A local-world network model based on inter-node correlation degree

Qi Xuan, Yanjun Li*, Tie-Jun Wu

National Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China

Received 8 September 2006; received in revised form 9 November 2006
Available online 18 December 2006

Abstract

A new local-world model is proposed in this paper to improve and extend the descriptive ability of the first geometrical local-world network model, i.e., the Gardeñes–Moreno (GM) model. A concept of correlation degree between nodes is introduced into the proposed model to build a local world of each node in a network. In consideration of the facts that each node has only a limited ability of information recognition and processing, and the improvement speed of this ability is normally far slower than the growth speed of a network, the local-world size is set limited and unchanged while the network grows. A series of theoretical analysis and numerical simulation are conducted in this paper, the results show that the correlation degrees follow a power-law distribution, and the proposed model can describe the small-world, scale-free, self-similarity and clustering properties for more comprehensive kinds of complex networks than the GM model and other existing local-world network models. The modelling and analysis of a supply chain system is discussed in this paper as a real-world example of our model.

Keywords: Complex network; Local-world model; Correlation degree; Self-similarity; Supply chain system

1. Introduction

One of the most important contributions of complex networks research [1] is that it reveals the fact that many biological [2–4], social [5,6], and technological [7–9] systems share some common network properties, such as scale-free [3–5,7,10], small-world [1,11], hierarchical and modular structure [2,3,6,7] and so on. A number of models have been proposed [2,12–14] to describe and analyze these properties, but most of them ignore the intrinsic correlations among network nodes which in fact may play a very important role in the evolution of a network structure. For example, in a world-wide airport network [7], the nodes (i.e., airports) occupy precise positions in two-dimensional Euclidean space, and the distance between two nodes greatly affects the efficiency and the strength of transportation along the connections (i.e., airlines) between them. Even in some non-spatial networks, e.g., in a scientific collaboration network (SCN) [15], the correlation
(i.e., research area overlaps) strength between two nodes (i.e., research institutes or scientists) may also have a significant effect on the behavior of the whole network.

In a real-world network, the ability or capability of each node is usually limited due to various kinds of physical constraints [16,17], which may have an inneglectable impact on the characteristics of the network. For instance, in a world trade web (WTW), an enterprise usually pays its attention on a subset of correlated companies in the web for its own business. When a node of this kind is added in a network, it will not be able/willing to obtain the global information about the network. In such a situation, a network growth rule only works in a local world around a newly added node, but not the whole network. Several local-world network models [17–20] have been introduced to help us understand this real-life phenomenon better. In these models, however, the nodes potential to be linked to a new node are selected randomly from all existing nodes in the network with equal probability, and the intrinsic correlations between the new node and the set of all existing nodes as a whole are not considered. This is obviously unreasonable in the real world. For example, an enterprise in a WTW always wants to contact with those companies in its relevant business fields; and a research institute in a SCN is likely to cooperate with those units in its associated research areas.

As an effort to fill the gap between the existing models and the systems in the real world, Gardeñes and Moreno [21] (GM) recently propose a new model based on the assumption that every node in a network has only the knowledge about the nodes in a predefined local area around it. It can be considered as the first work to establish a geometrical local-world network model. In their paper, however, the number of nodes in the local world is supposed to increase linearly as the network grows, and the estimates of some important network parameters, e.g., local clustering coefficients and degree–degree correlations, by using this model, significantly differ from the measurement in real-world networks [21].

In this paper, we introduce the concept of correlation degree between nodes into the GM model in order to measure the intrinsic correlations among nodes in a network. Intuitively, based on a generalized “position” of each node (it should be noted that in a non-spacial network the definition of a node position may also make sense) in a network, the closer the positions of two nodes, the larger the correlation degree is between them. Then, different from the GM model, we define the local world of a node by a fixed number of nodes with bigger correlation degrees to it, in consideration of the fact that each node has only a limited recognition and processing ability.

On the basis of those definitions, theoretical analysis as well as numerical simulations are conducted in this paper to reveal more statistical characteristics of complex networks than those in the GM model and other models. Investigation results show that, comparing with the existing local-world network models, more real-world network properties, such as power-law clustering function, power-law correlation degree distribution as a generalized version of power-law length distribution [4], and self-similarity, can be expressed more precisely by our model.

The rest of the paper is organized as follows: In Section 2, a local-world network model based on inter-node correlation degree is introduced. Theoretical analysis and numerical simulations are provided in Section 3. To exhibit the background of our model, a real-world example is introduced in Section 4. Our work is summarized in Section 5.

2. The local-world network model

Different from other local-world network models [17–20], the process of network growth in our model is composed of two stages. In the first stage, a new node $j$ is created and assigned with a parameter $a_j \in \mathbb{R}^n$ randomly to define the $n$-dimensional (in this paper we set $n = 1$ for descriptive simplicity) position of node $j$ in order to measure how close it is to the rest of the network. Then the correlation degree $h_{ij}$ between node $j$ and each existing node $i$ at position $a_i$ in the network is computed by

$$h_{ij} = |a_i - a_j|^{-1}, \quad i = 1, 2, \ldots$$ (1)

In the second stage of the network growth process, the preferential attachment (PA) rule [10] is applied to all the $M$ nodes in the local world of the new node in terms of their bigger correlation degrees with it, generating links between the new node and a subset of the local world in a probabilistic manner.
The detailed steps of the network growth process are listed as follows.

**Step 1:** Start at time $t = 0$ with a small number ($m_0$) of fully connected nodes. Assign to each of the nodes a random position $a_i$, $i = 1, 2, \ldots, m_0$, taken from a uniform distribution between $(0,1)$.

**Step 2:** At time $t + 1$, a new node $j$ is added into the network, given a random position $a_j$ between $(0,1)$, and connected to $m$ ($m \leq m_0$) existing nodes according to the following two sub-steps:

(a) Search through all $t + m_0$ nodes in the network, calculate the correlation degree $h_{ij}$ between node $j$ and existing node $i$, $i = 1, 2, \ldots, m_0 + t$ in the network according to Eq. (1), and choose $M$ ($M \geq m$) nodes with bigger correlation degrees to node $j$ to group them into a set $A_j$ as the local world of node $j$. Each member in the local world $A_j$ is regarded as a candidate for possible connection between it and node $j$. It should be noted that if $M > t + m_0$, we choose all the existing nodes in the network to form the $A_j$.

(b) Apply the PA rule to the local world $A_j$, i.e., for each node $i \in A_j$ calculate the connection probability $\Pi_{local}(j \rightarrow i)$ between it and the new node $j$ by

$$\Pi_{local}(j \rightarrow i) = \frac{k_i}{\sum_{l \in A_j} k_l},$$

where $k_i$ is the connection degree of node $l$, defined by the number of nodes directly linked to the node $l$ [10,18]. Select $m$ nodes from $A_j$ with the connection probabilities defined in Eq. (2) to connect them with the new node $j$.

**Step 3:** Repeat Step 2 for $T$ times. The final size of the network is $N = m_0 + T$.

The parameters $m_0$, $m$, $M$, and $T$ in the aforementioned network growth process are all predefined before the modelling process starts. The exact setting of those parameters will be discussed in the following section.

3. Model analysis

3.1. Correlation degree distribution

In this section, we investigate first the correlation degree distribution for the networks generated by our model. A correlation degree distribution $P(h < H)$ is defined as the average probability of finding a connection in the network with the correlation degree $h$ between its two end nodes less than $H$ for different values of $H$.

