PHYSICAL REVIEW E **73**, 036105 (2006)

Growth model for complex networks with hierarchical and modular structures

Qi Xuan, Yanjun Li,* and Tie-Jun Wu
National Laboratory of Industrial Control Technology, Institute of Intelligent Systems & Decision Making, Zhejiang University, Hangzhou 310027, China
(Received 30 August 2005; published 3 March 2006)

A hierarchical and modular network model is suggested by adding a growth rule along with the preferential attachment (PA) rule into Motter's modeling procedure. The proposed model has an increasing number of vertices but a fixed number of modules and hierarchical levels. The vertices form lowest-level modules which in turn constitute higher-level modules hierarchically. The creation of connections between two vertices in a single module or in two different modules of the same level obeys the PA rule. The structural characteristics of this model are investigated analytically and numerically. The results show that the degree distribution, the module size distribution, and the clustering function of the model possess a power-law property which is similar to that in many real-world networks. The model is then used to predict the growth trends of real-world networks with modular and hierarchical structures. By comparing this model with those real-world networks, an interesting conclusion is found: that many real-world networks are in their early stages of development, and when the growth time is large enough, the modules and levels of the networks will be ultimately merged.

DOI: 10.1103/PhysRevE.73.036105 PACS number(s): 89.75.Fb, 89.75.Da, 05.65.+b

I. INTRODUCTION

In the last few years, considerable efforts have been made to understand and model complex networks in the real world [1–4]. A number of complex networks have been investigated and categorized, in which two kinds have gained much attention due to their popularity in the real world [2–7].

One of the real-world networks which have been well studied is the scale-free (SF) network [1,4,7,8]. Those complex networks, such as in biological [8], social [4,9], and technological systems [10], are found to obey the power-law degree distribution i.e., $P(k) \sim k^{-\gamma}$ where the degree is the number of edges connecting to a given vertex. Barabási and Albert (BA) [4] successfully modeled such networks with the following principle: starting with a small number (m_0) of vertices, adding a new vertex at every time step, and connecting it to $m(m \leq m_0)$ different vertices which are selected with a probability linearly proportional to the degree of the target vertex. Such a selection rule is called the preferential attachment (PA) rule. Recently, many network models based on BA have been proposed and the dynamic properties of them have been explored extensively [1,11].

Another kind of real-world networks which have been well investigated is characterized by a high degree of clustering [2,11] and modular structure [5,12–24], both of which can be measured by a local clustering coefficient [2]. The local clustering coefficient for vertex i with degree k_i is defined as $C_i = 2n_i / k_i(k_i - 1)$, where n_i is the number of edges between the k_i neighbors of vertex i. A high degree of clustering means that the clustering coefficient C—i.e., the average of C_i over all the vertices—is significantly higher for these real-world networks than a regular network of similar size. Furthermore, a network can be considered modular and hierarchical when the clustering function $C(k)$, the average of C_i over the vertices with degree k, satisfies the relationship: $C(k) \sim k^{-\beta}$ and C remains finite for large system size N [8,13,14,18,22].

The BA model obeys the power-law degree distribution with $\gamma = 3$, but $C(k)$ is independent of k and C decreases with N, because the BA model does not contain modules [18]. Recently, Watts *et al.* [23] and Motter *et al.* [24] introduced social network models containing modular and hierarchical structure, where vertices form modules which are in turn grouped to become bigger modules hierarchically. Their models, however, are mainly focused on networks with a fixed number of vertices and thus without a growing property. Ravasz and Barabási (RB) introduced a hierarchical model in a deterministic way [13,14]; furthermore, Iguchi and Yamada completely analyzed the RB model and removed some flaws in its arguments [25]. Kim *et al.* [18] and Noh *et al.* [19] introduced growing models, where the number of vertices and modules increases with time. All of the models in the papers [13,14,18,19] obey the power-law degree distribution.

