

## Competition and multiscaling in evolving networks

G. BIANCONI<sup>1</sup> and A.-L. BARABÁSI<sup>1,2</sup>

<sup>1</sup> *Department of Physics, University of Notre Dame - Notre Dame, IN 46556, USA*

<sup>2</sup> *Institute for Advanced Studies, Collegium Budapest  
Szentháromság utca 2, H-1014 Budapest, Hungary*

(received 3 November 2000; accepted in final form 2 March 2001)

PACS. 05.65.+b – Self-organized systems.

PACS. 89.75.-k – Complex systems.

PACS. 89.75.Hc – Networks and genealogical trees.

**Abstract.** – The rate at which nodes in a network increase their connectivity depends on their fitness to compete for links. For example, in social networks some individuals acquire more social links than others, or on the www some webpages attract considerably more links than others. We find that this competition for links translates into multiscaling, *i.e.* a fitness-dependent dynamic exponent, allowing fitter nodes to overcome the more connected but less fit ones. Uncovering this fitter-gets-richer phenomenon can help us understand in quantitative terms the evolution of many competitive systems in nature and society.

The complexity of many systems can be attributed to the interwoven web in which their constituents interact with each other. For example, the society is organized in a social web, whose nodes are individuals and links represent various social interactions, or the www forms a complex web whose nodes are documents and links are URLs. While for a long time these networks have been modeled as completely random [1,2], recently there is increasing evidence that they in fact have a number of generic non-random characteristics, obeying various scaling laws or displaying short length-scale clustering [3–16].

A generic property of these complex systems is that they constantly evolve in time. This implies that the underlying networks are not static, but continuously change through the addition and/or removal of new nodes and links. Consequently, we have to uncover the dynamical forces that act at the level of individual nodes, whose cumulative effect determines the system's large-scale topology. A step in this direction was the scale-free model [8], that incorporates the fact that network evolution is driven by at least two coexisting mechanisms: 1) growth, implying that networks continuously expand by the addition of new nodes; 2) preferential attachment, mimicking the fact that a new node links with higher probability to nodes that already have a large number of links. With these two ingredients the scale-free model predicts the emergence of a power law connectivity distribution, observed in many systems [3,8–10], ranging from the Internet to citation networks. Furthermore, extensions of this model, including rewiring [11] or aging [12,13], have been able to account for more realistic aspects of the network evolution, such as the existence of various scaling exponents or cutoffs in the connectivity distribution.

The scale-free model neglects an important aspect of competitive systems: not all nodes are equally successful in acquiring links [17]. The model predicts that all nodes increase their connectivity in time as  $k_i(t) = (t/t_i)^\beta$ , where  $\beta = 1/2$  and  $t_i$  is the time at which node  $i$  has been added into the system. Consequently, the oldest nodes will have the highest number of links, since they had the longest timeframe to acquire them.

On the other hand, numerous examples convincingly indicate that in real systems a node's connectivity and growth rate does not depend on its age alone. For example, in social systems some individuals are better in turning a random meeting into a lasting social link than others. On the www some documents through a combination of good content and marketing acquire a large number of links in a very short time, easily overtaking older websites. Finally, some research papers in a short timeframe acquire a very large number of citations. We tend to associate these differences with some intrinsic quality of the nodes, such as the social skills of an individual, the content of a web page, or the content of a scientific article. We will call this the node's fitness, describing its ability to compete for links at the expense of other nodes.

In this paper we propose a simple model that allows us to investigate this competitive aspect of real networks in quantitative terms. Assuming that the existence of a fitness modifies the preferential attachment to compete for links, we find that different fitness translates into multiscaling in the dynamical evolution: the time dependence of a node's connectivity depends on the fitness of the node. We develop the continuum model for this competitive evolving network, allowing us to calculate  $\beta$  analytically and derive a general expression for the connectivity distribution. We find that the analytical predictions are in excellent agreement with the results obtained from numerical simulations.

*The fitness model.* – The examples discussed above indicate that nodes have different ability (fitness) to compete for links. To account for these differences we introduce a fitness parameter,  $\eta_i$ , that we assign to each node, and assume that it is unchanged in time (*i.e.*  $\eta_i$  represents a quenched noise) [18]. Starting with a small number of nodes, at every timestep we add a new node  $i$  with fitness  $\eta_i$ , where  $\eta$  is chosen from the distribution  $\rho(\eta)$ . Each new node  $i$  has  $m$  links that are connected to the nodes already present in the system. We assume that the probability  $\Pi_i$  that a new node will connect to a node  $i$  already present in the network depends on the connectivity  $k_i$  and on the fitness  $\eta_i$  of that node, such that

