Mapping Koch curves into scale-free small-world networks

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

(http://iopscience.iop.org/1751-8121/43/39/395101)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 202.120.224.18
The article was downloaded on 02/09/2010 at 01:53

Please note that terms and conditions apply.
Mapping Koch curves into scale-free small-world networks

Zhongzhi Zhang\textsuperscript{1,2}, Shuyang Gao\textsuperscript{1,2}, Lichao Chen\textsuperscript{3}, Shuigeng Zhou\textsuperscript{1,2}, Hongjuan Zhang\textsuperscript{2,4} and Jihong Guan\textsuperscript{5}

\textsuperscript{1} School of Computer Science, Fudan University, Shanghai 200433, People’s Republic of China
\textsuperscript{2} Shanghai Key Lab of Intelligent Information Processing, Fudan University, Shanghai 200433, People’s Republic of China
\textsuperscript{3} Electrical Engineering Department, University of California, Los Angeles, CA 90024, USA
\textsuperscript{4} Department of Mathematics, College of Science, Shanghai University, Shanghai 200444, People’s Republic of China
\textsuperscript{5} Department of Computer Science and Technology, Tongji University, 4800 Cao’an Road, Shanghai 201804, People’s Republic of China

E-mail: zhangzz@fudan.edu.cn, sgzhou@fudan.edu.cn and jhguan@tongji.edu.cn

Received 8 June 2010, in final form 7 July 2010
Published 1 September 2010
Online at stacks.iop.org/JPhysA/43/395101

Abstract
The class of Koch fractals is one of the most interesting families of fractals, and the study of complex networks is a central issue in the scientific community. In this paper, inspired by the famous Koch fractals, we propose a mapping technique converting Koch fractals into a family of deterministic networks called Koch networks. This novel class of networks incorporates some key properties characterizing a majority of real-life networked systems—a power-law distribution with exponent in the range between 2 and 3, a high clustering coefficient, a small diameter and average path length and degree correlations. Besides, we enumerate the exact numbers of spanning trees, spanning forests and connected spanning subgraphs in the networks. All these features are obtained exactly according to the proposed generation algorithm of the networks considered. The network representation approach could be used to investigate the complexity of some real-world systems from the perspective of complex networks.

PACS numbers: 89.75.Hc, 05.10.-a, 89.75.Fb, 61.43.Hv

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

1. Introduction
The past decade has witnessed a great deal of activity devoted to complex networks by the scientific community, since many systems in the real world can be described and characterized
by complex networks [1–4]. Prompted by the computerization of data acquisition and the increased computing power of computers, researchers have done a lot of empirical studies on diverse real networked systems, unveiling the presence of some generic properties of various natural and manmade networks: power-law degree distribution \( P(k) \sim k^{-\gamma} \) with the characteristic exponent \( \gamma \) in the range between 2 and 3 [5], small-world effect including a large clustering coefficient and small average distance [6], and degree correlations [7, 8].

The empirical studies have inspired researchers to construct network models with the aim to reproduce or explain the striking common features of real-life systems [1, 2]. In addition to the seminal Watts–Strogatz’s (WS) small-world network model [6] and Barabási–Albert’s (BA) scale-free network model [5], a considerable number of models and mechanisms have been developed to mimic real-world systems, including initial attractiveness [9], aging and cost [10], fitness model [11], weight or traffic driven evolution [13, 14], geographical constraint [15], accelerating growth [16, 17], coevolution [18] and visibility graph [19], to list a few. Although significant progress has been made in the field of network modeling and has led to a significant improvement in our understanding of complex systems, it is still a fundamental task and of current interest to construct models mimicking real networks and reproducing their generic properties from different angles [20].

In this paper, enlightened by the famous class of Koch fractals, we propose a family of deterministic mathematical networks, called Koch networks, which integrates the observed properties of real networks in a single framework. We derive analytically exact scaling laws for degree distribution, clustering coefficient, diameter, average distance or average path length (APL), degree correlations, even for spanning trees, spanning forests and connected spanning subgraphs. The obtained precise results show that Koch networks have rich topological features: they obey a power-law degree distribution with the exponent lies between 2 and 3; they have a large clustering coefficient and their diameter and APL grow logarithmically with the total number of nodes; and they may be either disassortative or uncorrelated.

This work unfolds an alternative perspective in the study of complex networks. Instead of searching generation mechanisms for real networks, we explore deterministic mathematical networks that exhibit some typical properties of real-world systems. As the classical Koch fractals are important for the understanding of geometrical fractals in real systems [21], we believe that Koch networks could provide valuable insights into real-world systems.

2. Network construction

In order to define the networks, we first introduce a classical fractal—Koch curve, which was proposed by von Koch [22]. The Koch curve, denoted as \( S_1(t) \) after \( t \) generations, can be constructed in a recursive way. To produce this well-known fractal, we begin with an equilateral triangle and let this initial configuration be \( S_1(0) \). In the first generation, we perform the following operations: firstly, we trisect each side of the initial equilateral triangle; secondly, on the middle segment of each side, we construct new equilateral triangles whose interiors lie external to the region enclosed by the base triangle; thirdly, we remove the three middle segments of the base triangle, upon which new triangles were established. Thus, we get \( S_1(1) \). In the second generation, for each line segment in \( S_1(1) \), we repeat the above procedure of three operations to obtain \( S_1(2) \). This process is then repeated for successive generations. As \( t \) tends to infinite, the Koch curve is obtained, and its Hausdorff dimension is \( d_f = \frac{\ln 4}{\ln 3} \) [23]. Figure 1 depicts the structure of \( S_1(2) \).

The Koch curve can be easily generalized to other dimensions by introducing a parameter \( m \) (a positive integer) [23, 24]. The generalization after \( t \) generations is denoted by \( S_m(t) \), which is constructed as follows [23]: start with an equilateral triangle as the initial configuration
Figure 1. The first two generations of the construction for the Koch curve.

$S_m(0)$. In the first generation, we perform the following operations similar to those described in the last paragraph: partition each side of the initial triangle into $2m + 1$ segments, which are consecutively numbered $1, 2, \ldots, 2m, 2m + 1$ from one endpoint of the side to the other; construct a new small equilateral triangle on each even-numbered segment so that the interiors of the new triangles lie in the exterior of the base triangle; remove the segments upon which triangles were constructed. In this way we obtain $S_m(1)$. Analogously, we can get $S_m(t)$ from $S_m(t - 1)$ by repeating recursively the procedure of the above three operations for each existing line segment in generation $t - 1$. In the infinite $t$ limit, the Hausdorff dimension of the generalized Koch curves $d_f = \frac{\ln(4m + 1)}{\ln(2m + 1)}$ [23]. Figure 2 shows the structure of $S_2(2)$.

The generalized Koch curves can be used as a basis of a new class of networks: sides (excluding those deleted) of the triangles of the Koch curves constructed at arbitrary generations are mapped to nodes, which are connected to one another if their corresponding sides in the Koch curves are in contact. For uniformity, the three sides of the initial equilateral triangle of $S_m(0)$ also correspond to three different nodes. We shall call the resultant networks Koch networks. Note that after establishing each side of a triangle constructed at a given generation of the Koch curves, although some segments of it will be removed at subsequent steps, we look at its remaining segments as a whole and map it to only one node. Figures 3 and 4 show two networks corresponding to $S_1(2)$ and $S_2(2)$, respectively.

Obviously, Koch networks have an infinite number of nodes. But in what follows we shall generally consider the network characteristics after a finite number of generations in the development of complete Koch networks. From our analytical results, we can quickly obtain the characteristics of the complete networks by taking the limit of large $t$. However, the numerical results are necessarily limited to networks with finite order (number of all nodes).

3. Generation algorithm

According to the construction process of the generalized Koch curves and the proposed method of mapping from Koch curves to Koch networks, we can introduce an iterative algorithm with
ease to create Koch networks, denoted by $K_{m,t}$ after $t$ generation evolutions. The algorithm is as follows: initially ($t = 0$), $K_{m,0}$ consists of three nodes forming a triangle. For $t \geq 1$, $K_{m,t}$ is obtained from $K_{m,t-1}$ by adding $m$ groups of nodes for each of the three nodes of every existing triangle in $K_{m,t-1}$. Each node group has two nodes. These two new nodes and their ‘mother’ nodes are linked to one another shaping a new triangle. In other words, to obtain $K_{m,t}$ from $K_{m,t-1}$, we replace each of the existing triangles of $K_{m,t-1}$ by the connected clusters on the rightmost side of figure 5. Figures 3 and 4 illustrate the growing process of the networks for two particular cases of $m = 1$ and $m = 2$, respectively. Note that in the peculiar case of $m = 1$, the networks under consideration reduce to the one previously studied in [25].

Let us compute the order and size (number of all edges) of the Koch networks $K_{m,t}$. To this end, we first consider the total number of triangles $L_{\Delta}(t)$ that exist at step $t$. By construction

Figure 2. The first two generations of the construction for the generalized Koch curve in the case of $m = 2$.

Figure 3. The network derived from $S_1(2)$.  