In this paper, we present a growing network model by adding the growth rule and the PA rule in the model of Motter *et al.* The proposed model is a growth model for complex systems with a fixed hierarchical and modular structure. And different from other methods, our approach could also be considered as a naturally generalized version of the BA model whose rules are considered to be very common in our society. The areas where our model can be more naturally applied are social systems, wide-area network (WAN) routing networks, large-scale logistic systems, intercontinental air transportation networks, etc., where the vertices are people, routers, local warehouse, and local airports, respectively, and the modules of different levels are cities, provinces, countries for social systems, local and regional routing subnets for WAN routing networks, local and transfer networks for logistic systems, local and connecting airlines for

*Author to whom all correspondence should be addressed. Electronic address: yjlee@iipc.zju.edu.cn

air transportation networks, and so on [24]. These systems have fixed hierarchical and modular structures which are all mainly determined by many geographic factors, but the vertices of them increase very quickly. Our model is designed according to the above system characteristics and is different from former studies. For example, the model of Watts *et al.* [23] and the model of Motter *et al.* [24] do not have the growth property; the RB model [13,14] has totally deterministic vertex growth characteristics which are not realistic in the real world; and in the model of Kim *et al.* [18] and the model of Noh *et al.* [19] both the modules and vertices increase with time. Therefore, these models cannot explain well the systems mentioned above. In those systems, the creation of connections between lower-level modules is easier and more frequent than that at higher levels. For example, in air transportation systems, opening a local airline is usually easier and more frequent than opening an intercontinental airline. In other words, comparing to the rapidly increasing number of infrastructural components, the upper-level organizations are much more stable, so the number of the upper-level organizations can be regarded as constants. In view of the dynamics of the vertex degrees and the local clustering coefficients, the proposed model exhibits both the properties of scale-free and modular structure. The degree distribution of the proposed model possesses a power-law distribution with $\gamma \in [2,3]$ and the clustering function $C(k) \sim k^{-\beta}$, consistent with most empirical data. We also briefly study the dynamics of module size which is defined as the number of members in a module. As the proposed model grows, the module size distribution shows a power-law distribution like the network model studied previously [19].

The rest of the paper is organized as follows. In Sec. II, we introduce the growing network model. The dynamic properties of the model can be changed by some parameters. Then analytic results about the dynamics of vertex degree and module size are provided in Sec. III, and the numerical simulation is given in Sec. IV. The results from the theoretical analysis and the numerical simulation are very close. As an application, the proposed model is used in Sec. V to predict the growth trends of a complex network with modular and hierarchical structure. Our work is summarized in Sec. VI.

II. GROWING NETWORK MODEL

In this section we present a growing network model with fixed modular and hierarchical structure. As is shown in Fig. 1, our model has a treelike structure. At the lowest level (i.e., the first level) of the network, there are several groups of vertices, forming modules of this level. Those modules are in turn grouped to constitute higher-level modules and so on. This process continues until at a certain level there is only one "big" module. This level is defined as the top level of the network, containing all the information about the network in a hierarchical way.

To simplify the description, we assume that, at all levels except the top level, every n module of the hth level forms a module of the $(h+1)$th level, where n is independent of h. Then a network with M levels will contain n^{M-1} lowest-level modules. The parameters M and n are fixed in our model.

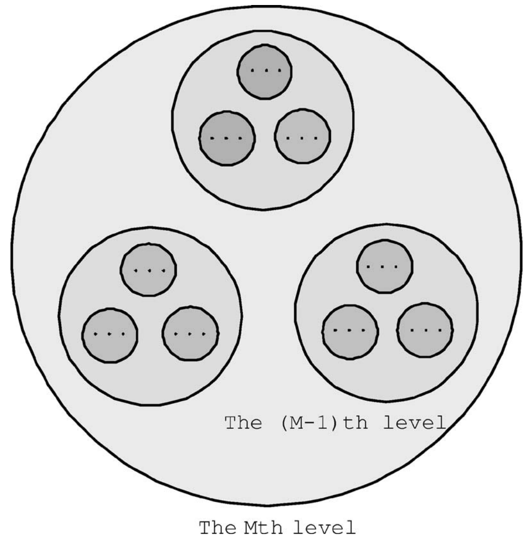

FIG. 1. The sketch map of the model with M levels.

Two stochastic events are considered in our model to make the network "grow"—i.e., the creation of a new vertex connected to some other existing vertices and the creation of a new connection attached between two existing vertices. According to the hierarchical and modular structure of our model, these events can be yielded by two kinds of operations: in-module connection and between-module connection. With the in-module connection a new vertex may be created in a module and connected to several vertices in the same module. With the between-module connection two vertices coming from different modules will be connected. All the operations are conducted in a random way: the connection type is determined with a predefined in-module-connection probability q_h for the hth-level modules, $h=1,2,\ldots,M$, and the vertex selection follows the PA rule.

The whole process of network growing can be described with the following top-down view. Initially we assume that there are m_0 fully connected vertices in each lowest-level module. We set $h=M$ and let Q_h be the currently selected module in the hth level in which a network growth operation will be conducted. Of course Q_M is the sole module at the top level.

(i) *Connection-type selection*. For the submodules of module Q_h at the hth level, the in-module connection type is selected with the probability q_{h-1} or the between-module connection type is selected with the probability $1-q_{h-1}$.

(ii) *Module selection*. A $(h-1)$th-level submodule of the module Q_h is selected with a uniform probability if the in-module connection type is chosen or two $(h-1)$th-level submodule of the module Q_h are selected also with a uniform probability if the between-module connection type is chosen.

(iii) *Vertex connection*. In the case of in-module connection, a new vertex is created if the $(h-1)$th-level submodule is at the lowest level and then connected to m existing vertices in this submodule by following the PA rule. Otherwise, if the $(h-1)$th level is not the lowest level, substitute h by $h-1$ and let this submodule be the selected module Q_h; then, return to the first step (connection type selection). In the case

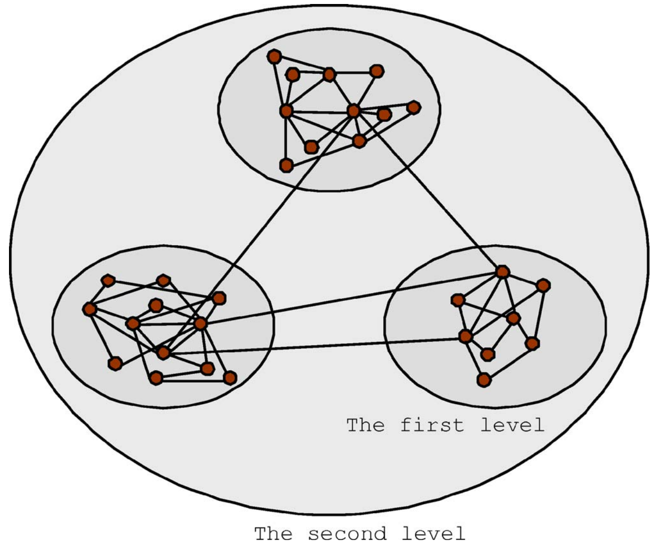

FIG. 2. (Color online) The network model with $M=2$, $n=3$, $m=m_0=2$, $q_1=0.9$, and growth time $T=30$. It is grown up to $N=31$ vertices.

of between-module connection, on the other hand, two existing vertices are selected, respectively, from the submodules determined in step 2 (module selection) in terms of the PA rule and then connected each other.

The above procedures can be repeated for a number of times to emulate network growing processes in the real world. Figure 2 demonstrates a configuration of our model with $M=2$, $n=3$, $m=m_0=2$, $q_1=0.9$, $T=30$, and grown up to $N=31$ where N is the number of total vertices.

As is discussed in Sec. I, in real-world networks, the creation of connections between lower-level modules is easier and more frequent than that at higher levels. This fact can be expressed in our model by the following inequality:

$$1-q_{h+1} \ll q_{h+1}(1-q_h). \tag{1}$$

Inequality (1) can be regarded as a constraint on the in-module connection probabilities q_h, $h=1,2,\ldots,M-1$. Two necessary conditions for inequality (1) holding are that $\forall h$, $q_h < q_{h+1}$; i.e., the in-module connection probability $\{q_1,q_2,\ldots,q_M\}$ must be a monotonically increasing sequence and $\forall h>1$, $q_h>0.5$. These conditions imply that, in order to preserve a modular and hierarchical structure, the events of vertex creation in a single module must occur much more frequently in the network-growing process than the creation of a connection attached between two existing vertices in different modules; otherwise, the clusters and hierarchies of the network will vanish soon.

The dynamical properties of the network model will be analyzed in the next two sections.

III. THEORETICAL ANALYSIS

In this section we will investigate some mathematical properties of the proposed model on the basis of the simplified assumption presented in Sec. II; i.e., except the lowest level in the model, all the modules at each level are composed of the same (fixed) number of their lower-level modules. We would like to investigate the characteristics of the network model when new vertices and connections are allowed to add into the network randomly. In this situation, the degree of each vertex and the number of vertices in each module in the model increase as the network grows, presenting a dynamical prospect.

In our model, all the modules at each level have the same uniform selection probability, and the in-module connection probability q_h, $h=1,2,\ldots,M-1$, satisfies inequality (1). As a result, the modules at the hth level can be considered to have the same size of module degree, denoted by $(\Sigma k_j)_h^M$, where k_j is the degree of vertex j in a module and M is the total number of levels in the network. We can easily get the relationship between the module degree of the hth level and that of the $(h+1)$th level:

$$\left(\sum k_j\right)_{h+1}^M = n\left(\sum k_j\right)_h^M. \tag{2}$$

According to the connection generation principle presented in Sec. II, the vertex degree dynamics in the hth-level module—i.e., the increasing rate of the connections of a vertex in the selected module at the hth level—denoted by $(\partial k_i/\partial t)_h^M$, satisfies the following recursive equations:

$$\left(\frac{\partial k_i}{\partial t}\right)_{h+1}^M = q_h\left(\frac{\partial k_i}{\partial t}\right)_h^M + \frac{2}{n^{M-h}}(1-q_h)\frac{k_i}{(\sum k_j)_h^M}, \tag{3}$$

$$\left(\frac{\partial k_i}{\partial t}\right)_1^M = \frac{1}{n^{M-1}}\frac{mk_i}{(\sum k_j)_1^M}. \tag{4}$$

The first term on the right-hand side of Eq. (3) represents the increasing rate of the connections within the selected module of the hth level, based on the in-module-connection probability q_h, and the second term concerns the increasing rate of the connections between those hth-level modules that are included in a module of the $(h+1)$th level, on the basis of the between-module-connection probability $1-q_h$. Equation (4) implies that the connections in any of the lowest-level modules obey the PA rule. But from Eqs. (2)–(4) one is unable to get the module degree dynamics $(\partial k_i/\partial t)_M^M$ from the top-level view, because the total number of connections, $(\Sigma k_j)_M^M$, in the network is unknown. Corresponding to Eq. (3), however, $(\Sigma k_j)_{h+1}^{h+1}$ can also be divided into two parts:

$$\left(\sum k_j\right)_{h+1}^{h+1} = q_h\left(\sum k_j\right)_h^h + 2(1-q_h)t, \tag{5}$$