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}. \quad (1)$$

This generalized preferential attachment [8] incorporates in the simplest possible way that fitness and connectivity jointly determine the rate at which new links are added to a given node, *i.e.* even a relatively young node with a few links can acquire links at a high rate if it has a large fitness parameter. To address the scaling properties of this model we first develop a continuum theory, allowing us to predict the connectivity distribution [8, 11, 12]. A node  $i$  will increase its connectivity  $k_i$  at a rate that is proportional to the probability (1) that a new node will attach to it, giving

$$\frac{\partial k_i}{\partial t} = m \frac{\eta_i k_i}{\sum_j k_j \eta_j}. \quad (2)$$

The factor  $m$  accounts for the fact that each new node adds  $m$  links to the system. If  $\rho(\eta) = \delta(\eta - 1)$ , *i.e.* all fitness are equal, (2) reduces to the scale-free model, which predicts that  $k_i(t) \sim t^{1/2}$  [8]. In order to solve (2) we assume that similarly to the scale-free model the time evolution of  $k_i$  follows a power law, but there is multiscaling in the system, *i.e.* the

dynamic exponent depends on the fitness  $\eta_i$ ,

$$k_{\eta_i}(t, t_0) = m \left( \frac{t}{t_0} \right)^{\beta(\eta_i)}, \quad (3)$$

where  $t_0$  is the time at which the node  $i$  was born. The dynamic exponent  $\beta(\eta)$  is bounded, *i.e.*  $0 < \beta(\eta) < 1$  because a node always increases the number of links in time ( $\beta(\eta) > 0$ ) and  $k_i(t)$  cannot increase faster than  $t$  ( $\beta(\eta) < 1$ ). We first calculate the mean of the sum  $\sum_j \eta_j k_j$  over all possible realizations of the quenched noise  $\{\eta\}$ . Since each node is born at a different time  $t_0$ , the sum over  $j$  can be written as an integral over  $t_0$ :

$$\begin{aligned} \left\langle \sum_j \eta_j k_j \right\rangle &= \int d\eta \rho(\eta) \eta \int_1^t dt_0 k_\eta(t, t_0) \\ &= \int d\eta \eta \rho(\eta) m \frac{(t - t^{\beta(\eta)})}{1 - \beta(\eta)}. \end{aligned} \quad (4)$$

Since  $\beta(\eta) < 1$ , in the  $t \rightarrow \infty$  limit  $t^{\beta(\eta)}$  can be neglected compared to  $t$ , thus we obtain

$$\left\langle \sum_j \eta_j k_j \right\rangle \stackrel{t \rightarrow \infty}{\approx} C m t (1 + O(t^{-\epsilon})), \quad (5)$$

where

$$\begin{aligned} \epsilon &= (1 - \max_\eta \beta(\eta)) > 0, \\ C &= \int d\eta \rho(\eta) \frac{\eta}{1 - \beta(\eta)}. \end{aligned} \quad (6)$$

Using (5), and the notation  $k_\eta = k_{\eta_i}(t, t_0)$ , the dynamic equation (2) can be written as

$$\frac{\partial k_\eta}{\partial t} = \frac{\eta k_\eta}{C t}, \quad (7)$$

which has a solution of form (3), given that

$$\beta(\eta) = \frac{\eta}{C}, \quad (8)$$

thereby confirming the self-consistent nature of the assumption (3). To complete the calculation we need to determine  $C$  from (6) after substituting  $\beta(\eta)$  with  $\eta/C$ ,

$$1 = \int_0^{\eta_{\max}} d\eta \rho(\eta) \frac{1}{\frac{C}{\eta} - 1}, \quad (9)$$

where  $\eta_{\max}$  is the maximum possible fitness in the system [19]. Apparently, (9) is a singular integral. However, since  $\beta(\eta) = \eta/C < 1$  for every value of  $\eta$ , we have  $C > \eta_{\max}$ , thus the integration limit never reaches the singularity. Note also that, since  $\sum_j \eta_j k_j \leq \eta_{\max} \sum_j k_j = 2m t \eta_{\max}$ , we have, using (5), that  $C \leq 2\eta_{\max}$ .

Finally, we can calculate the connectivity distribution  $P(k)$ , which gives the probability that a node has  $k$  links. If there is a single dynamic exponent  $\beta$ , the connectivity distribution follows the power law  $P(k) \sim k^\gamma$ , where the connectivity exponent is given by  $\gamma = 1/\beta + 1$ .

