PHYSICAL REVIEW E **77**, 045102(R) (2008)

Analytical results for stochastically growing networks: Connection to the zero-range process

P. K. Mohanty[1,*] and Sarika Jalan[2,†]

[1]*TCMP Division, Saha Institute of Nuclear Physics, 1/AF, Saltlake, Kolkata 700064, India*

[2]*Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Strasse 38, 01187 Dresden, Germany*

(Received 14 August 2007; revised manuscript received 8 February 2008; published 22 April 2008)

We introduce a stochastic model of growing networks where both the number of new nodes which join the network and the number of connections vary stochastically. We provide an exact mapping between this model and the zero-range process, and calculate analytically the degree distribution for any given evolution rule. We argue that this mapping can be used to infer a possible evolution rule for any given network. This is being demonstrated for a protein-protein interaction network of *Saccharomyces cerevisiae*.

DOI: 10.1103/PhysRevE.77.045102 PACS number(s): 89.75.Hc, 05.40.−a, 04.20.Jb, 89.20.−a

The study of networks has been gaining recognition as a fundamental tool in understanding the dynamical behavior and response of real systems. This recognition is arising from different fields such as biology, social systems, technological systems, etc. [1–5]. Different network models have been proposed to study and understand these systems, which have underlying network structure. The Erdös and Rényi random networks model is one of the oldest, and shows that the probability $[p(k)]$ of a node having degree k follows exponential distributions, $p(k) \propto \exp(-k)$ [6]. Many real world networks, however, show scalefree behavior, $p(k) \propto k^{-\gamma}$, with a few very striking examples of the World Wide Web (WWW) and cellular networks [7,8] (for a review of scalefree networks see [2]). In the WWW, the number of incoming links follows a power law with a value of $\gamma \sim 2.1$ [7], and analysis of metabolic networks of 43 organisms reveals that the number of chemical reactions (link) in which a substrate (node) is involved shows power law distribution, with the exponent varying between 2.0 and 2.4 [8].

To capture scalefree behavior of real world networks, Barabási-Albert (BA) proposed a growing network model based on the preferential attachment of the nodes [2,9]. In the BA model, each new node is connected with some old nodes with a probability *linearly* proportional to the degree of the node, $u(k) \propto (k+\beta)$. This model gives rise to the scalefree network with degree distribution following power law $p(k) \propto k^{-\gamma}$, a value of $\gamma = 3+\beta$ [10,11]. Since then, several variations of the BA algorithm have been proposed. An algorithm suggested by Dorgovtsev and Mendes based on the aging of the nodes also gives rise to scalefree behavior [12]. Krapivsky *et al.* also obtained an analytical solution for a different attachment function $u(k) \sim k^{\lambda}$ [10].

In this paper we introduce stochasticity to the growing network models. Starting with the few initially connected nodes, a network in our model evolves as follows. At each time step, n new nodes join the network and make m connections with existing nodes. Both m and n are taken as stochastic variables. Each new connection is made with a probability which depends on the degree of the node to be connected, and needs not be preferential. A special case of our model with linear connection probability and $n=1$ corresponds to the BA algorithm. Note that our evolution rule, being stochastic, naturally captures various stochastic effects which are always present during the evolution of any real system.

First we show an explicit mapping between our model and the zero-range process (ZRP), an exactly solvable model in nonequilibrium physics [13], which provides an exact relation between any attachment rule $u(k)$ and degree distribution $p(k)$ of the growing networks. So far, there have been partial attempts to solve the Barabási-Albert model. Dorgovtsev *et al.* have done analytical calculations for linear $u(k)$ [11] and for certain other forms [14]. Also, the exact result is known for $u(k) \sim k^{\lambda}$ [10]. Here, we provide exact degree distribution for any arbitrary evolution rule $u(k)$. This relation, being exact, can be inverted to infer a possible evolution rule for any given real-world network. Second, we show that the choice of stochastic parameters does not alter the degree distribution of the network. It only affects the correlations or statistical properties of the modules. Last, we apply our methodology to a real world network and derive a stochastic evolution rule which captures the exact degree distribution. We argue that this method can be used to generate a growing network with any desired degree distribution.

