Degree Distributions of Growing Networks

P. L. Krapivsky[1], G. J. Rodgers[2], and S. Redner[1]
[1] *Center for BioDynamics, Center for Polymer Studies, and Department of Physics, Boston University, Boston, MA 02215*
[2] *Department of Mathematical Sciences, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK*

The *in-degree* and *out-degree* distributions of a growing network model are determined. The in-degree is the number of incoming links to a given node (and *vice versa* for out-degree). The network is built by (i) creation of new nodes which each immediately attach to a pre-existing node, and (ii) creation of new links between pre-existing nodes. This process naturally generates correlated in- and out-degree distributions. When the node and link creation rates are linear functions of node degree, these distributions exhibit *distinct* power-law forms. By tuning the parameters in these rates to reasonable values, exponents which agree with those of the web graph are obtained.

PACS numbers: 02.50.Cw, 05.40.-a, 05.50.+q, 87.18.Sn

The world-wide web (WWW) is a rapidly evolving network which now contains nearly 10^9 nodes. Much recent effort has been devoted to characterizing the underlying directed graph formed by these nodes and their connecting hyperlinks – the so-called "web" graph [1–4]. In parallel with these developments, a variety of growing network models have recently been introduced and studied [5–14]. These model networks are built by sequentially adding both nodes and links in a manner which mimics the evolution of real network systems, with the WWW being the most obvious example.

One fundamental characteristic of any graph is the number of links at a node – the node degree. The growing network models cited above predict that the distribution of node degree has a power law form for growth rules in which the probability that a newly-created node attaches to a pre-existing node increases linearly with the degree of the "target" node [5–9]. This power law behavior strongly contrasts with the Poisson degree distribution of the classical random graph [15], where links are randomly created between any pair of pre-existing nodes in the network.

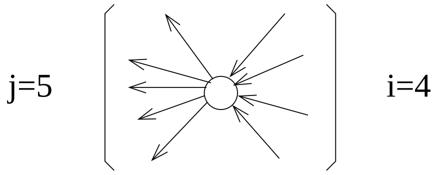

FIG. 1. A node with in-degree $i = 4$, out-degree $j = 5$, and total degree 9.

Since links in the web are directed, the *total* degree of a node may naturally be resolved into the *in-degree* – the number of incoming links to a node, and the *out-degree* – the number of outgoing links from a node (Fig. 1). While the total node degree and its distribution are now reasonably understood [6–9,12], much less is known about the joint distribution of in-degrees and out-degrees, as well as their correlation. Empirical measurements of the web graph indicate that the in- and out-degree distributions are both non-trivial and exhibit power-law behaviors with different exponents [2–4]. In this Letter, we solve for this joint distribution in a growing network model and obtain results which reproduce the degree distributions of the web graph.

For the model that we study, the network growth occurs by two distinct processes (Fig. 2):

(i) With probability p, a new node is introduced and it immediately attaches to one of the earlier target nodes in the network. The attachment probability depends only on the in-degree of the target.

(ii) With probability $q = 1 - p$, a new link is created between already existing nodes. The choices of the originating and target nodes depend on the out-degree of the originating node and the in-degree of the target node.

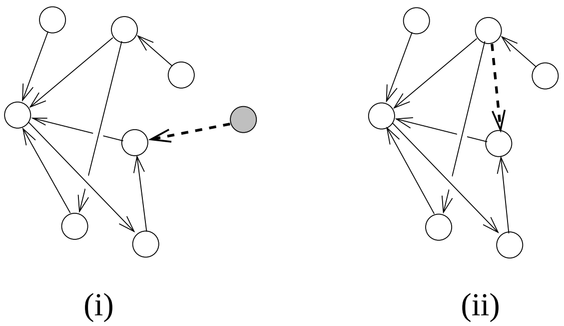

FIG. 2. Illustration of the growth processes in the growing network model: (i) node creation and immediate attachment, and (ii) link creation. In (i) the new node is shaded, while in both (i) and (ii) the new link is dashed.

If only process (i) was allowed, the out-degree of each node would be one by construction. The inclusion of process (ii) represents a simple extension of Simon's original model [5] and the growing network model of Barabási and Albert [6]. The relative rate of processes (i) and (ii) has been shown to drive a transition in the network structure [14]. We shall further show that the general model gives a non-trivial out-degree distribution which is distinct from the in-degree distribution.