According to Section 2, at time $t + 1$ the average local-world scale in a network is $M/(t + m_0)$ where the scale of a local world is defined as the multiplicative inverse of the smallest correlation degree in this local world. Then at time $t + 1$ the average probability of the new node linked to the members in its local world with the correlation degree less than $H$ can be calculated by

$$P_t(h < H) = P_t\left(|x - y| > \frac{1}{H}, x \in \left(0, \min\left(\frac{M}{t + m_0}, 1\right)\right), y \in \left(0, \min\left(\frac{M}{t + m_0}, 1\right)\right)\right)$$

$$= \begin{cases} 
(1 - \frac{1}{H})^2, & t < M - m_0; \\
(1 - \frac{t + m_0}{MH})^2, & M - m_0 \leq t < MH - m_0; \\
0, & t \geq MH - m_0. 
\end{cases}$$

Thus at the end of the network growth process, the correlation degree distribution of the whole network can be calculated by

$$P(h < H) = \int_0^T \frac{1}{T} P_t(h < H) \, dt,$$
size $N$. Plugging Eq. (3) into Eq. (4), we can get that when $MH - m_0 \leq T$,
\[
P(h < H) = \int_0^{M - m_0} \frac{1}{T} \left(1 - \frac{1}{H}\right)^2 dt + \int_{M - m_0}^{MH - m_0} \frac{1}{T} \left(1 - \frac{t + m_0}{MH}\right)^2 dt
\]
\[
= \frac{M - m_0}{T} \left(1 - \frac{1}{H}\right)^2 + \frac{MH}{3T} \left(1 - \frac{1}{H}\right)^3
\]
\[
\sim \frac{M}{3T} H.
\] (5)

Eq. (5) implies that the correlation degrees of a network generated by our model obey a power-law distribution when $H \leq (T + m_0)/M = N/M$. The simulation result shown in Fig. 1 validates this assertion.

3.2. Connection degree distribution

In the network growth process in Section 2, the local world $A_j$ of a new node $j$ is composed of $M$ ($M \geq m$) candidates from which $m$ nodes will be eventually selected to connect to the new node. When $M = m$, it implies that the new node will be linked to all nodes in its local world. Therefore, the PA rule in Eq. (2) will lose its effect on the model. In other words, the connection degree $k_i$ of node $i$, will obey
\[
\frac{\partial k_i}{\partial t} = \Pi'(i \in A_j) = \frac{m}{t + m_0},
\] (6)
where $\Pi'(i \in A_j)$ is the probability of node $i$ belonging to $A_j$, satisfying $\Pi'(i \in A_j) = M/(t + m_0)$. Eq. (6) results in an exponentially decayed connection degree distribution, i.e., $P(k) \sim \exp(-k/m)$.

If $M > m$, on the other hand, the PA rule is at least applicable locally. Then in average we can expect that the network will have similar node connection degree distribution as that created by the BA model proposed by Barabási and Albert [10,21]. In view of the fact that in our model the node positions obey a uniform distribution in (0,1), and are independent of their connection degrees (as is shown in Fig. 2), when the network

Fig. 1. Plots of the correlation degree distribution, defined as the average probability of finding a connection between two nodes with a correlation degree less than $H$, for networks with different values of $M = 2, 5, 10, 20$. Each network has $N = 5 \times 10^3$ nodes with $m = m_0 = 2$. 
growth time \( t \) is sufficiently large, we have

\[
\sum_{i \in \mathcal{A}_j} k_i = \frac{2mt + 2e_0}{t + m_0} M \approx 2mM,
\]

where \( e_0 \) is the number of connections in the initial network. Combining Eq. (7) with Eq. (2) leads to

\[
\frac{\partial k_i}{\partial t} = \frac{\partial}{\partial t} \left( i \in \mathcal{A}_j \right) \Pi_{\text{local}}(j \rightarrow i) \approx \frac{mM}{(t + m_0)} \frac{k_j}{2m} \approx \frac{k_i}{2t},
\]

which means that the connection degree distribution of the network is the same as that in the BA model [10,21], i.e., \( P(k) \sim 2m^2/k^3 \). The connection degree distributions for different values of \( M \) is demonstrated in Fig. 3.

### 3.3. Clustering coefficient

In order to estimate the local clustering coefficients [11] of a network created by our model, we define the total number of connections within the local world \( \mathcal{A}_j \) of node \( j \) as the local-world degree of node \( j \) when it is created, denoted by \( d(\mathcal{A}_j) \). This value is normally different from the total number of connections linked to all the members of \( \mathcal{A}_j \), i.e., \( \sum_{i \in \mathcal{A}_j} k_i \), because those connections with only one end-node in the set \( \mathcal{A}_j \) are not counted here.

Considering the creation order of the \( M \) nodes in \( \mathcal{A}_j \), we can figure that the \( v \)th generated node within the local world \( \mathcal{A}_j \) has chances to connect to \( v - 1 \) nodes that already exist in the local world, or to connect to \( M - (v - 1) \) nodes outside the local world. So in average it will connect to \( m(v - 1)/M \) nodes in \( \mathcal{A}_j \). This estimate is based on an assumption that the \( v - 1 \) nodes generated before the \( v \)th node in \( \mathcal{A}_j \) are within the local world of the \( v \)th generated node in \( \mathcal{A}_j \). It is statistically reasonable because of the following two facts: (1) the local-world scale defined in Section 3.1 decreases as the network grows since all the node positions are generated in a limited region \((0,1)\); (2) according to (1), the local-world scales of the nodes in \( \mathcal{A}_j \) are far larger than the scale of \( \mathcal{A}_j \) itself because those nodes are generated before node \( j \). Then the mean value of the...
The local-world degree of node $j$ can be calculated by

$$d(A_j) = \frac{1}{2} \frac{m(M - 1)}{M}. \quad (9)$$

We plot in Fig. 4 the distribution of the local-world degrees of a new node at different positions with different values of $M$. It can be derived that at most positions, those local-world degrees are similar to each other and close to the mean value $d(A_j)$. The sub-figure in the up left in Fig. 4 also validates the result of Eq. (9).

Curiously we compute the ratio between $d(A_j)$ and that of a fully connected network with the same size, i.e., $M$ nodes, as follows:

$$\rho \approx \frac{0.5m(M - 1)}{0.5M(M - 1)} = \frac{m}{M}. \quad (10)$$

Eq. (10) implies that, when $M$ is close to $m$, the subnet structure of a local world is similar to a fully connected network. In such a situation, we can expect that the connections in the local world are distributed uniformly among $M$ nodes, and

$$\frac{\partial e_i}{\partial k_i} \approx \frac{m}{M}(m - 1), \quad (11)$$

where $e_i$, called the density of node $i$, is the total number of connections among the nodes directly connected to node $i$. From Eq. (11) we can easily obtain that $e_i \sim m(m - 1)k_i/M$. Then the local clustering coefficient $C(k_i)$ of node $i$ can be calculated by

$$C(k_i) = \frac{2e_i}{k_i(k_i - 1)} \sim \frac{m(m - 1)}{M} k_i^{-1}. \quad (12)$$
That is to say, when \( m = M \) is close to 1, the clustering function \( C(k) \) as the average of \( C(k_i) \) over all the nodes \( i \) with connection degree \( k_i = k \) will follows the distribution \( C(k) \sim k^{-1} \), which is similar to that of most real-world networks. Fig. 5 shows that such a structure is very robust: the relationship \( C(k) \sim k^{-1} \) still holds even when the ratio \( m/M \) changes from 0.1 to 1. Furthermore, from Eq. (12), we find that the clustering coefficient
of the whole network, defined by the average value of the local clustering coefficients of all nodes in the network, is positively correlated to $m(m - 1)/M$, as is shown in Fig. 6.

On the other hand, when the number of connections allowed to be added to each new node is limited to a value far less than the number of nodes in its local world, i.e., $m \ll M$, similar to that in the BA model, the nodes with larger connection degrees will have more chances to increase their densities, i.e.,

$$\frac{\partial e_i}{\partial k_i} \sim k_i,$$

and in this case the clustering function $C(k)$ will be independent of the connection degree $k$ [21].

3.4. Average shortest path length

The shortest path length between two given nodes in a network generated by our model is defined the same as those in Refs. [1,11], i.e., the smallest number of connections included in any path between the two nodes, though in our model each node is given a position. To compare our model with the BA model and the GM model, we plot in Fig. 7 the average shortest path length versus the network size $N$ for four networks, two of them are created by our model with $M = 5$ and 10, respectively; the third one by the BA model, and the fourth by the GM model with parameter $\mu = 0.25$. It shows that the average shortest path lengths of all the networks obey the same law, i.e., $L(N) \sim \eta \ln(N)$.

With careful observation, however, we can find that the BA model and the GM model yield similar network behaviors, i.e., they have similar $\eta$ in $L(N)$, but the average shortest path lengths of the networks generated by our model increase more quickly, i.e., their $\eta$ are larger than those of the former two. A reasonable explanation for this phenomenon is that the local worlds in networks generated by the BA model and the GM model have almost fixed scales, while in our model the local-world scale decreases linearly as the network grows, which discourages long-range connections. On the other hand, at the early stage of the network growth, the local-world scale in our model may be larger than that in the GM model. As a result, the average shortest path
length of a network with small size $N$ created by our model may be shorter than that in the GM model. This phenomenon can also be seen in Fig. 7.

A small-world network can be characterized by a large network clustering coefficient and a logarithmically increasing average shortest path length of the network. The second property can be expressed by the BA model, the GM model and our model. Through the analysis in Section 3.3, we find that the average clustering coefficient of a network generated by our model is negatively correlated to $M$ which determines the local-world scale of a node, as is shown in Fig. 6. In consideration of the fact that, for a sufficiently large growth time, the local-world scales of both the BA model and the GM model are greater than that of our model, we can expect that the average clustering coefficient of a network generated by our model is greater than those by the BA model or the GM model. So compared with the BA model and the GM model, the small-world property can be described better by our model.

3.5. Self-similarity

Self-similarity is an interesting and popular network property in the real world, revealing a phenomenon that each part of a network topologically and recursively duplicates the structure of its parent network [6]. This property can be easily simulated by geometry-based network models such as the GM model and our model.

Given a network, a subnet can be extracted from it by the following two steps:

1. Select a node $i$ from the network as the seed of the subnet.
2. Choose each node $j$, satisfying in our model $h_{ij} > \theta$ where $\theta \geq \max(1/(1 - a_i), 1/a_i)$; or in the GM model $|a_i - a_j| < \xi$ where $\xi \leq \min(a_i, 1 - a_i)$, along with all the existing connections among those nodes as the subnet.

It should be noted that, according to the subnet extraction method mentioned above, whether a node is in a subnet relies only on its correlation degree with the seed of the subnet. So a node in the subnet may be an isolated point, in other words, it may have no any link to other nodes in this subnet.
Now we investigate the self-similarity from a network growth point of view. In the GM model, the self-similarity will be discovered only if the fixed local-world scale, measured by the parameter $\mu$ in that model, is very small. The reason is that, as a network grows, a new node in the extracted subnet is most likely to be connected to the subnet only when $\mu \ll \xi$, in other words, the subnet scale is larger than the local-world scale. In this case, we can say that the subnet is self-contained and the connecting rules in the subnet is similar to that in its parent network. As a result, it will have a similar structure as that of the whole network.

By reviewing the network growth process in Section 2, however, one can find that the self-similarity is an intrinsic property of the networks generated by our model; namely, at any network scale this property can be expressed well. In our model, the condition for a new node belonging to a subnet to be extracted is that the correlation degree between the new node and the seed of the subnet is greater than $\theta$. As the network grows, the probability that a new node belonging to the subnet is actually connected to any other nodes out of the subnet will approach 0. This is because there is a tendency that each node must be connected to the nodes within its local world, and the local-world scale is getting smaller and smaller along with the network growth process going on, whereas the subnet scale is fixed at $2^\xi$. As a result, the local world of a new node in the future will eventually “shrink into” the subnet, which leads to the fact that the growth of the subnet will be similar to that of the whole network in the future.