However, in this model we have a spectrum of dynamic exponents  $\beta(\eta)$ , thus  $P(k)$  is given by a weighted sum over different power laws. To find  $P(k)$  we need to calculate the cumulative probability that for a certain node  $k_\eta(t) > k$ ,

$$\begin{aligned} P(k_\eta(t) > k) &= P\left(t_0 < t \left(\frac{m}{k}\right)^{C/\eta}\right) \\ &= t \left(\frac{m}{k}\right)^{\frac{C}{\eta}}. \end{aligned} \quad (10)$$

Thus the connectivity distribution, *i.e.* the probability that a node has  $k$  links, is given by the integral

$$\begin{aligned} P(k) &= \int_0^{\eta_{\max}} d\eta \frac{\partial P(k_\eta(t) > k)}{\partial t} \\ &\propto \int d\eta \rho(\eta) \frac{C}{\eta} \left(\frac{m}{k}\right)^{\frac{C}{\eta}+1}. \end{aligned} \quad (11)$$

*Scale-free model.* – Given the fitness distribution  $\rho(\eta)$ , the continuum theory allows us to predict both the dynamics, described by the dynamic exponent  $\beta(\eta)$  (eqs. (8) and (9)), and the topology, characterized by the connectivity distribution  $P(k)$  (eq. (11)). To demonstrate the validity of our predictions, in the following we calculate these quantities for two different  $\rho(\eta)$  functions. As a first application, let us consider the simplest case, corresponding to the scale-free model, when all fitnesses are equal. Thus we have  $\rho(\eta) = \delta(\eta - 1)$ , which, inserted in (9), gives  $C = 2$ , which represents the largest possible value of  $C$ . Using (8) we obtain  $\beta = 1/2$  and from (11) we get  $P(k) \propto k^{-3}$ , the known scaling of the scale-free model. Thus the scale-free model represents an extreme case of the fitness model considered here, the connectivity exponent taking up the largest possible value of  $\gamma$ .

*Uniform fitness distribution.* – The behavior of the system is far more interesting, however, when nodes with different fitness compete for links. The simplest such case, which already offers nontrivial multiscaling, is obtained when  $\rho(\eta)$  is chosen uniformly from the interval  $[0, 1]$ . The constant  $C$  can be determined again from (9), which gives

$$\exp[-2/C] = 1 - 1/C, \quad (12)$$

whose solution is  $C^* = 1.255$ . Thus, according to (8), each node will have a different dynamic exponent, given by  $\beta(\eta) \sim \frac{\eta}{C^*}$ . Using (11) we obtain

$$P(k) \propto \int_0^1 d\eta \frac{C^*}{\eta} \frac{1}{k^{1+C^*/\eta}} \sim \frac{k^{-(1+C^*)}}{\log(k)}, \quad (13)$$

*i.e.* the connectivity distribution follows a generalized power law, with an inverse logarithmic correction.

To check the predictions of the continuum theory we performed numerical simulations of the discrete fitness model, choosing fitness with equal probability from the interval  $[0, 1]$ . Most important is to test the validity of the ansatz (3), for which we recorded the time evolution of nodes with different fitness  $\eta$ . As fig. 1 shows, we find that  $k_i(t)$  follows a power law for all  $\eta$ , and the scaling exponent,  $\beta(\eta)$ , depends on  $\eta$ , being larger for nodes with larger fitness. Equation (6) predicts that the sum  $\langle \sum_i \eta_i k_i \rangle / mt \rightarrow C^*$  in the  $t \rightarrow \infty$  limit, where  $C^*$  is given by (12) as  $C^* = 1.255$ . Indeed as the inset in fig. 1 shows, the discrete network model indicates that this sum converges to the analytically predicted value. Figure 1 allows us to determine

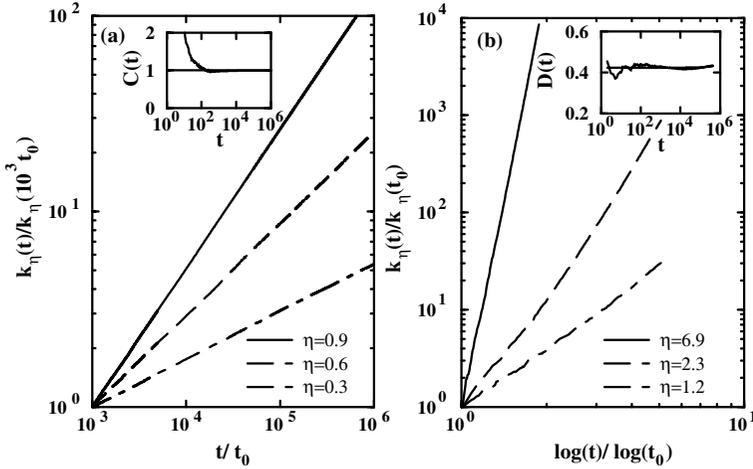


Fig. 1 – (a) Time dependence of the connectivity,  $k_\eta(t)$ , for nodes with fitness  $\eta = 0.3, 0.6$  and  $0.9$ . Note that  $k_\eta(t)$  follows a power law in each case and the dynamic exponent  $\beta(\eta)$ , given by the slope of  $k(t)$ , increases with  $\eta$ . While in the simulation the fitness of the nodes has been drawn uniformly, between  $[0, 1]$ , in the figure we show only the connectivity of three nodes with selected fitness. In the simulation we used  $m = 2$  and the shown curves represent averages over 20 runs. Inset: Asymptotic convergence of  $C(t) = (\sum_{i=1}^t \eta_i k_i)/t$  to the analytically predicted limit  $C^* = 1.255$ , shown as a horizontal line (see eq. (12)). (b) The same as (a) for the exponential fitness distribution demonstrating that  $k(t)$  follows (14). The inset shows the convergence of  $D(t) = (\sum_{i=1}^t \eta_i k_i)/(m \ln(t))$  to  $D^* = 0.45$ .

numerically the exponent  $\beta(\eta)$ , and compare it to the prediction (8). As the inset in fig. 2 indicates, we obtain excellent agreement between the numerically determined exponents and the prediction of the continuum theory. Finally, in fig. 2 we show the agreement between the prediction (13) and the numerical results for the connectivity distribution  $P(k)$ .

An interesting feature of the numerically determined connectivity distribution (fig. 2) is the appearance of a few nodes that have higher number of links than predicted by the connectivity distribution. Such highly connected hubs, appearing as a horizontal line with large  $k$  on the log-log plot, are present in many systems, including the www [3] or the metabolic network of a cell [20], clearly visible if we do not use logarithmic binning. This indicates that the appearance of a few “super hubs”, *i.e.* nodes that have connections in excess to that predicted by a power law, is a generic feature of competitive systems.

*Exponential fitness distribution.* – If the  $\rho(\eta)$  distribution has an infinite support, the integral (9) contains a singularity at  $\eta = C$ , and the self-consistent calculation cannot be applied. To recover the behavior of such systems, we studied numerically the case  $\rho(\eta) = e^{-\eta}$ . In a system with a finite support for which there is a  $\eta_{\max}$  such that  $\rho(\eta_{\max}) \neq 0$ , the system will reach  $\eta_{\max}$  within a finite time. That is, within a short timeframe a  $\eta'$  will appear that is infinitely close to  $\eta_{\max}$ , and the likelihood of finding an  $\eta > \eta'$  goes to zero. This is not the case for an infinite support: at any time there is a finite probability that an  $\eta > \eta_{\max}$  will appear, as, strictly speaking,  $\eta_{\max} = \infty$ . The average time required for a large  $\eta$  to appear scales as  $\tau(\eta) \sim 1/p(\eta) \sim e^\eta$ , indicating that  $\eta_{\max}$  scales as  $\eta_{\max} \sim \ln(t)$ , and  $\sum_i \eta_i k_i(t) \leq D \ln(t)t$ .

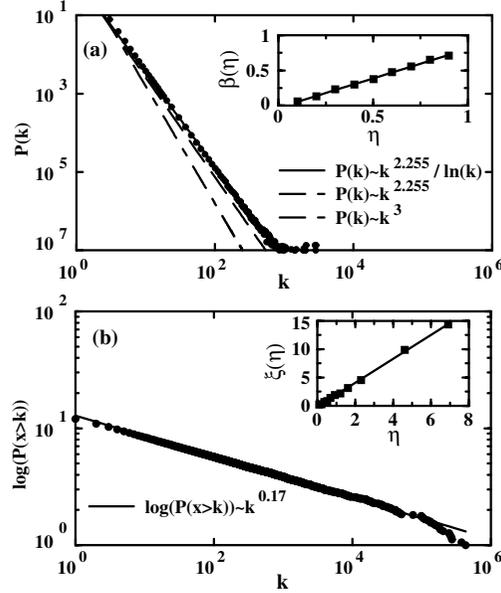


Fig. 2 – (a) Connectivity distribution in the fitness model, obtained for a network with  $m = 2$  and  $N = 10^6$  nodes and uniform fitness distribution. The upper solid line that goes along the circles provided by the numerical simulations corresponds to the theoretical prediction (13), with  $\gamma = 2.25$ . The dashed line corresponds to a simple fit  $P(k) \sim k^{-2.255}$  without the logarithmic correction, while the long-dashed curve correspond to  $P(k) \sim k^{-3}$ , as predicted by the scale-free model [8], in which all fitnesses are equal. Inset: The dependence of the dynamic exponent  $\beta(\eta)$  on the fitness parameter  $\eta$  in the case of a uniform  $\rho(\eta)$  distribution. The squares were obtained from the numerical simulations while the solid line corresponds to the analytical prediction  $\beta(\eta) = \eta/1.255$ . (b) