Eur. Phys. J. B **60**, 521–528 (2007)
DOI: 10.1140/epjb/e2008-00014-4

THE EUROPEAN PHYSICAL JOURNAL B

Accelerating networks: Effects of preferential connections

Y.-P. Jeon[1] and B.J. McCoy[2,a]

[1] Center for Advanced Engineering Fibers and Films, Clemson University, Clemson, SC 29634, USA
[2] Department of Chemical Engineering and Materials Science, University of California, Davis, California 95616, USA

Received 28 July 2007 / Received in final form 13 September 2007
Published online 16 January 2008 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2008

Abstract. Networks are commonly observed structures in complex systems with interacting and interdependent parts that self-organize. For nonlinearly growing networks, when the total number of connections increases faster than the total number of nodes, the network is said to accelerate. We propose a systematic model for the dynamics of growing networks represented by distribution kinetics equations. We define the nodal-linkage distribution, construct a population dynamics equation based on the association-dissociation process, and perform the moment calculations to describe the dynamics of such networks. For nondirectional networks with finite numbers of nodes and connections, the moments are the total number of nodes, the total number of connections, and the degree (the average number of connections per node), represented by the average moment. Size independent rate coefficients yield an exponential network describing the network without preferential attachment, and size dependent rate coefficients produce a power law network with preferential attachment. The model quantitatively describes accelerating network growth data for a supercomputer (Earth Simulator), for regulatory gene networks, and for the Internet.

PACS. 89.75.Fb Structures and organization in complex systems – 05.65.+b Self-organized systems – 87.23.Ge Dynamics of social systems

1 Introduction

A complex system can be defined as a system with many interacting and interdependent parts having emergent self-organization [1]. New technologies and rapidly changing societies, as well as biological evolution, increase the need for a better understanding of these complex systems and their structure. Data for complex systems [2] often reveal network structure, consisting of many connections among many nodes [1,3,4]. Among the models for network simulation, the Erdos and Renyi model [5,6] generates undirected random connection networks. The small-world network model [7] is an interpolation between regular lattice models and random graphs [8]. Compared to the random graph model the small world network has a much larger clustering coefficient [7], which is the probability that two randomly chosen nodes have a connection with each other. The Barabasi-Albert network model [9], a preferential growth model producing a power law structure, clarified the time dependence of power law networks. Generally, networks evolve with time [1,3,4,10], typically growing unless they undergo breakage of connections or removal of nodes. For growing networks, the number of nodes and number of connections are typically increasing with time [2,3,9,10]. As few networks show linear growth with their size [11], the majority of networks grow nonlinearly, with the total number of connections increasing faster than the total number of nodes. Such networks are called accelerating networks [3,12,13].

The present aim is to propose a different approach for the dynamics of linear and nonlinear growing networks based on distribution kinetics [14,15]. For a class of growing network models [2,12,16], where the addition of new nodes leads to power law structure [9,17,18], the distribution follows a power law, $p(\xi) \sim \xi^{-\lambda}$. Examples of power law networks [19] are communication networks, such as the World Wide Web and the Internet [20,21], citation networks in the scientific literature [22,23], collaboration networks [24,25], and metabolic reaction networks [11].

The current model, allowing multiple connections between any two nodes, is more general than the connection-limited networks, allowing only one connection between any two nodes. An example of the model network is an airline network where several flying routes exist from one airport to another. For connection-limited networks, the most effective structure is the saturated state, where all nodes are connected with all other nodes [12]. Such saturated structures, where a clustering coefficient is unity, may not be realistic even for the connection-limited networks, because real-world complex systems are trade-offs between cost and efficiency, and unlinked node pairs may be present. For instance, if a computer is directly connected to the main server, surfing the Internet will be more

[a] e-mail: bjmccoy@ucdavis.edu

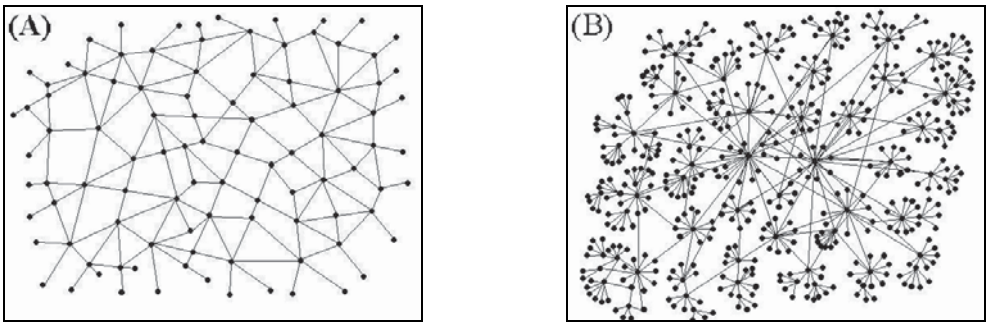

Fig. 1. Two types of networks: (A) random (exponential) and (B) scale-free (power law) networks.

than 100 times faster than the normal broadband high-speed Internet connection. For billions of Internet users and given the cost of direct connections, this would be unrealistic; thus the Internet, a representative communication network, follows the second best effective structure, the power law network.