4
(see figure 5), this quantity increases by a factor of \(3m + 1\), i.e. \(L_\Delta(t) = (3m + 1) L_\Delta(t - 1)\). Considering \(L_\Delta(0) = 1\), we have \(L_\Delta(t) = (3m + 1)^t\). Denote \(L_v(t)\) and \(L_e(t)\) as the numbers of nodes and edges created at step \(t\), respectively. Note that each triangle in \(K_{m,t-1}\) will give rise to 6\(m\) new nodes and 9\(m\) new edges at step \(t\); then one can easily obtain \(L_v(t) = 6m L_\Delta(t - 1) = 6m (3m + 1)^{t-1}\) and \(L_e(t) = 9m L_\Delta(t - 1) = 9m (3m + 1)^{t-1}\), both of which hold for arbitrary \(t > 0\). Then, the total numbers of nodes \(N_t\) and edges \(E_t\) present at step \(t\) are

\[
N_t = \sum_{i=0}^{t} L_v(t_i) = 2 (3m + 1)^t + 1 \quad (1)
\]

and

\[
E_t = \sum_{i=0}^{t} L_e(t_i) = 3 (3m + 1)^t, \quad (2)
\]

respectively. Thus, the average degree is

\[
\langle k \rangle = \frac{2 E_t}{N_t} = \frac{6 (3m + 1)^t}{2 (3m + 1)^t + 1}, \quad (3)
\]
which is approximately 3 for large $t$, showing that Koch networks are sparse as most real-life networks [1–4].

4. Topological properties

Now we study some relevant characteristics of the Koch networks $K_{m,t}$, focusing on degree distribution, clustering coefficient, diameter, average distance, degree correlations, spanning trees, spanning forests and connected spanning subgraphs. We emphasize that this is the first analytical study for counting spanning trees, spanning forests and connected spanning subgraphs in scale-free networks.

4.1. Degree distribution

Let $k_i(t)$ be the degree of node $i$ at time $t$. When node $i$ enters the network at step $t_i$ ($t_i \geq 0$), it has a degree of 2, namely $k_i(t_i) = 2$. To determine $k_i(t)$, we first consider the number of triangles involving node $i$ at step $t$ that is denoted by $L/Delta_1(i, t)$. These triangles will give rise to new nodes linked to node $i$ at step $t + 1$. Then at step $t_i$, $L/Delta_1(i, t_i) = 1$. By construction, for any triangle involving node $i$ at a given step, it will lead to $m$ new triangles passing by node $i$ at a next step. Thus, $L/Delta_1(i, t) = (m + 1)L/Delta_1(i, t - 1)$. Considering the initial condition $L/Delta_1(i, t_i) = 1$, we have $L/Delta_1(i, t) = (m + 1)^{t - t_i}$. On the other hand, each triangle passing by node $i$ contains two links connected to $i$; therefore, we have $k_i(t) = 2L/Delta_1(i, t)$. Then we obtain

$$k_i(t) = 2L/Delta_1(i, t) = 2(m + 1)^{t - t_i}.$$  \hspace{1cm} (4)

In this way, at time $t$ the degree of the arbitrary node $i$ of Koch networks has been computed explicitly. From equation (4), it is easy to see that at each step the degree of a node increases $m$ times, i.e.

$$k_i(t) = (m + 1)k_i(t - 1).$$  \hspace{1cm} (5)

Equation (4) shows that the degree spectrum of Koch networks is discrete. Thus, we can get the degree distribution $P(k)$ of the Koch networks via the cumulative degree distribution [3] given by

$$P_{\text{cum}}(k) = \frac{1}{N_t} \sum_{\tau \leq t_i} L/Delta_1(\tau) = \frac{2 \times (3m + 1)^{\frac{1}{2}} + 1}{2 \times (3m + 1)^{\frac{1}{2}} + 1}.$$  \hspace{1cm} (6)

Substituting $t_i = t - \frac{\ln(\frac{k}{2})}{\ln(m + 1)}$ in this expression gives

$$P_{\text{cum}}(k) = \frac{2 \times (3m + 1)^{\frac{1}{2}} \times \left(\frac{k}{2}\right)^{\frac{1}{2}} \ln(\frac{k}{2})}{2 \times (3m + 1)^{\frac{1}{2}} + 1}.$$  \hspace{1cm} (7)

In the infinite $t$ limit, we obtain

$$P_{\text{cum}}(k) = 2 \frac{\ln(\frac{k}{2})}{\ln(m + 1)} \times k^{-\frac{\ln(\frac{k}{2})}{\ln(m + 1)}}.$$  \hspace{1cm} (8)

So the degree distribution follows a power-law form $P(k) \sim k^{-\gamma}$ with the exponent $\gamma = 1 + \frac{\ln(\frac{k}{2})}{\ln(m + 1)}$ belonging to the interval [2, 3]. When $m$ increases from 1 to infinite, $\gamma$ decreases from 3 to 2. It should be stressed that the exponent of degree distribution of most real scale-free networks also lies in the same range between 2 and 3.
4.2. Clustering coefficient

By definition, the clustering coefficient \[ C_i \] of a node \( i \) with degree \( k_i \) is the ratio between the number of triangles \( e_i \) that actually exist among the \( k_i \) neighbors of node \( i \) and the maximum possible number of triangles involving \( i \), \( k_i(k_i - 1)/2 \), namely \( C_i = 2e_i/[k_i(k_i - 1)] \). For Koch networks, we can obtain the exact expression of the clustering coefficient \( C(k) \) for a single node with degree \( k \). By construction, for any given node having a degree \( k \), there are just \( e = \frac{k^2}{4} \) triangles connected with this node; see also equation (4). Hence there is a one-to-one corresponding relation between the clustering coefficient of a node and its degree: for a node of degree \( k \),

\[
C(k) = \frac{1}{k - 1},
\]

which shows a power-law scaling \( C(k) \sim k^{-1} \) in the large limit of \( k \), in agreement with the behavior observed in a variety of real-life systems [26].

After \( t \) step growth, the average clustering coefficient \( C_t \) of the whole network \( K_{m,t} \), defined as the mean of \( C_j \)'s over all nodes in the network, is given by

\[
C_t = \frac{1}{N_t} \sum_{r=0}^{t} \left[ \frac{1}{G_r - 1} \times L_r(r) \right],
\]

where the sum runs over all the nodes of all generations and \( G_r \) is the degree of those nodes created at step \( r \), which is given by equation (4). In the limit of large \( N_t \), equation (10) converges to a nonzero value \( C \), as reported in figure 6. For \( m = 1, 2 \) and 3, \( C \) is 0.82008, 0.88271 and 0.91316, respectively. As \( m \) approaches infinite, \( C \) converges to 1. Thus, \( C \) increases with \( m \): when \( m \) grows from 1 to infinite, \( C \) increases from 0.82008 to 1. Therefore, for the full range of \( m \), the the average clustering coefficient of Koch networks is very high.
4.3. Diameter

Most real networks are small-world, i.e. their average distance grows logarithmically with network order or slower. Here the average distance means the minimum number of edges connecting a pair of nodes, averaged over all node pairs. For a general network, it is not easy to derive a closed formula for its average distance. However, the whole family of Koch networks has a self-similar structure, allowing for analytically calculating the average distance, which approximately increases as a logarithmic function of the network order. We leave the detailed exact derivation about the average distance to the next subsection.

Here we provide the exact result of the diameter of \( K_{m,t} \) denoted by \( \text{Diam}(K_{m,t}) \) for all parameters \( m \), which is defined as the maximum of the shortest distances between all pairs of nodes. Small diameter is consistent with the concept of small-world. The obtained diameter also scales logarithmically with the network order. The computation details are presented as follows.

Clearly, at step \( t = 0 \), \( \text{Diam}(K_{m,0}) \) is equal to 1. At each step \( t \geq 1 \), we call newly created nodes at this step as active nodes. Since all active nodes are attached to those nodes existing in \( K_{m,t-1} \), so one can easily see that the maximum distance between any active node and those nodes in \( K_{m,t-1} \) is not more than \( \text{Diam}(K_{m,t-1}) + 1 \) and that the maximum distance between any pair of active nodes is at most \( \text{Diam}(K_{m,t-1}) + 2 \). Thus, at any step, the diameter of the network increases by 2 at most. Then we get \( 2(t + 1) \) as the diameter of \( \text{Diam}(K_{m,t}) \).

Equation (1) indicates that the logarithm of the order of \( \text{Diam}(K_{m,t}) \) is proportional to \( t \) in the large limit \( t \). Thus the diameter \( \text{Diam}(K_{m,t}) \) grows logarithmically with the network order, showing that the Koch networks are small-world.

4.4. Average path length

Using a method similar to but different from those in the literature [27, 28], we now study analytically the average path length \( d_i \) of the Koch networks \( K_{m,t} \). It follows that

\[
    d_i = \frac{D_{\text{tot}}(t)}{N_t(N_t - 1)/2},
\]

where \( D_{\text{tot}}(t) \) is the total distance between all couples of nodes, i.e.

\[
    D_{\text{tot}}(t) = \sum_{i \in K_{m,t}, j \in K_{m,t}, i \neq j} d_{ij}(t),
\]

where \( d_{ij}(t) \) is the shortest distance between nodes \( i \) and \( j \) in the networks \( K_{m,t} \).

Note that Koch networks have a self-similar structure, which allows us to address \( D_{\text{tot}}(t) \) analytically. This self-similar structure is obvious from an equivalent network construction method: to obtain \( K_{m,t} \), one can make \( 3m + 1 \) copies of \( K_{m,t-1} \) and join them at the hubs (namely nodes with largest degree). As shown in figure 7, the network \( K_{m,t+1} \) may be obtained by the juxtaposition of \( 3m + 1 \) copies of \( K_{m,t} \), which are labeled as \( K_{m,t}^1, K_{m,t}^2, \ldots, K_{m,t}^{3m+1} \), respectively.

We continue by exhibiting the procedure of the determination of the total distance and present the recurrence formula, which allows us to obtain \( D_{\text{tot}}(t+1) \) of the \( t + 1 \) generation from \( D_{\text{tot}}(t) \) of the \( t \) generation. From the obvious self-similar structure of Koch networks, it is easy to see that the total distance \( D_{\text{tot}}(t + 1) \) satisfies the recursion relation

\[
    D_{\text{tot}}(t + 1) = (3m + 1) D_{\text{tot}}(t) + \Omega_t,
\]
Figure 7. Second construction method of Koch networks that highlights self-similarity. The graph after \( t + 1 \) construction steps, \( K_{m,t+1} \), consists of \( 3m + 1 \) copies of \( K_{m,t} \) denoted as \( K_{m,t}^{\theta} \) \((\theta = 1, 2, 3, \ldots, 3m, 3m + 1)\), which are connected to one another as above.

where \( \Omega_t \) is the sum over all shortest paths whose endpoints are not in the same \( K_{m,t}^{\theta} \) branch. The solution of equation (13) is

\[
D_{\text{tot}}(t) = (3m + 1)^{t-1}D_{\text{tot}}(1) + \sum_{\tau=1}^{t-1}(3m + 1)^{t-\tau-1}\Omega_\tau.
\]

(14)

All the paths contributing to \( \Omega_t \) must go through at least one of the three edge nodes (i.e. the gray nodes X, Y and Z in figure 7) at which the different \( K_{m,t}^{\theta} \) branches are connected. The analytical expression for \( \Omega_t \), called the length of crossing paths, is found below.

Let \( \Omega_{\alpha,\beta}^t \) be the sum of the lengths of all shortest paths with endpoints in \( K_{m,t}^{\alpha} \) and \( K_{m,t}^{\beta} \). Based on whether or not two branches are adjacent, we sort the crossing path length \( \Omega_{\alpha,\beta}^t \) into two classes: if \( K_{m,t}^{\alpha} \) and \( K_{m,t}^{\beta} \) meet at an edge node, \( \Omega_{\alpha,\beta}^t \) rules out the paths where either endpoint is that shared edge node. For example, each path contributed to \( \Omega_{1,2}^t \) should not end at node X. If \( K_{m,t}^{\alpha} \) and \( K_{m,t}^{\beta} \) do not meet, \( \Omega_{\alpha,\beta}^t \) excludes the paths where either endpoint is any edge node. For instance, each path contributed to \( \Omega_{2,m+2}^t \) should not end at node X or Y. We can easily compute that the numbers of the two types of crossing paths are \( 3m^2 + 3m \) and \( 3m^2 \), respectively. On the other hand, any two crossing paths belonging to the same class have identical length. Thus, the total sum \( \Omega_t \) is given by

\[
\Omega_t = \frac{3m^2 + 3m}{2} \Omega_{t-1}^1 + 3m^2 \Omega_{t-1}^{2,m+2}.
\]

(15)

In order to determine \( \Omega_{t}^{1,2} \) and \( \Omega_{t}^{2,m+2} \), we define

\[
s_t = \sum_{i \in K_{m,t}, i \neq X} d_{i,X}(t).
\]

(16)

Considering the self-similar network structure, we can easily know that at time \( t + 1 \), the quantity \( s_{t+1} \) evolves recursively as

\[
s_{t+1} = (m + 1)s_t + 2m [s_t + (N_t - 1)]
\]

\[
= (3m + 1)s_t + 4m (3m + 1)t.
\]

(17)
Using \( s_0 = 2 \), we have
\[
s_t = (4mt + 6m + 2) (3m + 1)^{t-1}. \tag{18}
\]

Having obtained \( s_t \), the next step is to compute the quantities \( \Omega^{1,2}_t \) and \( \Omega^{2,mt+2}_t \) given by
\[
\Omega^{1,2}_t = \sum_{i \in K^{1,2}_{m,t}, j \notin X} d_{ij} (t + 1)
= \sum_{i \in K^{1,2}_{m,t}, j \notin X} [d_{iX} (t + 1) + d_{jX} (t + 1)]
= (N_t - 1) \sum_{i \in K^{1,2}_{m,t}, j \notin X} d_{iX} (t + 1) + (N_t - 1) \sum_{j \in K^{1,2}_{m,t}, i \notin X} d_{jX} (t + 1)
= 2(N_t - 1) \sum_{i \in K^{1,2}_{m,t}, j \notin X} d_{iX} (t + 1)
= 2(N_t - 1) s_t, \tag{19}
\]
and
\[
\Omega^{2,mt+2}_t = \sum_{i \in K^{2,mt+2}_{m,t}, i \notin X} d_{ij} (t + 1)
= \sum_{i \in K^{2,mt+2}_{m,t}, i \notin X} [d_{iX} (t + 1) + d_{XY} (t + 1) + d_{jY} (t + 1)]
= 2(N_t - 1) s_t + (N_t - 1)^2, \tag{20}
\]
where \( d_{XY} (t + 1) = 1 \) has been used. Substituting equations (19) and (20) into equation (15), we obtain
\[
\Omega_t = (9m^2 + 3m)(N_t - 1) s_t + 3m^2 (N_t - 1)^2
= 12m(2mt + 4m + 1)(3m + 1)^{t-1}. \tag{21}
\]
Inserting equations (21) for \( \Omega_t \) into equation (14), and using \( D_{tot}(1) = 48m^2 + 21m + 3 \), we can exactly obtain the expression for \( D_{tot}(t) \) as
\[
D_{tot}(t) = \frac{(3m + 1)^{t-1}}{3} [3m + 5 + (24mt + 24m + 4)(3m + 1)^t]. \tag{22}
\]
By inserting equation (22) into equation (11), one can obtain the analytical expression for \( d_t \):
\[
d_t = \frac{3m + 5 + (24mt + 24m + 4)(3m + 1)^t}{3(3m + 1) [2(3m + 1)^t + 1]}, \tag{23}
\]
which approximates \( \frac{4mt}{(3m + 1)^t} \) in the infinite \( t \), implying that the APL shows a logarithmic scaling with network order. This again shows that the Koch networks exhibit a small-world behavior. We have checked our analytic result for \( d_t \) given in equation (23) against numerical calculations for different \( m \) and various \( t \). In all the cases we obtain complete agreement between our theoretical formula and the results of numerical investigation, see figure 8.
4.5. Degree correlations

Degree correlation is a particularly interesting subject in the field of network science [7, 8, 29–32] because it can give rise to some interesting network structure effects. An interesting quantity related to degree correlations is the average degree of the nearest neighbors for nodes with degree \( k \), denoted as \( k_{nn}(k) \), which is a function of the node degree \( k \) [30, 31]. When \( k_{nn}(k) \) increases with \( k \), it means that nodes have a tendency to connect to the nodes with a similar or larger degree. In this case the network is defined as assortative [7, 8]. In contrast, if \( k_{nn}(k) \) is decreasing with \( k \), which implies that the nodes of large degree are likely to have near neighbors with small degree, then the network is said to be disassortative. If correlations are absent, \( k_{nn}(k) = \text{const.} \).

We can exactly calculate \( k_{nn}(k) \) for Koch networks using equations (4) and (5) to work out how many links are made at a particular step to nodes with a particular degree. By construction, we have the following expression:

\[
\begin{align*}
   k_{nn}(k) &= \frac{1}{L_v(t_i)k(t_i, t)} \left( \sum_{t_i'}^{t_i} m L_v(t_i')k(t_i', t_i - 1)k(t_i', t) \right. \\
   &\quad \left. + \sum_{t_i'}^{t_i} m L_v(t_i')k(t_i, t_i' - 1)k(t_i', t) \right) + 1
\end{align*}
\]

for \( k = 2(m + 1)^{-\frac{1}{m}} \). Here the first sum on the right-hand side accounts for the links made to nodes with a larger degree (i.e. \( t_i' < t_i \)) when the node was generated at \( t_i \). The second sum describes the links made to the current smallest degree nodes at each step \( t_i' > t_i \). The last term 1 accounts for the link connected to the simultaneously emerging node. In order to compute equation (24), we distinguish two cases according to the parameter \( m \): \( m = 1 \) and \( m \geq 2 \).

When \( m = 1 \), we have

\[
k_{nn}(k) = t + 2.
\]
Thus, in the case of $m = 1$, the networks show the absence of correlations in the full range of $t$. From equations (25) and (1) we can easily see that for large $t$, $k_{nn}(k)$ is approximately a logarithmic function of the network order $N_t$, namely $k_{nn}(k) \sim \ln N_t$, exhibiting a similar behavior as that of the BA model [31] and the two-dimensional random Apollonian network [32].

When $m \geq 2$, equation (24) is simplified to

$$k_{nn}(k) = \frac{3m+1}{m-1} \left[ \frac{(m+1)^2}{3m+1} \right]^t (\frac{k}{2})^{-\frac{3(m+1)^2}{3m+1}} - \frac{m+3}{m-1} + \frac{2m}{m+1} \ln \left( \frac{k}{2} \right).$$

(26)

Thus after the initial step $k_{nn}(k)$ grows linearly with time. Writing equation (26) in terms of $k$, it is straightforward to obtain

$$k_{nn}(k) = \frac{3m+1}{m-1} \left[ \frac{(m+1)^2}{3m+1} \right]^t \frac{m+3}{m-1} + \frac{2m}{m+1} \ln \ln \left( \frac{k}{2} \right).$$

(27)

Therefore, $k_{nn}(k)$ is approximately a power-law function of $k$ with negative exponent, which shows that the networks are disassortative. Note that $k_{nn}(k)$ of the Internet exhibits a similar power-law dependence on the degree $k_{nn}(k) \sim k^{-\omega}$, with $\omega = 0.5$ [30].

4.6. Spanning trees, spanning forests and connected spanning subgraphs

Spanning trees, spanning forests and connected spanning subgraphs are important quantities of networks, and the enumeration of these interesting quantities in networks is a fundamental issue [33–37]. However, explicitly determining the numbers of these quantities in networks is a theoretical challenge [38]. Fortunately, the peculiar construction of Koch networks makes it possible to derive exactly the three variables.

4.6.1. Spanning trees. By definition, a spanning tree of any connected network is a minimal set of edges that connect every node. The problem of spanning trees is closely related to various aspects of networks, such as reliability [39, 40], optimal synchronization [41] and random walks [42]. Thus, it is of great interest to determine the exact number of spanning trees [43]. In what follows we will examine the number of spanning trees in Koch networks.

Note that in the Koch networks $K_{m,t}$ there are $L_\triangle(t) = (3m+1)^t$ triangles, but there are no cycles of length more than 3. For each of $L_\triangle(t) = (3m+1)^t$ triangles, to assure that its three nodes are in one tree, only two edges of it must be present. Obviously, there are three possibilities for this. Thus, the total number of spanning trees in $K_{m,t}$, denoted by $N_{ST}(t)$, is

$$N_{ST}(t) = 3^{L_\triangle(t)} = 3^{(3m+1)^t}. \quad (28)$$

We proceed to represent $N_{ST}(t)$ as a function of the network order $N_t$, with the aim to provide the relation governing the two quantities. From equation (1), we have $(3m+1)^t = \frac{N_t-1}{2}$. This expression allows one to write $N_{ST}(t)$ in terms of $N_t$ as

$$N_{ST}(t) = 3^{(N_t-1)/2}. \quad (29)$$

Thus, the number of spanning trees in $K_{m,t}$ increases exponentially with the network order $N_t$, which means that there exists a constant $E_{ST}$, called as the entropy of spanning trees, describing this exponential growth [34]:

$$E_{ST} = \lim_{N_t \to \infty} \frac{\ln N_{ST}(t)}{N_t} = \frac{1}{2} \ln 3. \quad (30)$$

In addition to the above analytical computation, according to the previously known result [44], one can also obtain numerically but exactly the number of spanning trees, $N_{ST}(t)$, by
computing the nonzero eigenvalues of the Laplacian matrix associated with the networks $K_{m,t}$ as
\[
N_{ST}(t) = \frac{1}{N_t} \prod_{i=1}^{N_t-1} \lambda_i(t),
\]
(31)
where $\lambda_i(t) (i = 1, 2, \ldots, N_t - 1)$ are the $N_t - 1$ nonzero eigenvalues of the Laplacian matrix, denoted by $\mathbf{L}_t$, for the networks $K_{m,t}$, which is defined as follows: its non-diagonal element $l_{ij}(t)$ ($i \neq j$) is $-1$ (or $0$) if nodes $i$ and $j$ are (or not) directly linked to each other, while the diagonal entry $l_{ii}(t)$ is exactly the degree of node $i$.

Using equation (31), we have calculated directly the number of spanning trees in the networks $K_{m,t}$, and the results from equation (31) are fully consistent with those obtained from equation (28), showing that our analytical formula is right. It should be stressed that although expression (31) seems compact, it is involved in the computation of the eigenvalues of a matrix of order $N_t \times N_t$, which makes heavy demands on time and computational resources. Thus, it is not acceptable for large networks. In particular, by virtue of the eigenvalue method it is difficult and even impossible to obtain the entropy $E_{ST}$. Our analytical computation can get around the two difficulties, but is only applicable to peculiar networks.

4.6.2. Spanning forests. To define spanning forests, we first recall the definition for a spanning subgraph. A spanning subgraph of a network is a subgraph having the same node as of the network but having partial or all edges of the original graph. A spanning forest of a network is a spanning graph of it that is a disjoint union of trees (here an isolated node is consider as a tree), i.e. a spanning graph without any cycle. The enumeration of spanning forests is very interesting since it corresponds to the partition function of the $q$-state Potts model [45] in the limit of $q \to 0$. For a general network, it is very hard to count the number of its spanning subgraphs. But below we will show that for the Koch networks $K_{m,t}$, the number of spanning subgraphs, $N_{SF}(t)$, can be obtained explicitly.

Analogous to the enumeration of spanning trees, for each triangle in $K_{m,t}$, to guarantee the absence of cycle among its three nodes, at least one edge must be removed. And there are total seven possibilities for deleting the edges of a triangle. Then the number of spanning forests in $K_{m,t}$ is
\[
N_{SF}(t) = 7^{L_{\triangle}(t)} = 7^{(3m+1)t},
\]
(32)
which can be rewritten as a function of the network order $N_t$ as
\[
N_{SF}(t) = 7^{(N_t-1)/2}.
\]
(33)
Therefore, $N_{SF}(t)$ also grows exponentially in $N_t$, which allows for defining the entropy of the spanning forests of Koch networks as the limiting value [46]:
\[
E_{SF} = \lim_{N_t \to \infty} \frac{\ln N_{SF}(t)}{N_t} = \frac{1}{2} \ln 7.
\]
(34)
Thus, we have obtained the rigorous results for the number of spanning forests in Koch networks and its entropy.

4.6.3. Connected spanning subgraphs. As the name suggests, a connected spanning subgraph of a connected network is a spanning subgraph of the network, which remains connected. By applying a method similar to that given above, we can compute the number of connected spanning subgraphs in the Koch networks $K_{m,t}$, which is denoted by $N_{CSS}(t)$. For any triangle in $K_{m,t}$, to ensure the connectedness of its three nodes, at most one edge can be
deleted. In other words, two or three edges of it should be present. On the other hand, for an arbitrary triangle in $K_{m,t}$, the number of its connected spanning subgraphs is 4. Then, the number of connected spanning subgraphs is

$$N_{CSS}(t) = 4^{L_{△}(t)} = 4^{(3m+1)}.$$  \hfill (35)

which can be further recast in terms of the network order $N_t$ as

$$N_{CSS}(t) = 4^{(N_t−1)/2}.$$  \hfill (36)

Clearly $N_{CSS}(t)$ increases exponentially in $N_t$. Thus, the entropy of the connected spanning subgraphs of Koch networks is

$$E_{CSS} = \lim_{N_t \to \infty} \frac{\ln N_{CSS}(t)}{N_t} = \ln 2.$$  \hfill (37)

5. Conclusions

Networks having a structure complicated enough to display nontrivial properties for real-life systems and statistical models but simple enough to reveal analytical insights are few and far between. In this paper, we have presented a mapping that converts the Koch curves into complex networks, which have many important general properties observed in real networks. Our networks are characterized by closed-form, exact formulas for various properties, especially for numbers for spanning trees, spanning forests and connected spanning subgraphs. The rigorous solutions show that the resulting graphs have a heavy-tailed degree distribution with the general exponent $\gamma \in [2, 3]$, logarithmical diameter and average path length with network order, large clustering coefficient, and degree correlations. Our analytical technique could guide and shed light on related studies for deterministic network models by providing a paradigm for computing the structural features.

In addition to the analytical method and the nontrivial topological characterizations of our proposed model, another contribution of our work is mapping the Koch fractals to a family of graphs. With the mapping method, a natural bridge between the theory of network science and the classic fractals has been built. This network representation technique could find application to some real systems making it possible to explore the complexity of real-life networked systems in biological and information fields within the framework of complex network theory. Recently, a similar recipe has been adopted for investigating the navigational complexity of cities [47], which has also proven useful to the study of polymer physics [48].

Finally, it should be stressed that for the special case of $m = 1$, the network is completely uncorrelated; it is thus of potential interest as a null substrate network to corroborate the dynamical behavior of complex systems since the analytic solution of dynamical processes is usually available only for uncorrelated networks [49–52], the generation algorithm of which is often more difficult than one would expect a priori [53].

Before closing the paper, it should be mentioned that in a previous paper, Barabási et al proposed a deterministic model, hereafter called the BRV model [54], which is the progenitor of deterministic network models. Although the BRV model is also scale-free and small-world [55], its degree exponent $\gamma$ is a constant $1 + \ln 3/\ln 2$, and its clustering coefficient is zero because there is no triangle in the BRV model. In this context, our model addressed here can mimic real systems better than the BRV model.

Acknowledgments

We would like to thank Yichao Zhang for his assistance. This research was supported by the National Natural Science Foundation of China under grants nos 60704044, 60873040 and
60873070, the National Basic Research Program of China under grant no 2007CB310806, and Shanghai Leading Academic Discipline Project no B114. SG also acknowledges the support by Fudan’s Undergraduate Research Opportunities Program, and HZ acknowledges the support by Shanghai Key Laboratory of Intelligent Information Processing, China, grant no IPL-09-017.

References

[23] Schneider J E 1965 Math. Mag. 38 144
[31] Zhang Z Z and Zhou S G 2007 Physica A 380 621
[34] Lyons R, Peres Y and Schramm O 2001 Ann. Probab. 29 1
[39] Boesch F T 1986 J. Graph Theory 10 339