We begin our analysis by determining the average node degree; this can be found in a direct manner and does not require the specification of the attachment and link creation probabilities. Let $N(t)$ be the total number of nodes in the network, and let $I(t)$ and $J(t)$ be the total in-degree and out-degree, respectively. According to the two basic growth processes enumerated above, at each time step these degrees evolve according to one of the following two possibilities

$$(N, I, J) \to \begin{cases} (N+1, I+1, J+1) & \text{probability } p, \\ (N, I+1, J+1) & \text{probability } q. \end{cases} \tag{1}$$

That is, with probability p a new node and new directed link are created (Fig. 2) so that the number of nodes and both node degrees increase by one. Conversely, with probability q a new directed link is created and the node degrees each increase by one, while the total number of nodes is unchanged. As a result,

$$N(t) = pt, \qquad I(t) = J(t) = t, \tag{2}$$

from which we immediately conclude that the average in- and out-degrees, $\mathcal{D}_{\rm in} \equiv I(t)/N(t)$ and $\mathcal{D}_{\rm out} \equiv J(t)/N(t)$, are both time independent and equal to $1/p$.

To determine the joint degree distributions, we need to specify: (i) the *attachment rate* $A(i,j)$, defined as the probability that a newly-introduced node links to an existing node with i incoming and j outgoing links, and (ii) the *creation rate* $C(i_1, j_1|i_2, j_2)$, defined as the probability of adding a new link from a (i_1, j_1) node to a (i_2, j_2) node. We restrict the form of these rates to those which we naturally expect to occur in systems such as the web graph. First, we assume that the attachment rate depends only on the in-degree of the target node, $A(i,j) = A_i$. We also assume that the link creation rate depends only on the out-degree of the node from which it emanates and the in-degree of the target node, that is, $C(i_1, j_1|i_2, j_2) = C(j_1, i_2)$.

On general grounds, the attachment rate A_i and the creation rate $C(j,i)$ should be increasing functions of i. Similarly, $C(j,i)$ should be an increasing function of j. For example the designer of a new web page is more likely to construct hyperlinks to well-known pages rather than to obscure pages. Similarly, a web page with many outgoing hyperlinks is more likely to create even more hyperlinks. We have found that the degree distributions exhibit qualitatively different behaviors depending on whether the asymptotic dependence of the rates A_i and $C(j,i)$ on both i and j grow slower than linearly, linearly, or faster than linearly. The first and last cases lead to either rapidly decaying degree distributions or to the dominance of a single node; this same behavior was already found for the total node degree [8,12]. The most interesting behavior arises for asymptotically linear rates, and we focus on this class of models in our investigations.

Specifically, we consider a growing network model with the attachment and creation rates which are shifted linear functions in all indices (*linear-bilinear* rates)

$$A_i = i + \lambda, \qquad C(j,i) = (i+\lambda)(j+\mu). \tag{3}$$

An intuitively natural feature of this model that both the attachment and creation rates have the same dependence on the popularity of the target node. The parameters λ and μ in the rates of Eq. (3) must obey the constraints $\lambda > 0$ and $\mu > -1$ to ensure that the corresponding rates are positive for all permissible values of in- and out-degrees, $i \geq 0$ and $j \geq 1$.

As the network grows, the joint degree distribution, $N_{ij}(t)$, defined as the average number of nodes with i incoming and j outgoing links, builds up. To solve for $N_{ij}(t)$, we shall use the rate equation approach, which has recently been applied to simpler versions of growing networks [8,9,12]. When the attachment and creation rates are given by Eq. (3), the degree distribution $N_{ij}(t)$ evolves according to the rate equations

$$\begin{aligned} \frac{dN_{ij}}{dt} &= (p+q)\left[\frac{(i-1+\lambda)N_{i-1,j} - (i+\lambda)N_{ij}}{I + \lambda N}\right] \\ &\quad + q\left[\frac{(j-1+\mu)N_{i,j-1} - (j+\mu)N_{ij}}{J+\mu N}\right] + p\,\delta_{i0}\delta_{j1}. \end{aligned} \tag{4}$$

The first group of terms on the right-hand side account for the changes in the in-degree of target nodes. These changes arise by simultaneous creation of a new node and link (with probability p) or by creation of a new link only (with probability q). For example, the creation of a link to a node with in-degree i leads to a loss in the number of such nodes. This occurs with rate $(p+q)(i+\lambda)N_{ij}$, divided by the appropriate normalization factor $\sum_{i,j}(i+\lambda)N_{ij} = I + \lambda N$. The factor $p+q=1$ in Eq. (4) has been written to make explicit the two type of relevant processes. Similarly, the terms in the second group of terms accounts for changes in the out-degree. These occur due to the creation of new links between already existing nodes – hence the prefactor q. The last term accounts for the continuous introduction of new nodes with no incoming links and one outgoing link. As a useful self-consistency check, we can easily verify that the total number of nodes, $N = \sum_{i,j} N_{ij}$, obeys $\dot{N} = p$, in agreement with Eq. (2). In the same spirit, the total in- and out-degrees, $I = \sum_{i,j} iN_{ij}$ and $J = \sum_{i,j} jN_{ij}$, obey $\dot{I} = \dot{J} = 1$.

By solving the first few of Eqs. (4), it is clear that the N_{ij} grow linearly with time. Accordingly, we substitute $N_{ij}(t) = t\,n_{ij}$, as well as $N = pt$ and $I = J = t$, into Eqs. (4) to yield a recursion relation for n_{ij}. Using the shorthand notations,

$$a = q\,\frac{1+p\lambda}{1+p\mu} \quad \text{and} \quad b = 1 + (1+p)\lambda,$$

the recursion relation for n_{ij} simplifies to

$$\begin{aligned} [i + a(j+\mu) + b]n_{ij} &= (i-1+\lambda)n_{i-1,j} \\ &\quad + a(j-1+\mu)n_{i,j-1} \\ &\quad + p(1+p\lambda)\delta_{i0}\delta_{j1}. \end{aligned} \tag{5}$$

We first consider the in-degree and out-degree distributions, $N_i^{\rm in}(t)=\sum_j N_{ij}(t)$ and $N_j^{\rm out}(t)=\sum_i N_{ij}(t)$. Because of the linear time dependence of the nodes degrees, we may write $N_i^{\rm in}(t)=tf_i$ and $N_j^{\rm out}(t)=tg_j$, where the densities f_i and g_j satisfy

$$(i+b)f_i=(i-1+\lambda)f_{i-1}+p(1+p\lambda)\delta_{i0}, \quad (6)$$

$$\left(j+\frac{1}{q}+\frac{\mu}{q}\right)g_j=(j-1+\mu)g_{j-1}+p\frac{1+p\mu}{q}\delta_{j1}, \quad (7)$$

respectively. The solution to these recursion formulae may be immediately written down in terms of the ratio of the Euler gamma functions

$$f_i=F\,\frac{\Gamma(i+\lambda)}{\Gamma(i+b+1)}, \quad (8)$$

$$g_j=G\,\frac{\Gamma(j+\mu)}{\Gamma(j+1+q^{-1}+\mu q^{-1})}. \quad (9)$$

The prefactors F and G are found by matching these forms for f_i and g_j with $f_0=p(1+p\lambda)/b$ and $g_1=p(1+p\mu)/(1+q+\mu)$, as obtained directly from Eqs. (6) and (7).

From the asymptotics of the gamma function, the asymptotic behavior of the in- and out-degree distributions have the power law forms,

$$f_i\sim i^{-\nu_{\rm in}}, \qquad g_j\sim j^{-\nu_{\rm out}}, \quad (10)$$

with the following exponents

$$\nu_{\rm in}=2+p\lambda, \qquad \nu_{\rm out}=1+q^{-1}+\mu p q^{-1}. \quad (11)$$

These exact formulae for the exponents of the in- and out-degree distributions constitute one of our primary results. Interestingly, the in-degree exponent depends on λ (an in-degree feature) while the out-degree exponent depends on μ (an out-degree feature). Notice also that both $\nu_{\rm in}$ and $\nu_{\rm out}$ are greater than 2.