2 Model

We begin with the general concept of frequency distribution of nodes and connections, in which nodes and connections are added and established for network growth, and removed and eliminated for breakage [14]. To exemplify fundamental ideas, we will show even simple models can produce diverse behaviors. Connections (links or edges) are binary interactions between nodes that are nondirectional and of indeterminate length. For such connections, the nodal-linkage distribution $p(\xi, t)d\xi$ is defined as the number of nodes at time t with number of connections in the interval ξ to $\xi + d\xi$. Because ξ is a positive integer, the distribution is discrete, but for a large number of connections one can substitute the discrete distribution with a continuous distribution (replacing summations for the discrete distribution with integrals for a continuous distribution). The distribution of the number of connections in the interval ξ to ξ+dξ can be expressed as $1/2\ \xi p(\xi, t)d\xi$. Since each connection is associated with two nodes in nondirectional networks, the factor 1/2 appears. This is similar to polymer molecular weight distributions [26], where the number of macromolecules having mass in the range $(x, x + dx)$ is $p(x)dx$ and the mass of macromolecules in the same interval is $xp(x)dx$.

The moments of the nodal-linkage distribution are defined as,

$$p^{(n)}(t) = \int p(\xi, t)\xi^n d\xi \tag{1}$$

where the integration limits are determined by the domain of $p(\xi, t)$. From the definition, the total number of nodes and total number of connections are $p^{(0)}(t)$ and $1/2\ p^{(1)}(t)$, respectively, and the average number of connections per node, $1/2\ p^{avg}(t)$, where $p^{avg}(t) = p^{(1)}(t)/p^{(0)}(t)$. The variance and polydispersity index in terms of the second moment are defined as $p^{var} = p^{(2)}/p^{(0)} - p^{avg2}$ and $p^{pd} = p^{(2)}p^{(0)}/p^{(1)2}$, providing further information for the character and shape of the distribution.

To estimate the maximum number of connections for a connection-limited network, consider a simple saturated network with four nodes (three connections each where any two nodes have only one connection between them). Three connections for the first node, two connections for the second node, and one connection for the third node can be counted without repeating, and the summation of these gives the total number of connections. Therefore, the maximum possible number of connections, $1/2p^{(1)}(t)$, for a connection-limited network expressed as an arithmetic progression from 1 up to $(p^{(0)}(t)-1)$ is $1/2\ p^{(0)}(t)(p^{(0)}(t)-1)$ [3,13].

We will focus on the network growth for two classifications: exponential networks (single-scale), randomly connected in the absence of preferential attachment, and power law (scale-free) networks constructed by preferential attachment (Fig. 1). The distribution of exponential networks, for instance, Gaussian, binomial, or Poisson distributed networks, is unimodal (peaked) with well-defined moments. Therefore, statistical properties such as mean and variance can be easily defined and measured. The distribution of power law networks has a power law expression, $p(\xi) \sim \xi^{-\lambda}$, where λ is usually a positive constant. The moments of such networks are not defined on the interval $(0, \infty)$, because they do not have an inherent scaling factor. However, if the evolving power law network has an expanding finite domain, the moment integral in equation (1) can be defined within the domain. For that reason, moments of the power law networks within a finite domain are defined as,

$$p^{(n)}(t) = \int_0^{\xi_m(t)} p(\xi, t)\xi^n d\xi \tag{2}$$

where $\xi_m(t)$ represents the maximum number of connections, generally a function of time.

Our aim is to develop a framework that determines the evolution of the two types of network. We will approach the problem with knowledge from distribution kinetics based on population dynamics, which has proven a productive approach to polymerization and depolymerization [27], particulate fragmentation and aggregation ([26]), and crystal growth and dissolution [28]. By this method, we can obtain solutions for numerous interesting systems,

and show the effect of the parameters that govern the network evolution.

3 Distribution kinetics

We describe network structure based on the association-disassociation process, written below as a reaction-like expression. Connections are added one at a time to available nodes, with the possibility that connected nodes, or indeed entire networks, might coalesce by such connecting processes. The addition or removal of connections can be written as a reversible rate process,

$$P(\xi) + P(\xi') \underset{k_d(\xi)}{\overset{k_g(\xi)}{\rightleftharpoons}} P(\xi+1) + P(\xi'+1) \tag{3}$$

where $P(\xi)$ schematically represents a node with ξ connections. A connection can be formulated by the interaction between two nodes. We propose the rate coefficients $k_g(\xi)$ and $k_d(\xi)$ for addition (growth) and removal (dissociation), respectively, by adopting power expressions, which are generally considered to depend on the number of connections,

$$k_g(\xi) = \gamma\xi^{\lambda} \quad \text{and} \quad k_d(\xi) = \kappa\xi^{\nu} \tag{4}$$

where the constants γ, κ, λ, and ν are positive definite. The process of equation (3) is unchanged by replacing $\xi - 1$ with ξ or $\xi' - 1$ with ξ'. In addition to expressions for connection formulation or removal, expressions for rates of node generation or loss are also required in the population dynamics equation. Constructing the governing equations for networks is similar to polymer [27], and crystallization [28] kinetics. For example, the loss of $P(\xi)$ on the left-hand side of equation (3) is the product between $p(\xi, t)$ and $p(\xi', t)$, however, if $P(\xi')$ is abundant, we can assume that $P(\xi)$ is the limiting reactant. Therefore, the loss of $P(\xi)$ is expressed as,

$$k_g(\xi)p(\xi, t). \tag{5}$$

Similarly, the removal of a connection between $P(\xi)$ and $P(\xi')$ is proportional to $p(\xi, t)$. The node addition or node removal rate with ξ_i connections can be expressed as $I_i(t)\delta(\xi - \xi_i)$, or can be incorporated into boundary conditions. Based on these preliminary concepts we write the population dynamics equation for equation (3) with generation and loss terms,

$$\begin{aligned}\partial p(\xi,t)/\partial t = {} & \gamma[(\xi-1)^{\lambda}p(\xi-1,t) - \xi^{\lambda}p(\xi,t)] \\ & + \kappa[(\xi+1)^{\nu}p(\xi+1,t) - \xi^{\nu}p(\xi,t)] + \Sigma_{i=0} I_i(t)\delta(\xi-\xi_i).\end{aligned} \tag{6}$$

By substituting rate coefficients in equation (4), we obtain equation (6), similar to a master equation with first-order kinetics [29]. Similar rate equation approaches for the network analysis were introduced previously [16,30]. The population dynamics equation for network growing process is similar to how crystallization [28] or polymerization [26] affords growth by monomer addition for clusters or polymers. Because we will describe node insertion using the boundary conditions, we set the source terms to zero, $I_i = 0$. We expand the distribution in $\xi \pm 1$ in a series around ξ, as in other distribution kinetics applications [26,28], and obtain a Fokker-Planck equation from equation (6),

$$\begin{aligned}\partial p(\xi,t)/\partial t = {} & \partial[(k_d(\xi) - k_g(\xi))p(\xi,t)]/\partial\xi \\ & + 1/2\partial^2[(k_d(\xi) + k_g(\xi))p(\xi,t)]/\partial\xi^2 + \ldots\end{aligned} \tag{7}$$

where we have omitted third- and higher-order terms.

4 Exponential networks: absence of preferential attachment

We keep up to the second-order terms in equation (7),

$$\begin{aligned}\partial p(\xi,t)/\partial t = {} & \partial[(k_d(\xi) - k_g(\xi))p(\xi,t)]/\partial\xi \\ & + 1/2\partial^2[(k_d(\xi) + k_g(\xi))p(\xi,t)]/\partial\xi^2.\end{aligned} \tag{8}$$

A one-dimensional random walk and its expression as a convective diffusion equation are similar to equations (6) through (8), suggesting how a Gaussian distribution for the exponential network is obtained when the rate coefficients are constants [31,32], $k_g(\xi) = \gamma$ and $k_d(\xi) = \kappa$. For this case, connections are randomly established and yield exponential networks. By substituting a "velocity", $v = (\gamma - \kappa)$, and a "diffusivity", $D = (\gamma + \kappa)/2$, into equation (8), the convective diffusion equation can be expressed as,

$$\partial p(\xi,t)/\partial t = -v\partial[p(\xi,t)]/\partial\xi + D\partial^2[p(\xi,t)]/\partial\xi^2. \tag{9}$$

In stochastic theory, equation (9) is also called the Fokker-Planck or Kolmogorov forward equation, and diffusivity and velocity correspond to the constant infinitesimal mean and variance [33]. By a Fourier transformation of equation (9) and the initial condition, $p(\xi, t = 0) = p_o^{(0)}\delta(\xi - \xi_o)$, which means $p_o^{(0)}$ nodes with ξ_o connections exist initially, the exact solution can be obtained. Two additional boundary conditions are required to solve the forward equation. The first condition is $p(\xi \to \infty, t) = 0$, which means no node can have an unlimited number of connections (there should be a maximum number of connections per node). Because ξ is a positive integer ($\xi \geq 0$) for the network systems, a typical second condition, $p(\xi \to -\infty, t) = 0$, is not realistic. If the solution peak, located at ξ_o initially, is far enough away from $\xi = 0$, the boundary condition does not affect the peak. In convective diffusion theory [32], if the Peclet number for this system defined as $N_{Pe} = v\xi_o/D = 2(\gamma - \kappa)\xi_o/(\gamma + \kappa)$ is much greater than 1, the boundary condition [34] can be approximated by $p(\xi \to 0, t) = 0$, which ensures that the network excludes nodes without connection. As ξ_o, where the initial distribution is positioned, becomes larger, the Peclet number is larger.

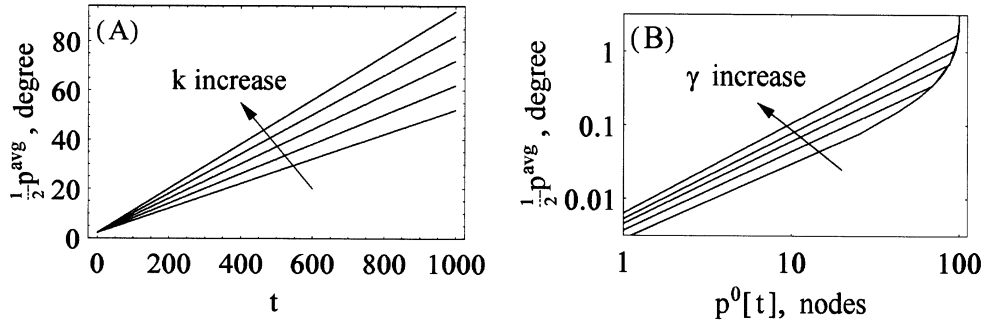

Fig. 2. Evolution of the Gaussian (A) and Poisson (B) distributed network growth based on the moment results in equations (11)–(13) (A) and (17)–(19) (B): $p_o^{(0)} = 100$, $\xi_o = 4$, and the growth rate $k = \gamma - \kappa$, increases in steps of 0.02 from 0.1 to 0.2 for (A), and $p_o^{(0)} = 100$ and γ increases in steps of 3 from 1 to 13 for (B).

The solution for the convective diffusion equation, equation (9), can be approximated by a Gaussian distribution for ξ and t, if $N_{Pe} \gg 1$,

$$p(\xi, t) = [p_o^{(0)}/(4\pi/N_{Pe})^{1/2}] \exp[-(\xi/\xi_o)^2 N_{Pe}/4]. \quad (10)$$

By applying integration in equation (2), the moments of equation (10) are readily found as

$$p^{(0)}(t) = p_o^{(0)} \quad (11)$$

$$p^{(1)}(t) = p_o^{(0)}[(\gamma - \kappa)t + \xi_o]. \quad (12)$$

The average moment and the variance are

$$p^{avg}(t) = (\gamma - \kappa)t + \xi_o \quad (13)$$

$$p^{var}(t) = (\gamma + \kappa)t. \quad (14)$$

If the growth rate coefficient, $k_g(\xi) = \gamma$, is greater than the dissolution rate coefficient, $k_d(\xi) = \kappa$, nodes are being connected, the average number of connections, corresponding to the degree, $1/2p^{avg}(t)$, increases, and the network grows. If $\gamma < \kappa$, the average moment of connections decreases and the network deteriorates by connection removal. For either growth or breakage, the network variance increases according to equation (14). As described in equations (11) and (13), the degree for this network with fixed number of nodes does not change with network size because the number of nodes is constant.

The discrete Poisson distribution for the exponential network derives from more constrained conditions: irreversible network growth ($k_d = 0$), lack of source terms ($I_i = 0$), and constant rate coefficient ($k_g = \gamma$). With these restrictions, equation (6) can be written as,

$$\partial p(\xi, t)/\partial(\gamma t) = -p(\xi, t) + p(\xi - 1, t) \quad (15)$$

where only positive integers ($\xi > 0$) are considered. Equation (15) is a first-order difference-differential equation similar (but not identical) to governing equations in chain polymerization [27] and stirred-tank cascade modeling [35]. The initial and boundary conditions are $p(\xi, t = 0) = p_o^{(0)}\delta_{0\xi}$ and $p(\xi < 0, t) = 0$; the initial condition is expressed in terms of the Kronecker delta, representing unconnected node insertion, and the boundary condition ensures the variable ξ is positive. Equation (15) can be solved by Laplace transformation and the solution is similar to a Poisson distribution [27],

$$p(\xi, t) = p_o^{(0)} e^{-\gamma t}(\gamma t)^{\xi+1}/(\xi + 1)!. \quad (16)$$

The moments for the distribution by the calculation in equation (2) are,

$$p^{(0)}(t) = p_o^{(0)}(1 - e^{-\gamma t}) \quad (17)$$

$$p^{(1)}(t) = p_o^{(0)}(\gamma t - 1 + e^{-\gamma t}). \quad (18)$$

The average moment, which has a long time limit, γt, and the variance are,

$$p^{avg}(t) = (\gamma t - 1 + e^{-\gamma t})/(1 - e^{-\gamma t}) \quad (19)$$

$$p^{var}(t) = \gamma t e^{\gamma t}(e^{\gamma t} - \gamma t - 1)/(1 - e^{\gamma t})^2. \quad (20)$$

The variance also shows a linear time behavior (degree $\sim \gamma t$) for large values of time. Thus, similar to the Poisson distribution, the average and variance of equation (16) asymptotically reach the same expressions at long time.

Figure 2 shows how the degree of Gaussian and Poisson distributed exponential networks grows with time (Fig. 2A) and the number of nodes (Fig. 2B). With the absence of preferential attachment, the Gaussian distributed network in Figure 2A exemplifies how the total number of connections among a constant number of nodes increases with time, and the Poisson distributed networks in Figure 2B (log-log coordinates) demonstrates nonlinear accelerating growth of degree with network size (the total number of connections grows faster than the total number of nodes). As plotted in Figure 2B, if the growth rate, $k = \gamma - \kappa$, increases, the degree with a constant number of nodes increases. The diverging behavior of degree explains that the model allows multiple connections between nodes. The linear behavior on log-log coordinates (Fig. 2B) implies that the degree is proportional to a power of network size.

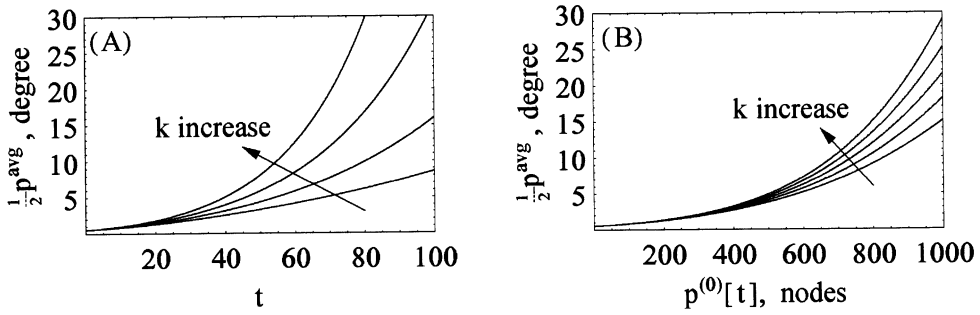

Fig. 3. Evolution of moments of the power law distributed network (Eqs. (24)–(26)) with initial condition, $p(\xi, t = 0) = 0$ and boundary condition, $p(\xi = 1, t) = p_o^{(0)} e^{t/\tau}$. The growth rate k increases in steps of 0.01 from 0.11 to 0.14 for parameters $\lambda = 1$, $p_o^{(0)} = 100$, $\xi_m = 1000$, $\tau = 10$.

5 Power law networks: effect of preferential attachment

Many complex systems have power law size distributions. The well-known mechanism to produce power law networks is preferential attachment [9] where a new node introduced into networks preferentially connects with highly connected ones. To describe the evolution of power law distributed networks, we truncate equation (7) to first-order with the power expression of the rate coefficients (Eq. (4), with $\lambda = \nu \neq 0$).

$$\partial p(\xi, t)/\partial t) \partial [G p(\xi, t)]/\partial \xi = 0 \tag{21}$$

where $G = (\gamma - \kappa)\xi^{\lambda}$ is the growth rate. This partial differential equation, similar to a continuity equation, is common to population balance modeling [36], and can be solved by Laplace transformation. We examine two different cases with different initial conditions. We first consider the initial condition $p(\xi, t = 0) = 0$, which means initially no nodes exist, and the boundary condition, $p(\xi = 1, t) = p_o^{(0)} e^{t/\tau}$, which represents the number of nodes with one connection increasing exponentially. The distribution for this set of initial and boundary conditions can be obtained by Laplace transformation as well as the method of characteristics,

$$p(\xi, t) = p_o^{(0)} \xi^{-\lambda} \exp[t/\tau + (1 - \xi^{1-\lambda})/k\tau(1 - \lambda)] u[t - (\xi^{1-\lambda} - 1)/k(1 - \lambda)] \tag{22}$$

where $u(x)$ is the unit step function defined as $u(x < 0) = 0$ and $u(x \geqslant 0) = 1$. The moments of equation (22) can be obtained by integration (Eq. (2)).

The analytical solution for general power λ is complicated, and we will show the evolution and network size dependence graphically. Here we list the simplest moment results where node connection probability is linearly proportional to node degree, $\lambda = 1$. For this linearly proportional connection probability, the nth moment obtained by integration within a finite domain is,

$$p^{(n)}(t) = p_o^{(0)} k\tau[\exp(t/\tau) - \exp(nkt)]/(1 - nk\tau). \tag{23}$$

Thus the zeroth, first, and average moments are,

$$p^{(0)}(t) = p_o^{(0)} k\tau[\exp(t/\tau) - 1] \tag{24}$$

$$p^{(1)}(t) = p_o^{(0)} k\tau[\exp(kt) - \exp(t/\tau)]/(k\tau - 1) \tag{25}$$

and

$$p^{avg}(t) = [\exp(kt) - \exp(t/\tau)]/\{(k\tau - 1)[\exp(t/\tau) - 1]\}. \tag{26}$$

As explained, the zeroth moment, $p^{(0)}(t)$, represents the total number of nodes, the total number of connections can be represented by $1/2\ p^{(1)}(t)$, and the degree is defined using the average moment, as degree $= 1/2p^{avg}(t) = 1/2p^{(1)}(t)/p^{(0)}(t)$. Clustering coefficient for the connection-limited networks can be defined as $C = 1/2p^{(1)}(t)/[p^{(0)}(t)(p^{(0)}(t) - 1/2]$, which is a ratio between the actual and maximum number of connections. Only a particular network [11] shows linear growth, where the number of connections is linearly proportional to network size. However, many networks display accelerating nonlinear growth, where the total number of connections grows faster than the total number of nodes, for example, the communication networks (the WWW and Internet), citation networks, and collaboration networks [20–25]. We now demonstrate such nonlinear, accelerating network growth showing evolution of a power law distribution. Figure 3A shows the time dependence of power law networks with proposed initial and boundary conditions, and Figure 3B demonstrates nonlinear accelerating growth of the degree with network size.

