nature **393**, 440 (1998)
7. P.L. Krapivsky, S. Redner, F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000)
8. S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000)
9. R.Albert, A.-L. Barabási, Phys. Rev. Lett. **85**, 5234 (2000)
10. S.N. Dorogovtsev, J.F.F. Mendes, EPL **52**, 33 (2000)
11. L.A.N. Amaral, A. Scala, M. Barthélémy, H.E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000)
12. G. Bianconi, A.-L. Barabási, EPL **54**, 436 (2001)
13. F. Chung, L.Y. Lu, T.G. Dewey, D.J. Galas, Biology **10**, 677 (2003)
14. W.X. Wang, B.H. Wang, B. Hu, G. Yan, Q. Ou, Phys. Rev. Lett. **94**, 188702 (2005)

15. Z.Z. Zhang, S.G. Zhou, L.J. Fang, J.H. Guan, Y.C. Zhang, EPL **79**, 38007 (2007)
16. A.F. Rozenfeld, R. Cohen, D. ben-Avraham, S. Havlin, Phys. Rev. Lett. **89**, 218701 (2002)
17. Z.Z. Zhang, L.L. Rong, F. Comellas, J. Phys. A **39**, 3253 (2006)
18. J.S. Andrade Jr, H.J. Herrmann, R.F.S. Andrade, L.R. da Silva, Phys. Rev. Lett. **94**, 018702 (2005)
19. J.P.K. Doye, C.P. Massen, Phys. Rev. E **71**, 016128 (2005)
20. Z.Z. Zhang, F. Comellas, G. Fertin, L.L. Rong, J. Phys. A **39**, 1811 (2006)
21. T. Zhou, G. Yan, B.H. Wang, Phys. Rev. E **71**, 046141 (2005)
22. Z.Z. Zhang, L.L. Rong, F. Comellas, Physica A **364**, 610 (2006)
23. Z.Z. Zhang, L.L. Rong, S.G. Zhou, Phys. Rev. E **74**, 046105 (2006)
24. S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, Phys. Rev. E **65**, 066122 (2002)
25. F. Comellas, G. Fertin, A. Raspaud, Phys. Rev. E **69**, 037104 (2004)
26. Z.Z. Zhang, L.L. Rong, S.G. Zhou, Physica A **377**, 329 (2007)
27. P.G. Lind, J.A.C. Gallas, H.J. Herrmann, Phys. Rev. E **70**, 056207 (2004)
28. E. Bollt, D. Ben-Avraham, New J. Phys. **7**, 26 (2005)
29. M.C. González, A.O. Sousa, H.J. Herrmann, Eur. Phys. J. B **49**, 253 (2006)
30. S.N. Dorogovtsev, J.F.F. Mendes, A.N. Samukhin, Phys. Rev. E **63**, 062101 (2001)
31. S.N. Dorogovtsev, Phys. Rev. E **67**, 045102(R) (2003)
32. P.L. Krapivsky, S. Redner, Phys. Rev. E **63**, 066123 (2001)
33. J.C. Willis, G.U. Yule, Nature **109**, 177 (1922)
34. G.U. Yule, Phil. Trans. R. Soc. B **213**, 21 (1925)
35. M.E.J. Newman, Contemp. Phys. **46**, 323 (2005)
36. H.A. Simon, Biometrika **42**, 425 (1955)
37. E. Ravasz, A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003)
38. A. Barrat, R. Pastor-Satorras, Phys. Rev. E **71**, 036127 (2005)
39. F. Comellas, H.D. Rozenfeld, D. Ben-Avraham, Phys. Rev. E **72**, 046142 (2005)
40. M. Hinczewski, A.N. Berker, Phys. Rev. E **73**, 066126 (2006)
41. M.E.J. Newman, Proc. Natl. Acad. Sci. U.S.A. **97**, 404 (2001)
42. S. Battiston, M. Catanzaro, Eur. Phys. J. B **38**, 345 (2004)
43. J. Sienkiewicz, J.A. Holyst, Phys. Rev. E **72**, 046127 (2005)
44. S.M.G. Caldeira, T.C.P. Lobão, R.F.S. Andrade, A. Neme, J.G.V. Miranda1, Eur. Phys. J. B **49**, 523 (2006)
45. J.-L. Guillaume, M. Latapy, Inf. Process. Lett. **90**, 215 (2004)
46. J.-L. Guillaume, M. Latapy, Physica A **371**, 795 (2006)
47. A. Erbas, A. Tuncer, B. Yücesoy, A.N. Berker, Phys. Rev. E **72**, 026129 (2005)
48. M. Hinczewski, Phys. Rev. E **75**, 061104 (2007)