$$\left(\sum k_j\right)_1^1 = 2mt, \tag{6}$$

where t is the network growth time. If we define

$$a_1 = m, \tag{7}$$

$$b_1 = 2m, \tag{8}$$

$$a_{h+1} = q_h a_h + 2(1-q_h), \tag{9}$$

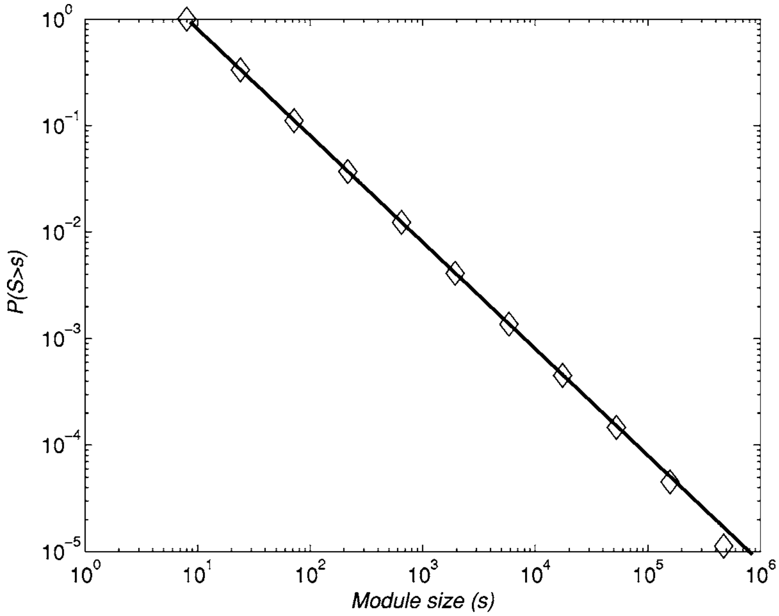

FIG. 3. Plots of the module size distribution $P(S>s)$ of the model with $M=10$, $n=3$, and 8 vertices in each smallest module. The slope of the solid line is 1.0.

$$b_{h+1} = q_h b_h + 2(1 - q_h), \quad (10)$$

then from Eqs. (2)–(6), the degree dynamics of the network can be easily described as

$$\left(\frac{\partial k_i}{\partial t}\right)_M^M = \frac{a_M}{b_M}\frac{k_i}{t}, \quad (11)$$

where a_M and b_M can be derived from Eqs. (7)–(10). The result allows us to conclude that the distribution function $P_M(k) \sim k^{-\lambda_M}$—i.e., the power-law distribution, with exponent

$$\lambda_M = 1 + b_M/a_M. \quad (12)$$

Because it satisfies $1 \leqslant b_M/a_M \leqslant 2$, we can find that the exponent $\lambda_M \in [2,3]$ which fits most empirical data in [4,8–10,26].

The fact that the exponent λ_M is between 2 and 3 can be easily understood. When the number of new connections created within the lowest-level modules are far more than those between different modules—i.e., the in-module-connection probability $q_h \to 1$, for all $h=1,2,\ldots,M-1$—the model will be close to the BA model and the exponent $\lambda_M \to 3$. Whereas if at some level—say, the hth level—most of the new connections are created between modules—i.e., $q_h \to 0$—or the number of levels in the network is very large—i.e., $M \to \infty$—then, $\lambda_M \to 2$.

Since the modules at any level have an identical selection probability P_h^S, all the modules of this level can be considered to have the same module size. We define S_h and R_h as the module size and the number of modules at the hth level, respectively. From the model structure, it can be easily derived that

$$S_{h+1} = nS_h, \quad (13)$$

$$R_{h+1} = \frac{1}{n}R_h. \quad (14)$$

From Eqs. (13) and (14), we can get

$$\ln S_h = -\ln R_h + a_0, \quad (15)$$

where a_0 is independent of h. And according to the definitions of S_h and R_h, we have

$$P(S_h) \sim R_h. \quad (16)$$

Combining Eq. (15) with Eq. (16), we know that the module size obeys the power-law distribution

$$P(S_h) \sim S_h^{-1}. \quad (17)$$

Numerical analysis also shows that $P(S>s)$ almost satisfies

$$P(S > s) \sim s^{-1}, \quad (18)$$

which is shown in Fig. 3. Such a result is similar to that of some real-world networks [19,27].