We next consider the initial condition $p(\xi, t = 0) = p_o^{(0)} \xi^{-\lambda}$, where connections are initially distributed as a power law, and the exponentially increasing boundary condition, $p(\xi = 1, t) = p_o^{(0)} e^{t/\tau}$. The network distribution for this set of initial and boundary conditions is,

$$p(\xi - t) = p_o^{(0)} \xi^{-\lambda} \{1 - \{1 - \exp[t/\tau + (1 - \xi^{1-\lambda})/k\tau(1 - \lambda)]\} u[t - (\xi^{1-\lambda} - 1)/k(1 - \lambda)]\}. \tag{27}$$

When $\lambda = 1$, the general nth moment obtained by integration is,

$$p^{(n)}(\tau) = p_o^{(0)}[\exp(nkt) - nk\tau \exp(t/\tau) + \xi_m^n (nk\tau - 1)]/[n(nk\tau - 1)] \tag{28}$$

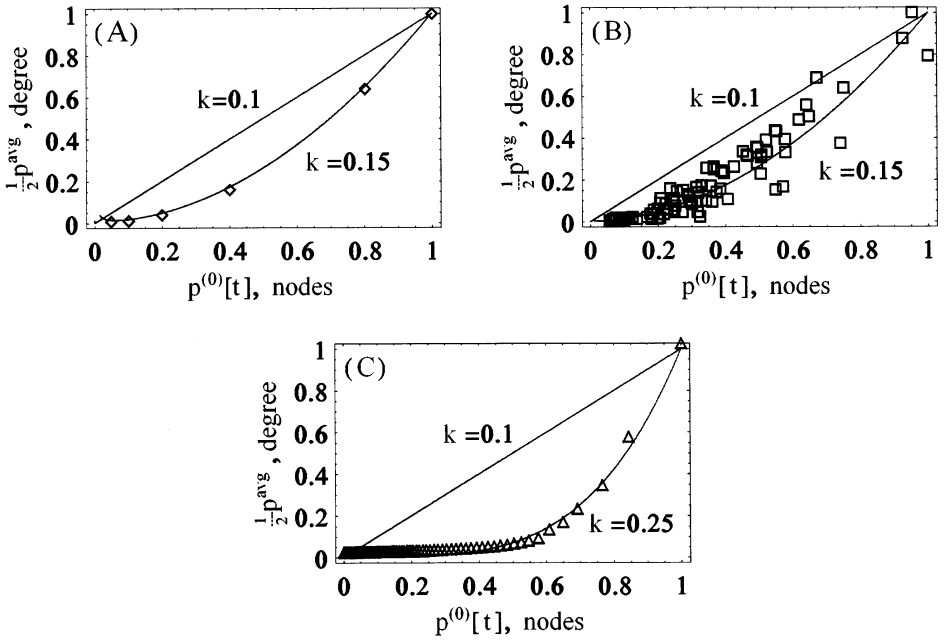

Fig. 4. The scaled degree versus total number of nodes for the model (Eqs. (24)–(26) with $p_o^{(0)} = 100$, $\xi_m = 1000$, and $\tau = 20$). The axes are normalized from 0 to 1 with values at $t = 100$. The lines are model predictions and symbols indicate data [12] for (A) Earth simulator ($\lambda = 2$), (B) regulatory gene networks ($\lambda = 2$), and (C) the Internet in 1999 ($\lambda = 2.2$). The predictions with $k = 0.15$ and 0.25 demonstrate that the model can describe real-world accelerating network growth. The predictions with $k = 0.1$ ($k\tau/2 = 1$) show the most effective network growth path for connection-limited networks.

and the zeroth, first, and average moments are,

$$p^{(0)}(t) = p_o^{(0)}\{\ln(\xi_m) - k[t + \tau(1 - \exp(t/\tau))]\} \quad (29)$$

$$p^{(1)}(t) = p_o^{(0)}[\exp(kt) - k\tau\exp(t/\tau) + \xi_m(k\tau - 1)]/(k\tau - 1) \quad (30)$$

and

$$p^{avg}(t) = [\exp(kt) - k\tau\exp(t/\tau) + \xi_m(k\tau - 1)]/\{(k\tau - 1)[\ln(\xi_m) - k\{t + \tau[1 - \exp(t/\tau)]\}]\}. \quad (31)$$

If k and τ are large, equations (28)–(31) reduce to equations (23)–(26), and the degree shows nonlinear accelerating behavior similar to Figure 3.

As expressed in equations (23)–(31) and Figure 3, the model for power law networks contains two parameters, k and τ, which allow quantitative description of many nonlinearly growing systems: As k (growth rate) increases, the network size and nonlinearity, the ratio between the total number of connections and nodes, increases. As τ (node addition intensity) increases the total number of connections exceeds the maximum number of connections of connection-limited networks, $p^{(0)}(t)(p^{(0)}(t) - 1)/2$, indicating multiple connections between nodes.

It is also interesting that when $\tau k/2 = 1$, the number of nodes represented by the zeroth moments in equations (24) and (29) is approximated as $p^{(0)}(t) \sim \exp[kt/2]$. The number of connections expressed by the first moments in equations (25) and (30) is written as $1/2p^{(1)}(t) \sim (\exp[kt/2])^2$, and therefore the number of connections increases quadratically with network size, $1/2p^{(1)}(t) \sim (p^{(0)}(t))^2$. Because the degree is defined by the average moment (degree$= 1/2p^{avg}(t)$ and $p^{avg}(t) = p^{(1)}(t)/p^{(0)}(t)$), this shows linear behavior with the total number of nodes, thus, degree $\sim p^{(0)}(t)$. The quadratic increase of number of connections with number of nodes was reported to apply for supercomputers and regulatory gene networks [12].

A vector-parallel high-performance computer developed by ESRDC (the Earth Simulator Research and Development Center)/NEC, the Earth Simulator, is registered as the world's fastest supercomputer with 35.61 TFlops (trillion operations per second) according to Linpack benchmark test results. The Earth Simulator consists of 640 supercomputers, one at each node, with 8 vector processors for a total of 5120 processors connected by a high-speed network with 12.3 GBytes data transfer speed. Recent research [12,37] indicates that more than 400 000 connections are required to connect 640 nodes by 83 000 wires.

For the gene regulatory network of single-celled prokaryotic organisms, general arguments on the network control indicate that regulatory gene number grows relatively fast as genome size increases; the number of transcriptional regulators scales quadratically with the total number of genes. Recent studies [38,39] on the gene regulatory network reveal that the percentage of regulatory genes increases from 2.5% to 9% as the network size grows from 4000 to 8000 bacterial genes. The model predictions are compared with the accelerating behaviors of the

Earth Simulator and the gene regulatory network [12] in Figure 4.

The Internet is an essential example of a growing network with technological and economical importance. A well-known property of the Internet is that the distribution of connections follows a power law behavior with power exponent 2.15–2.2 [15]. The Internet also shows a nonlinear accelerating growth behavior.

For an accelerating network with n connections among m nodes, for $b > 0$ the total number of connections is increased more than bn if the number of nodes increases bm. We have plotted this accelerating behavior in Figure 4 as the normalized average number of connections per node versus the total number of nodes

Real-world complex systems such as the Earth Simulator, the Internet, bacteria, or business organizations show nonlinear accelerating growth behavior [12], because of connection and organization costs. As a test we compare model results with data for a supercomputer (Earth Simulator), for regulatory gene networks [12], and for the Internet in 1999 [40]. As explained for the connection-limited networks, the most connected network is the most effective one, for example, when a new node introduced into networks has connections with all other nodes. This will appear as a straight line on a scaled plot of degree with network size. Predicting network growth path by using the model solution is straightforward and convenient as pictured in the scaled plots, Figures 4A–4C with different k. We normalized the axes, the degree and the number of nodes in Figure 4 by using values at $t = 100$ (which is sufficient time to see nonlinear accelerating behavior). When the parameter k is 0.1, the straight line, which corresponds to the most effective network for the connection-limited network, was obtained, i.e., $\tau k/2 = 1.0$. The model prediction for data in Figure 4 shows nonlinear growth of degree, $1/2p^{avg}$, with network size, $p^{(0)}(t)$, through quadratic increases in the total number of connections, $1/2p^{(1)}(t)$. The model adequately describes data for accelerating network growth of the Earth Simulator, regulatory gene networks ($k = 0.15$), and the Internet in 1999 ($k = 0.25$). As indicated in Figure 4, as k approaches to 0.1, which represents the most effective network, the network becomes more efficient, and in this regard, we presume that the Earth Simulator and the regulatory gene networks have more effective structures than the Internet.

6 Conclusions

What kind of mechanism do growing networks follow? How do they construct and maintain their exponential or power law structures? To answer these questions, we have suggested a model based on population balance dynamics (distribution kinetics). The approach shows how continuous distributions can describe network dynamics and how either exponential or power law networks can be constructed. Our aim has been to introduce a generalized model for growing networks.

For the distributions in population balance kinetics, we have proposed a growing network model with and without preferential attachment, i.e., power law and exponential networks, respectively. The model with size-independent rate coefficients, $k_g(\xi) = \gamma$ and $k_d(\xi) = \kappa$, yields Gaussian or Poisson distributed exponential networks. The model with the size-dependent rate coefficients, $k_g(\xi) = \gamma\xi^{\lambda}$ and $k_d(\xi) = \kappa\xi^{\nu}$, produces power law networks, and the nonlinear network size dependence of the number of connections describes how such power law networks evolve. We explored the nonlinear growth of power law degree distribution with time and network size. Our model for power law network evolution has two parameters: τ, which controls node addition, and the key parameter, k, which is the difference between pre-factors of the rate coefficient $(\gamma - \kappa)$ and manages network-growing intensity. For real-world complex systems, k reflects changes of the internal or external conditions of networks, such as the accumulation of connection load [41,42], which increases connecting and organization costs, or technological innovations, which diminish the costs.

Moment results show the time dependence of the distribution, $p(\xi, t)$, for either linearly or nonlinearly growing networks. Power law expressions with unlimited ξ do not have proper integrals, and thus their moments are indeterminate. But as we have demonstrated, truncated power laws evolve to increasingly larger values of ξ, and thus moment integrals can be defined for finite time. A general moment equation was derived by the integration of the moment definition in equation (2) when integer values of λ and ν are identical, $\lambda = \nu$. The moments, $p^{(n)}(t)$, $p^{avg}(t)$, and $p^{var}(t)$, have complicated expressions that can be derived by computer algebra. We also derived time and network size dependent behaviors of the model using moment solutions, determined based on the initial and boundary conditions.

The growing network model based on distribution dynamics can describe growing network systems and represent data. The degree distributions of the network model for exponential and power law networks increase with time and network size through nonlinear accelerating growth (Figs. 2 and 3). The proposed model, allowing multiple connections between any two nodes, is general compared to the connection-limited networks, which allow only one connection between two nodes. We have demonstrated that the model is also able to describe accelerating nonlinear growth of networks by plotting degree and the total number of connections with network size.

The aim of this paper was to present a model for accelerating network evolution by developing a framework for the dynamics of linear and nonlinear growing networks. We focused on accelerating networks [12] and how the number of nodes and connections and the degree evolve in time. The model in its present state treats only those quantities that can be represented as moments of the distribution, $p(\xi, t)$, and thus not quantities such as node separation or clustering coefficient, which have been computed via Monte Carlo simulations [43]. At present, no single model quantitatively describes all possible phenomena associated with networks. Population balance (nonequilibrium distribution kinetics) modeling can incorporate

additional qualities as added parameters and variables in the distribution function, and this is the subject of continuing investigation.

References

1. J.M. Ottino, AIChE. J. **49**, 292 (2003)
2. S.N. Dorogovtsev, J.F.F. Mendes, Phys. Rev. E **63**, 025101 (2001)
3. R.L. Albert, A.L. Barabasi, Revs. Mod. Phys. **74**, 47 (2002)
4. S.H. Strogatz, Nature **410**, 268 (2001)
5. P. Erdos, A. Renyi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960)
6. B. Bollobas, *Random Graphs* (Academic Press, London, 1985)
7. D.J. Watts, S.H. Strogatz, Nature **393**, 440 (1998)
8. A. Vazquez, e-print: `arXiv:cond-mat/0006132` (2000)
9. A.L. Barabasi, R.L. Albert, Science **286**, 509 (1999)
10. A.L. Barabasi et al., Physica A **311**, 590 (2002)
11. H. Jeong et al., Nature (London) **407**, 651 (2000)
12. J.S. Mattick, M.J. Gagen, Science **307**, 856 (2005)
13. S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002)
14. Y.-P. Jeon, B.J. McCoy, Comput. Chem. Eng. **29**, 2429 (2005)
15. Y.-P. Jeon, B.J. McCoy, Phys. Rev. E **72**, 037104 (2005)
16. S.N. Dorogovtsev, J.F.F. Mendes, Phys. Rev. E **62**, 1842 (2000)
17. B.A. Huberman, L.A. Adamic, Nature **401**, 131 (1999)
18. A.L. Barabasi, R.L. Albert, H. Jeong, Physica A **272**, 173 (1999)
19. S.N. Dorogovtsev, J.F.F. Mendes, Phys. Rev. E **63**, 056125 (2001)
20. P. Baran, *Introduction to Distributed Communications Networks*, RM-3420-PR (1964) (`http://www.rand.org/publications/RM/RM3420/`)
21. B.A. Huberman, P.L.T. Pirolli, J.E. Pitkow, R.J. Lukose, Science **280**, 95 (1998)
22. J. Lahererre, D. Sornette, Eur. Phys. J. B **2**, 525 (1998)
23. S. Render, Eur. Phys. J. B **4**, 131 (1998)
24. L.A.N. Amaral, A. Scala, M. Barthelemy, H.E. Stanley, Proc. Natl. Acad. Sci. USA **97**, 11149 (2000)
25. M.E.J. Newman, Proc. Natl. Acad. Sci. USA **98**, 404 (2001)
26. G. Madras, B.J. McCoy, J. Coll. Interf. Sci. **246**, 356 (2002)
27. B.J. McCoy, G. Madras, Chem. Eng. Sci. **56**, 2831 (2001)
28. G. Madras, B.J. McCoy, J. Chem. Phys. **117**, 8042 (2002)
29. N.G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, 1992)
30. P.L. Krapivsky, S. Redner, F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000)
31. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd edn. (Wiley, New York, 1957)
32. S. Chandrasekhar, Revs. Mod. Phys. **15**, 1 (1943)
33. D.R. Cox, M.D. Miller, *The Theory of Stochastic Processes* (Wiley, New York, 1965)
34. O. Levenspiel, W.K. Smith, Chem. Eng. Sci. **6**, 227 (1957)
35. N.A. Dotson, R. Galvan, R.L. Laurence, M. Tirrell, *Polymerization Process Modeling* (New York, VCH Publishers Inc, 1996)
36. A.D. Randolph, M.A. Larson, *Theory of Particulate Processes*, 2nd edn. (Academic Press, New York, 1986)
37. T. Sato, Nucl. Phys. B (Proc. Suppl.) **129&130**, 102 (2004)
38. M.J. Gagen, J.S. Mattick, Theory in Bioscience **123**, 381 (2005)
39. L.J. Croft, M.J. Lercher, M.J. Gagen, J.S. Mattick et al., Genome Biol. **5**, P2 (2003)
40. R. Pastor-Satorras, A. Vazquez, A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001)
41. P. Holme, B.J Kim, Phys. Rev. E **65**, 066109 (2002)
42. Y. Moreno, R. Pastor-Satorras, A. Vazquez, A. Vespignani, Europhys. Lett. **62**, 292 (2003)
43. S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks* (Oxford, New York, 2003)