The analytical form of the degree distributions greatly simplifies in the region of the parameter space where $\nu_{\rm in}=\nu_{\rm out}$. In this region, $a=1$ and $\mu+b=2\lambda$. Therefore, the recursion relation in Eq. (5) reduces to

$$\begin{aligned}(i+j+2\lambda)n_{ij}=&(i-1+\lambda)n_{i-1,j}+\\ &+(j-1+\mu)n_{i,j-1}\\ &+p(1+p\lambda)\delta_{i0}\delta_{j1}.\end{aligned} \quad (12)$$

This is much simpler than the general recursion of Eq. (5), since the degrees i and j appear now with equal prefactors. This feature allows us to transform Eq. (12) into a constant-coefficient recursion relation. Indeed, the substitution

$$n_{ij}=\frac{\Gamma(i+\lambda)\,\Gamma(j+\mu)}{\Gamma(i+j+2\lambda+1)}\,m_{ij} \quad (13)$$

reduces (12) to

$$m_{ij}=m_{i-1,j}+m_{i,j-1}+\gamma\,\delta_{i0}\delta_{j1}, \quad (14)$$

with

$$\gamma=p(1+p\lambda)\,\frac{\Gamma(1+2\lambda)}{\Gamma(\lambda)\,\Gamma(\mu+1)}. \quad (15)$$

We may solve Eq. (14) easily by the generating function technique. Multiplying Eq. (14) by x^iy^j and summing over all $i\geq 0, j\geq 1$, we find the generating function

$$\mathcal{M}(x,y)\equiv\sum_{i=0}^{\infty}\sum_{j=1}^{\infty}m_{ij}x^iy^j=\frac{\gamma y}{1-x-y}. \quad (16)$$

Expanding this latter expression we then obtain

$$m_{ij}=\gamma\,\frac{\Gamma(i+j)}{\Gamma(i+1)\,\Gamma(j)}. \quad (17)$$

Combining Eqs. (13) and (17) gives the joint in- and out-degree distribution

$$n_{ij}=\gamma\,\frac{\Gamma(i+\lambda)\,\Gamma(j+\mu)\,\Gamma(i+j)}{\Gamma(i+1)\,\Gamma(j)\,\Gamma(i+j+2\lambda+1)}. \quad (18)$$

While the general form is unwieldy, we can easily understand the significance of this result by considering simplifying limiting cases. In the limit where both $i\to\infty$ and $j\to\infty$, we find

$$n_{ij}=\gamma\,\frac{i^{\lambda-1}j^{\mu}}{(i+j)^{2\lambda+1}}. \quad (19)$$

The limiting behaviors of n_{ij} when both i and j are large and their ratio is very different from unity are

$$n_{ij}\to\gamma\times\begin{cases} i^{-\lambda-2}j^{\mu} & 1\ll j\ll i,\\ i^{\lambda-1}j^{\mu-2\lambda-1} & 1\ll i\ll j.\end{cases} \quad (20)$$

Notice the basic feature that the in-degrees and out-degrees of a node are correlated. In contrast, if these degrees were uncorrelated, we would instead have $n_{ij}=f_ig_j\sim(ij)^{-\nu_{\rm in}}$.

Another interesting feature of the degree distribution becomes evident if we fix the in-degree i and vary the out-degree j. We then find that n_{ij} reaches a maximum value when $j=i\mu/[2+(1+p)\lambda]$ (here we consider large i and assume that $\mu>0$). Correspondingly, the average out-degree always scales linearly with the in-degree, $\langle j\rangle=i(\mu+1)/[(1+p)\lambda]$ (here the coefficient is always positive). Thus popular nodes – those with large in-degree – also tend to have large out-degrees. A dual property also holds: Nodes with large out-degree – those where many links originate – also tend to be popular.

Let us now compare our theoretical results with the empirical observations for the world-wide web. The relevant results for the node degrees are [4]

$$\nu_{\rm in}\approx 2.1, \quad \nu_{\rm out}\approx 2.7, \quad \mathcal{D}_{\rm in}=\mathcal{D}_{\rm out}\approx 7.5, \quad (21)$$

Setting the observed value $\mathcal{D}_{\text{in}} = \mathcal{D}_{\text{out}} = 7.5$ to p^{-1} (see the discussion following Eq. (2)) we see that the predictions of Eq. (11) matches the observed value of the in- and out-degree exponents when $\lambda = 0.75$ and $\mu = 3.55$, respectively.

We have also investigated an even simpler version of this growing network model with node creation rate $A_i = i + \lambda$, as above, but with link creation rate $C(j, i) = j + \mu$, which does not depend on the popularity of the target node i (*linear-linear* rates). For this model, the rate equations for the evolution of the number of nodes with degrees (i, j) have a similar structure to Eqs. (4) and they can be solved by the same as those applied to the network with linear-bilinear growth rates. We find that the in- and out-degree distributions again have power-law forms. Moreover, the out-degree exponent is still given by Eq. (11), while the value of the in-degree exponent is now $\nu_{\text{in}} = 1 + \lambda + p^{-1}$. If we set $p^{-1} = 7.5$ to reproduce the correct average degree of the web graph, we see that ν_{in} must be larger than 8.5. Therefore this linear-linear rate model cannot match the empirical observations from the web. Thus while the solutions for the degree distributions for linear-bilinear and linear-linear growing network models are formally similar, only the former appears to be a viable candidate as an appropriate description for the link structure of the web graph.

Parenthetically, we can also solve the general growing network with both constant node creation rate and constant link creation rate, $A_i = 1$ and $C(j, i) = 1$. While not necessarily a realistic model, it provides a useful exactly solvable case. By following the basic steps of the rate equation approach we find, for the (normalized) joint degree distribution

$$n_{ij} = \frac{p^2 q^{j-1}}{2^{i+j}} \, \frac{\Gamma(i+j)}{\Gamma(i+1)\Gamma(j)}, \tag{22}$$

from which we deduce that the (normalized) in- and out-degree distributions are $f_i = p^2/(1+p)^{i+1}$ and $g_j = p^2 q^{j-1}$.

In summary, we have studied a growing network model which is built by the sequential events of either: (i) node creation which immediately attaches to a pre-existing node, or (ii) link creation between pre-existing nodes. The combination of these two processes naturally leads to non-trivial in-degree and out-degree distributions. Many of the structural properties of the resulting network have been computed exactly by solving the rate equations for the evolution of the number of nodes with given in- and out-degree. In particular, for link attachment rate linear in the degree of the target node and link creation rate which is linear in the degrees of both the two adjacent nodes, we have shown that power-law in- and out-degree distributions are dynamically generated. By choosing the parameters of the growth rates in a natural manner these exponents can be brought into accord with recent measurements of the web graph. The model also makes a number of predictions concerning the joint degree distribution. In particular, we have shown that significant correlations between in-degrees and out-degrees spontaneously develop. Qualitatively these predictions agree with everyday experience – the assumption of independence of in- and out-degrees of a node is certainly flawed. Quantitative measurements of correlations in the web graph might greatly help to test the existing models and possibly to construct a more realistic model of the world-wide web.

We are grateful for financial support of this work from NSF grant DMR9978902 and ARO grant DAAD19-99-1-0173 (PLK and SR), and a grant from EPSRC (GJR).

---

[1] B. A. Huberman, P. L. T. Pirolli, J. E. Pitkow, and R. Lukose, Science **280**, 95 (1998); S. M. Maurer and B. A. Huberman, nlin.CD/0003041.

[2] J. Kleinberg, R. Kumar, P. Raphavan, S. Rajagopalan, and A. Tomkins, in: *Proceedings of the International Conference on Combinatorics and Computing*, Lecture Notes in Computer Science, Vol. 1627 (Springer-Verlag, Berlin, 1999).

[3] S. R. Kumar, P. Raphavan, S. Rajagopalan, and A. Tomkins, in: *Proceedings of the 25th Very Large Databases Conference, Edinburgh, Scotland, 1999* (Morgan Kaufman, Orlando, FL, 1999).

[4] A. Broder, R. Kumar, F. Maghoul, P. Raphavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, Computer Networks **33**, 309 (2000).

[5] H. A. Simon, Biometrica **42**, 425 (1955).

[6] A. L. Barabási and R. Albert, Science **286**, 509 (1999).

[7] S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **62**, 1842 (2000); *cond-mat*/0012009.

[8] P. L. Krapivsky, F. Leyvraz, and S. Redner, Phys. Rev. Lett. **85**, 4629 (2000).

[9] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).

[10] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *cond-mat*/0007235.

[11] G. Bianconi and A. L. Barabási, *cond-mat*/0011029.

[12] P. L. Krapivsky and S. Redner, *cond-mat*/0011094.

[13] B. Tadić, *cond-mat*/0011442.

[14] R. Albert and A.-L. Barabási, Phys. Rev. Lett. **85**, 5234 (2000); S. N. Dorogovtsev and J. F. F. Mendes, *cond-mat*/0005050.

[15] B. Bollobás, *Random Graphs* (Academic Press, London, 1985); S. Janson, T. Luczak, and A. Rucinski, *Random Graphs* (Wiley, New York, 2000).