A generic algorithm for a growing network would be as follows. Starting from a small connected network, one brings n new nodes at each iteration time t and then each of these n nodes connects to m existing nodes. In general, n and m are stochastically varying integers in $(1, n_{max})$ and $(1, m_{max})$, respectively, and their distributions are $\eta(n)$ and $h(m)$. These variations are not just the generalizations of [9]; it is quite natural that a variable number of nodes joins realistic networks and make connections which vary from one node to the other. The probability that any given new node i makes a link with one of the existing nodes j is $w[k(j),t]$, where $k(j)$ is the degree of j and $\Sigma_j w[k(j),t]=1$.

Our first task would be to find the steady state degree distribution $p(k)$ of these generic networks as $t \to \infty$. Let $M(k,t)$ be the number of nodes having k links at time t. Since $\Sigma_j w(k(j),t)=1$, we may take $w(k,t)=u(k)/v(t)$ where

$$v(t) = \sum_k u(k) M(k,t). \tag{1}$$

Here $u(k)$ is considered to be a generic function, and need not be an increasing function corresponding to the preferen-

*pk.mohanty@saha.ac.in

†sarika@pks.mpg.de

tial attachment discussed in [9,10]. The size of the network, and thus $M(k,t)$, increases with iteration time t. The growth rate of $M(k,t)$ is

$$\frac{dM(k,t)}{dt} = \bar{m}\bar{n}\left[\frac{u(k-1)}{v(t)}M(k-1,t) - \frac{u(k)}{v(t)}M(k,t)\right] + \bar{n}h(k), \tag{2}$$

where $\bar{n}=\sum_{n=1}^{n_{max}} n\,\eta(n)$ is the average number of nodes which join the network in each iteration step t. Equation (2) is constrained by $M(0,t)=0$, which ensures that every node in the network has nonzero links. The initial condition is $M(k,0)=m_{max}\delta_{k,1}$, i.e., we start with a network of size m_{max} to avoid multiple connections. Note that (2) must be supplemented by the growth rate of nodes,

$$\frac{dN(t)}{dt} = \bar{n}. \tag{3}$$

In general, $\bar{n}$ depends on t (particularly when η explicitly depends on t). First let us take $\bar{n}$ to be independent of t. The degree distribution $p(k)$ in the steady state is defined as

$$p(k) = \lim_{t\to\infty}\left\langle \frac{M(k,t)}{N(t)} \right\rangle, \tag{4}$$

where averaging $\langle\cdots\rangle$ is done over realizations. Clearly the steady state is reached only if $M(k,t)\propto N(t)$ as $t\to\infty$. Thus, in the steady state we must have

$$M(k,t) = p(k)N(t). \tag{5}$$

One can safely assume that this product form (5) holds even for $t<\infty$ (evidence in favor of this *ansatz* is provided later in this article), which leads Eq. (2) to

$$\frac{1}{\bar{m}}\frac{v(t)}{N(t)} = \frac{u(k-1)p(k-1) - u(k)p(k)}{p(k) - h(k)}. \tag{6}$$

The left- (right-) hand side of the above equation is independent of $k(t)$, which can be satisfied only by a constant. Taking this constant to be α,

$$p(k) = \frac{u(k-1)}{\alpha + u(k)}p(k-1) + \frac{\alpha h(k)}{\alpha + u(k)}, \tag{7}$$

$$\alpha = \frac{1}{\bar{m}}\frac{v(t)}{N(t)} = \frac{1}{\bar{m}}\sum_k u(k)p(k). \tag{8}$$

Before proceeding to solve the difference Eq. (7) let us note a few things. First, $p(k)$ is in fact normalized, which can be proved by summing Eq. (7) for all k. Second, $\bar{n}$ do not appear in these equations, implying that $\bar{n}$ can be fixed to any arbitrary value without changing the degree distribution $p(k)$. However, since $\bar{n}$ governs the evolution dynamics of the network it is expected to affect the correlations [even though the steady state $p(k)$ is the same].

Equation (7) can be solved to obtain

$$p(k) = \frac{\alpha}{u(k)}\sum_{m=1}^{m_{max}} h(m)\prod_{j=m}^{k}\frac{u(j)}{\alpha + u(j)}. \tag{9}$$

Here, we have used the natural boundary condition $p(0)=0$. Although (9) provides an exact expression of $p(k)$, the main difficulty still remains in finding α, which has to be self-consistently determined using (7) and (8).

Now, let us apply this exact result (9) to the most well-studied growing networks, where only one node having m_0 links joins the existing network at each time step, i.e., $\bar{n}=1$ and $h(m)=\delta_{m,m_0}$. Thus, only a single term $m=m_0$ in Eq. (9) survives under the sum, and we have $p(k)=0$ for $k<m_0$. For $k\geq m_0$,

$$p(k) = \frac{\alpha}{u(k)}\prod_{j=m_0}^{k}\frac{u(j)}{\alpha + u(j)}. \tag{10}$$

If we use the BA algorithm with preferential attachment rule $u(k)=k+\beta$, the degree distribution becomes

$$p(k) = \alpha\frac{\Gamma(\alpha+\beta+m_0)}{\Gamma(1+\alpha+\beta+k)}\frac{\Gamma(\beta+k)}{\Gamma(\beta+m_0)}, \tag{11}$$

which can be used further to obtain $\alpha=2+\beta/m_0$ from (8). Clearly, for the large values of k, $p(k)\sim k^{-1-\alpha}$ giving rise to $\gamma=3+\beta/m_0$. In the original formulation of Barabási-Albert [9] the attachment rule is $u(k)=k$ (i.e., $\beta=0$) which, thus, generates a scalefree network with $\gamma=3$.

In the following we discuss the mapping of our growing network model with the ZRP. First we consider the model without any stochasticity. In the ZRP, particles hop between the sites of a lattice with rate $w(k)$ where k is the occupancy of the departure site. The steady-state distribution of particles $\pi(k)$ in the ZRP can be calculated exactly as $\pi(k)=\mathcal{N}\Pi_{j=1}^{k}w(k)^{-1}$, where $\mathcal{N}$ is a normalization constant. Comparing the steady state distribution of the ZRP with (10), one can identify that $\pi(k)=p(k)u(k)$ with transfer rate of particles in the ZRP, satisfying

$$w(k) = \begin{cases} 1+\dfrac{\alpha}{u(k)} & \text{for } k\geq m_0 \\ 1 & \text{for } k<m_0. \end{cases} \tag{12}$$

Obviously, Eq. (8) is nothing but the normalization condition for $\pi(k)$. Such a mapping is useful, because, for any growing network with attachment rule $u(k)$, Eq. (12) finds an equivalent particle transfer rate $w(k)$ of the ZRP, and since steady state distribution of particles $\pi(k)$ is known exactly for the ZRP [13] one effectively finds degree distribution $p(k)\sim\pi(k)/u(k)$.

To explain the importance of this mapping, let us take the example where $u(k)=k^{\lambda}$. In this case, the ZRP predicts [13] the following three different possibilities: (a) $0<\lambda<1$: $\pi(k)$ is a stretched exponential and thus $p(k)\sim\exp[-\alpha k^{1-\lambda}/(1-\lambda)]k^{-\lambda}$, (b) $\lambda=1$: in this case one gets $p(k)\sim k^{-(\alpha+1)}$, and (c) $\lambda>1$: here $\pi(k)$ is more complex but asymptotically reaches a constant, resulting in asymptotic degree distribution $p(k)\sim k^{-\lambda}$. All these results agree with the $u(k)=k^{\lambda}$ case considered in [10].

Let us emphasize at this point that although writing a

closed form expression of $p(k)$ for generic $u(k)$ is difficult, asymptotic behavior can be obtained easily using (10) or (14). One simple way is to take the continuum limit, by choosing $x=k/K$, where K is the maximum possible links (an arbitrarily large number). The difference equation (7) then becomes a differential equation

$$-\frac{1}{p(x)}\frac{d}{dx}p(x)u(x)=\alpha,$$

with boundary condition $p(x_0)=\alpha/u(x_0)$, where $x_0=m_0/K$. A formal solution is then

$$p(x)=\frac{\alpha}{u(x_0)}\frac{1}{u(x)}\exp\left(-\alpha\int^{x}\frac{dx'}{u(x')}\right), \tag{13}$$

$$\alpha=\frac{1}{x_0}\int dk\ u(x)p(x). \tag{14}$$

It is easy to check that the above equations provide correct asymptotic values for exactly solvable cases, $u(k)=k+\beta$ and $u(k)=k^{\lambda}$. As far as exact derivation of $p(k)$ is concerned, one may numerically implement (9) and (8), i.e., by iterating (9) and (8), and assuming an initial α. In most cases, we observe that α converges rapidly (within 15 iterations) to a constant.

This mapping to the ZRP holds for the stochastically growing network as well. To establish this point, let us consider a generalized zero range process $\mathcal{Z}_\nu$, where a single particle can be transferred from a site to its neighbor with rate $w(k)$ if and only if the site has $k\geq\nu$ particles. A special case $\mathcal{Z}_1$ is the standard ZRP. In the steady state of $\mathcal{Z}_\nu$, distribution of particles is simply $\pi(k)=\Pi_{j=\nu}^{k}w(k)^{-1}$ up to a normalization constant. If we take an ensemble of ZRP $\{\mathcal{Z}_\nu\}$ with $\nu=1,2\ldots\nu_{max}$, which may be realized by taking several rings each having a different dynamics $\mathcal{Z}_\nu$, clearly the steady state distribution of particles becomes

$$\pi(k)=\mathcal{N}\sum_{\nu=1}^{\nu_{max}}h(\nu)\left[\prod_{j=\nu}^{k}\frac{1}{w(j)}\right], \tag{15}$$

where $h(\nu)$ is the weight of $\mathcal{Z}_\nu$ in the ensemble and $\mathcal{N}$ is the normalization constant. This equation may be compared with (9) for drawing an analogy between the degree distribution of stochastically growing networks and the ZRP. Note that for large $k(\gg\nu_{max})$, $p(k)$ solely depends on the $[\cdots]$ of Eq. (15). Thus, asymptotic degree distribution of a growing network is not affected by the stochasticity as long as m_{max}, the maximum number of links a new node can make, is not too large.

Now let us discuss how one can infer about a possible evolution rule from any given degree distribution. Equation (9) can be inverted to obtain

$$u(k)=\frac{1}{p(k)}\sum_{i=1}^{k}[h(i)-p(i)], \tag{16}$$

which is consistant with Eq. (8). Note that by equating the right-hand side of (8) to α one gets $h(i)-p(i)=u(i-1)p(i-1)-u(i)p(i)$, which gives (16) when summed for $i=1$ to $i=k$. Here, α appears as an (irrelevant) multiplicative constant which can be dropped. Equation (16) provides meaningful insight about the possible evolution rule of any given real world network. For illustration, we take the protein-protein interaction (PPI) network for *Saccharomyces cerevisiae* (yeast) [15]. The largest connected part has $N=3930$ nodes and $M=7725$ links. The degree distribution of this network is shown in Fig. 1. The average degree of this network is 3.93. To model this, we take $h(m)=\Sigma_{i=1}^{3}a_i\delta_{m,i}$, where $\{a_i\}$ are constrained by the conditions (i) normalization: $a_1+a_2+a_3=1$, and (ii) average degree: $3.93=2(a_1+2a_2+3a_3)$. We choose to work with $a_1=0.4$, which gives $a_2=0.234$ and $a_3=0.366$, and evaluate $u(k)$ for this network (inset of Fig. 1) using (16) which fits well with a linear function $u(k)=1.5(k-0.8)$. The corresponding degree distribution turns out to be $p(k)\sim k^{-2.2}$, which is a scalefree distribution as expected from a linear $u(k)$.

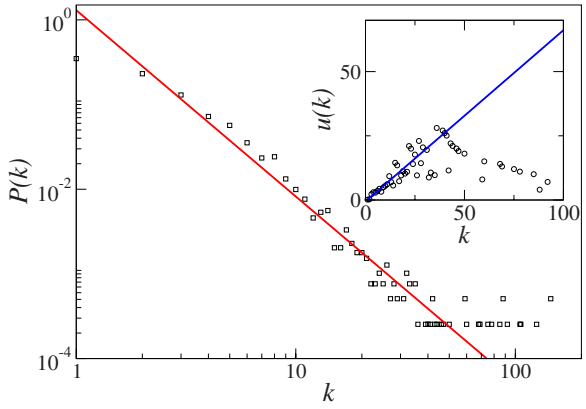

FIG. 1. (Color online) Degree distribution for the PPI network for *Saccharomyces cerevisiae* [15]. The evolution rule $u(k)$ derived using (16) is shown in the inset. The solid line here (inset) is a linear fit $u(k)=k-0.8$, for which one expects $p(k)\sim k^{-2.2}$. A solid line with slope −2.2 is drawn in the main figure to compare $p(k)$ with the theory.

Finally, we turn our attention to the other stochastic parameter $\eta(n)$, namely, the distribution of number of nodes joining the network during each iteration time step t. We have seen in (9) that $\eta(n)$ does not alter the degree distribution. Being in the dynamics, however, it affects the correlations or the statistical properties of modular structures in the network. To illustrate this point, we generate a network with $u(k)=k+0.5$, $h(m)=\delta_{m,4}$ and $\eta(n)=q\delta_{n,1}+(1-q)\delta_{n,5}$, and measure the clustering coefficient for different q. As explained in Fig. 2, we find that the clustering coefficient changes only marginally with q.

Our analysis here relies on the fact that Eq. (5) holds for large networks (as $t\to\infty$). Let us check the validity of (5) in detail. From (8) it is clear that $v(t)$ is proportional to $N(t)$ which can be obtained from (3). First, we numerically evaluate $v(t)$ for a few different networks (starting from two nodes which are connected) and compare them with the theoretical results (3). If the number of new nodes n is a stochastic variable, then $N(t)=\bar{n}t+2$ is linear. However, one can introduce an explicit time dependence in n to get nonlinear $N(t)$. For example, if $n(t)=\sqrt{t}$, we have $N(t)=t^{3/2}+2$ and thus $v(t)\propto t^{3/2}$. In Fig. 3 we plot numerically measured $v(t)$ in log scale for two different cases: (a) $n=0.6\delta_{n,1}+0.4\delta_{n,2}$ and (b)

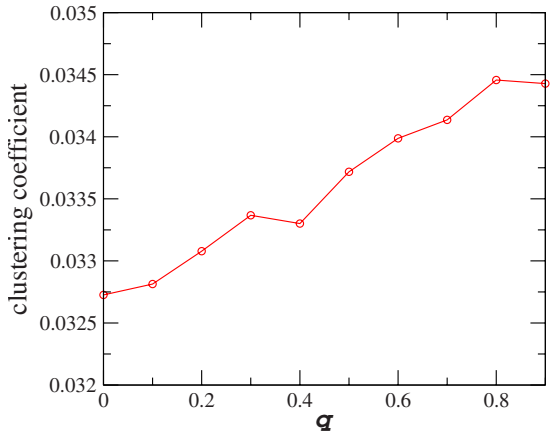

FIG. 2. (Color online) Clustering coefficient changes with a stochastic parameter q (see text). Other parameters are $u(k)=k+0.5$, $N=1000$, $m=4$, $\eta(n)=q\delta_{n,1}+(1-q)\delta_{n,5}$, and the clustering coefficient is averaged over 1000 realizations.