IV. NUMERICAL SIMULATION

The analytic result in Sec. III has been validated by numerical simulation. In the first example, we set $M=4$, $n=3$, $q_1=0.7$, $q_2=0.8$, $q_3=0.9$, $m=m_0=2$, and $T=7000$, and in the second example, $M=5$, $n=3$, $q_1=0.7$, $q_2=0.9$, $q_3=0.95$, $q_4=0.995$, $m=m_0=2$, and $T=6000$. We present the numerical data in Fig. 4. It shows that the results of the two examples satisfy the power-law degree distribution with a different exponent; the first has the exponent $\gamma \approx 2.5$ and the second the exponent $\gamma \approx 2.6$, which are very close to the analytic results from Eq. (12).

We also investigated the clustering function $C(k)$ of the proposed model with various parameters. Roughly speaking, $C(k)$ is likely to obey the distribution of $k^{-\beta}$. In the early stage of the growing process (i.e., each smallest module has a relative small number of vertices), β is close to 1, as is shown in Figs. 5(a) and 5(d). As the network grows, β decreases from 1 to 0, shown in Figs. 5(a)–5(c). The reason is that as the network grows, the connection between modules

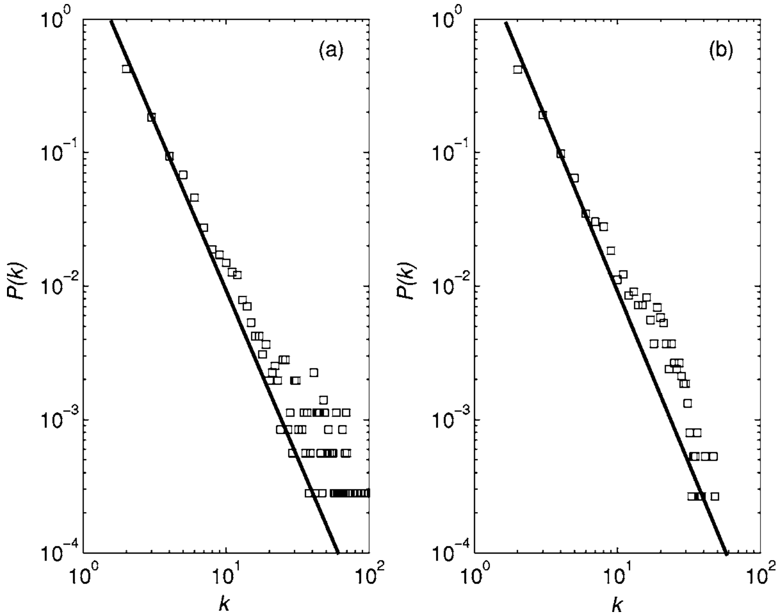

FIG. 4. Plots of the degree distribution $P(k)$: (a) The first example with $M=4$, $n=3$, $q_1=0.7$, $q_2=0.8$, $q_3=0.9$, $m=m_0=2$, $T=7000$, and $N=3562$. The slope of the solid line is 2.5. (b) The second example with $M=5$, $n=3$, $q_1=0.7$, $q_2=0.9$, $q_3=0.95$, $q_4=0.995$, $m=m_0=2$, $T=6000$, and $N=3775$. The slope of the solid line is 2.6.

becomes stronger and the connection within the lowest-level modules is weaker due to the increasing number of vertices. So the clustering phenomenon in the network is not so clear as in the early growing stage. The result means that many of the real-world networks are in their early growing stage because β is close to 1 for them [18]. The length of the early stage is mainly determined by the number of levels, M; the in-module-connection probability q_h; and the number of connections, m, added while a vertex is created. The larger the parameters, the longer the early stage is. So some real-world networks may have to spend a very long early stage.

V. TREND PREDICTION USING OUR MODEL

Here we will study the growth trends of those real-world networks by using our network model with corresponding parameters. We consider a medium size network with $M=4$, $n=3$, $q_1=0.5$, $q_2=0.95$, and $q_3=0.995$ and plot in Fig. 6 the change of the exponents of $C(k)$ of any module at each level (except the first level, because the modules at the first level do not have a modular structure) with respect to the increasing number of vertices in the network. In Fig. 6 we find that as the network grows, lower modules lose their intermodular structure quickly, and when the growth time T is large enough, the modules (and consequently the hierarchies) of the network will eventually merge. That is to say, for example, one day, in human society the administrative organizations governing villages, cities, provinces, and countries will vanish as a result of the rapid development of communication among people. In fact, even at present, we have seen a phenomenon that the boundaries of different cities in the developed countries are getting more and more blurry with

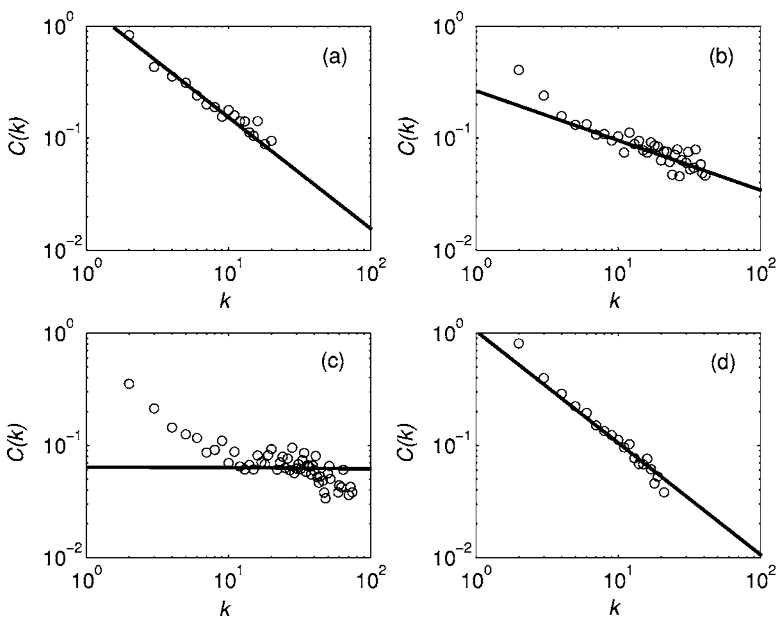

FIG. 5. The clustering function $C(k)$ with the parameters $m=m_0=2$, $q_1=0.7$, $q_2=0.8$, and $q_3=0.9$: (a) $M=4$, $n=3$, $T=400$, and $N=240$; (b) $M=4$, $n=3$, $T=1600$, and $N=859$; (c) $M=4$, $n=3$, $T=3200$, and $N=1632$; (d) $M=4$, $n=5$, $T=1600$, and $N=1049$. The slopes of solid lines are 1.0 in (a) and (d), 0.5 in (b), and 0 in (c).

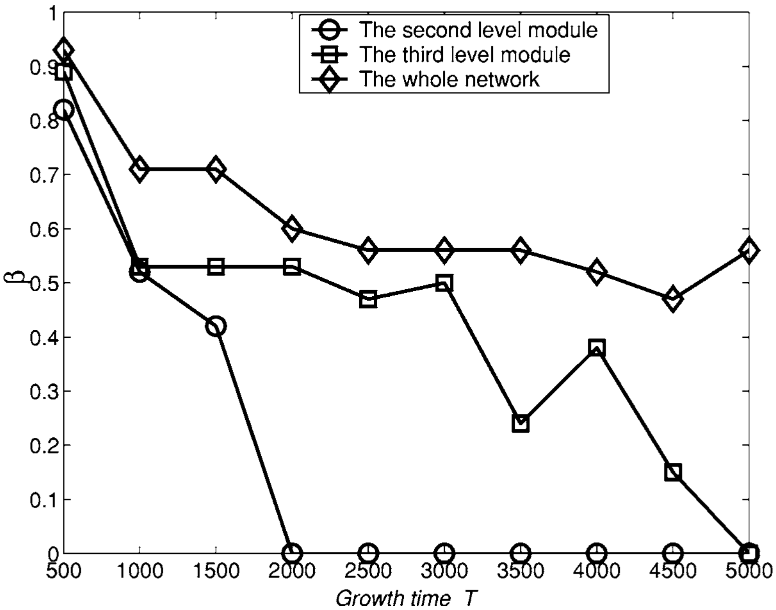

FIG. 6. Plots of the trends of exponents for modules at each level except the lowest level in a network with $M=4$, $n=3$, $q_1=0.5$, $q_2=0.95$, and $q_3=0.995$.

the development of local and long-distance transportation systems.