As an example, we extract four subnets with the same seed position $a_i = 0.1$ and $\theta = 10$ from different networks generated by our model with $M = 2, 5, 10, 20$, respectively. We plot the connection degree distributions and the clustering functions of the subnets in the bottom left in Figs. 3 and 5. In these figures we can see that the subnets and their parent networks have very similar connection degree distributions and clustering functions, which implies that the networks generated by our model are self-similar.

4. An example

In this section, a real-world example is demonstrated as a background of our model, though due to limited space the discussion will be focused on the key points of the application.

A supply chain system is a commonly used model in the field of economical research, describing the supply-demand relations among enterprises linked via commodity flows. When a large number of enterprises are involved in a supply chain system, its structure will be characterized by a network. A large amount of research work have be done on the properties and behaviors of various kinds of supply chain systems [22–26], among which the complex networks theory plays a distinctive role in the analysis of the structural evolution of supply networks due to enterprises entering into or quitting from a market [22].

Based on our model, a supply chain system can be expressed by a local-world network, in which each node represents an enterprise, and a connection between two nodes stands for a commodity flow channel between them. The position of a node in this network indicates the status and the influence of the enterprise in the market, described for instance by its product varieties, supply demand, geographic location, as well as market share, commercial credit, financial condition, and R&D ability, etc. [27], based on which the correlation degree between two nodes can be defined to estimate the probability of these enterprises to establish a cooperative relation.

When a new company is set up in the market, it has to establish the supply-demand relations with other existing enterprises for its raw material supply and products sale. Due to various kinds of constraints on its financial, material and manpower resources, the number of potential partners it is searching for is limited. A limited number of enterprises in the supply network may be put into consideration in terms of their bigger correlation degrees with the new company, which in our model constitute the local world of the company. Several substantive cooperation channels may be finally established, implying that this company has merged into the supply chain system.

The network growth process will be going on from time to time, resulting in network structure changes. The analysis on those changes may reveal the development trend of the supply chain system, usually very important for investors, enterprise managers and government officers concerning this system.

By our model, several important properties of the supply chain system can be analyzed. For instance, a scale-free correlation degree distribution implies that in the supply network there may exist a few of vital commodity flow channels that have significant influence on the throughput of the whole system, although the
enterprises at the ends of those channels have only lower correlation degrees. A scale-free connection degree distribution means that there may exist some dominant enterprises, i.e., the enterprises with large number of cooperators, over which a large part of commodity flows in the network will traverse. A scale-free clustering function represents a modular structure of the supply network in which frequent cooperation activities among those enterprises with bigger correlation degrees yield a few of alliances. And self-similarity reveals the fact that, at different organization levels, e.g., at regional, national, or international levels, the structural features of the supply network are very similar to those of its local portions. Other properties such as local clustering coefficient and average shortest path length can also be used to measure the flexibility and the responsive speed of the supply chain system to a market [22].

5. Summary

The model proposed in this paper is straightly derived from the Barabási–Albert model by introducing the concept of correlation degree to quantitate the potential relationship between two nodes in a network, and constraining the PA rule in a local set of nodes composed of a limited number of neighbors with bigger correlation degrees to each newly added node. Theoretical analysis and numerical simulation show that the correlation degree distribution as a generalized version of length distribution obeys the power law, and the proposed model can describe the small-world, scale-free, self-similarity and clustering properties for more comprehensive kinds of complex networks than other local-world network models.

The concept of the local world plays an important role in our model, but for descriptive simplicity, the size of local world is set to a fixed number, which is not realistic. A part of our future work is to generalize local world size to a bounded number or a stochastic variable. Furthermore, in the real world, the calculation of correlation degree may be much more complex, which is also an interesting topic and belongs to our future studies.

Acknowledgments

We would like to thank all the members in our research group in the Institute of Intelligent Systems and Decision Making, Zhejiang University at Yuquan Campus, for the valuable discussion about the ideas presented in this paper. This work has been supported by the National 973 Program of China (Grant no. 2002CB312200).

References