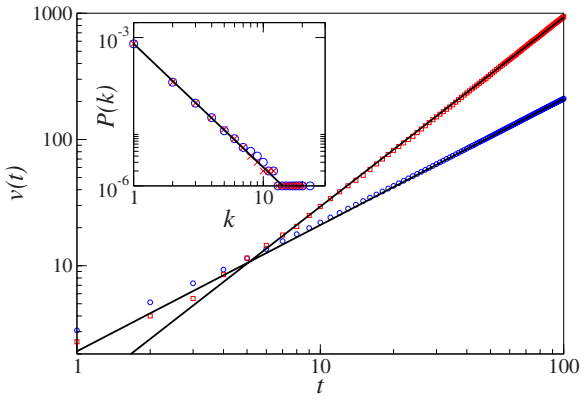

FIG. 3. (Color online) Log scale plot of $v(t)$ for two different cases: (a) $\eta(n)=0.6\delta_{1,n}+0.4\delta_{2,n}$, and (b) $n(t)=\sqrt{t}$. It is expected from (3) and (8) that $v(t)$ is linear in the first case, whereas for (b) $v(t)\sim t^{3/2}$. Solid lines with slope 1 and 1.5 are drawn for comparison. For both cases, $u(k)=k-0.5$ and $m=1$, and averaging is done over 1000 realizations. The degree distribution $p(k)\sim k^{-2.5}$ (inset) is identical for both cases.

$n(t)=\sqrt{t}$, both agree well with (3). Although $N(t)$ is quite different, $p(k)$ (shown in the inset) was found to be same as expected. For both cases the evolution rule is $u(k)=k-0.5$ and thus we have $p(k)\sim k^{-2.5}$. To conclude, Eq. (5) holds quite well, except for first few iteration steps $t<10$. For large networks, however, the number of nodes which join in the first few iteration steps is vanishingly small as compared to the size of the network, and hence does not affect the network properties as a whole.

In summary, we introduce a generic model of a stochastically growing network and show that this model can easily be mapped to the ZRP, thus enabling us to derive an exact relation between the degree distribution of the network and its evolution function. This relation can be used to derive the analytical form of the degree distribution for any arbitrary evolution rule. Conversely, for a given network data we can infer a possible evolution rule. Our evolution rule produces exact degree distribution, as obtained from the given network data, even for small k values. We demonstrate this by taking an example of a real world PPI network and deriving a possible evolution rule to this network.

Based on our exact calculations we expect to obtain a better understanding of the evolution of real world networks. Also, since the ZRP is exactly solvable, mapping of the ZRP with network growth models opens up a platform to study the interplay between evolution rules and steady state degree distribution.

One of us (P.K.M.) acknowledges MPIPKS for hospitality and support.

[1] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002), and references therein.
[3] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002); A. Clauset, M. E. J. Newman, and C. Moore, Phys. Rev. E **70**, 066111 (2004); M. E. J. Newman, Soc. Networks **27**, 39 (2005); Proc. Natl. Acad. Sci. U.S.A. **103**, 8577 (2006).
[4] R. Guimerá and L. A. N. Amaral, Nature (London) **433**, 895 (2005).
[5] L. da. F. Costa, F. A. Rodrigues, G. Travieso, and P. Villas Boas, Adv. Phys. **56**, 167 (2007).
[6] P. Erdös and A. Rényi, Publ. Math. Inst. Hungar. Acad. Sci. **5**, 17 (1960).
[7] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).
[8] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, Nature (London) **407**, 651 (2000).
[9] A.-L. Barabási and R. Albert, Science **286**, 509 (1999); A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).
[10] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[11] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[12] S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **62**, 1842 (2000).
[13] For a review on ZRP, see M. R. Evans and T. Hanney, J. Phys. A **38**, R195 (2005).
[14] S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **63**, 056125 (2001).
[15] Data published by COSIN Project Group, http://www.cosin.org/extra/data/proteins/