VI. SUMMARY

This paper proposes a hierarchical and modular network model by adding the growth principle and the PA principle into the model of Motter *et al.* The proposed model has increasing vertices, a fixed number of modules, and a hierarchical structure. The vertices of the model form the lowest-level modules which in turn constitute higher-level modules hierarchically. The creation of connections between two vertices in a single module or in two different modules of the same level obeys the PA rule. With theoretical analysis and numerical simulation, it is shown that the degree distribution, the module size distribution, and the clustering function of the model possess a power-law property which is similar to that in many real-world networks. The model has been used to predict the growth trends of the real-world networks with modular and hierarchical structures. By comparing this model with those real-world networks, an interesting conclusion is found that many real-world networks are in their early stages of development, and when growth time is large enough, the modules and levels of the networks will be ultimately merged.

It is noteworthy that, as the lower modules disappear, the number of network levels will decrease and the between-module-connection probability of original higher-level modules will increase correspondingly. This can be considered as an interim process between two relatively stable stage of the network growing, i.e., the initial stage at which there is a very clear modular and hierarchical structure and the ultimate stage at which all the infrastructures and superstructures vanish. Unfortunately, our model is unable to characterize this interim period. This is an interesting topic about the development of complex networks, and we will study it in our subsequent work [28].

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 valuable discussions about the ideas presented in this paper. This work was supported by the National 973 Program of China (Grant No. 2002CB312200).

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[3] A. Barrat and M. Weigt, Eur. Phys. J. B **13**, 547 (2000).
[4] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[5] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002).
[6] C. M. Song, S. Havlin, and H. A. Makse, Nature (London) **433**, 392 (2005).
[7] L. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas, e-print cond-mat/0505185.
[8] A.-L. Barabási and Z. N. Oltvai, Nature (London) **5**, 101 (2004).
[9] S. Eubank, H. Guclu, V. S. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, Nature (London) **429**, 180 (2004).
[10] A. Barrat, M. Barthélemy, R. P. Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004).
[11] S. A. Pandit and R. E. Amritkar, Phys. Rev. E **60**, R1119 (1999).
[12] A. Clauset, Phys. Rev. E (to be published).
[13] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and

A.-L. Barabási, Science **297**, 1551 (2002).
[14] E. Ravasz and A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003).
[15] L. F. Costa, Phys. Rev. Lett. **93**, 098702 (2004).
[16] H. J. Zhou, Phys. Rev. E **67**, 041908 (2003).
[17] S.-W. Son, H. Jeong, and J. D. Noh, e-print cond-mat/0502672.
[18] D.-H. Kim, G. J. Rodgers, B. Kahng, and D. Kim, Physica A **351**, 671 (2005).
[19] J. D. Noh, H.-C. Jeong, Y.-Y. Ahn, and H. Jeong, Phys. Rev. E **71**, 036131 (2005).
[20] E. M. Jin, M. Girvan, and M. E. J. Newman (unpublished).
[21] B. Skyrms and R. Pemantle, Proc. Natl. Acad. Sci. U.S.A. **97**, 9340 (2000).
[22] A. Grönlund and P. Holme, Phys. Rev. E **70**, 036108 (2004).
[23] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Science **296**, 1302 (2002).
[24] A. E. Motter, T. Nishikawa, and Y.-C. Lai, Phys. Rev. E **68**, 036105 (2003).
[25] K. Iguchi and H. Yamada, Phys. Rev. E **71**, 036144 (2005).
[26] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[27] A. Arenas, L. Danon, A. D. Guilera, P. M. Gleiser, and R. Guimerà, Eur. Phys. J. B **38**, 373 (2004).
[28] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).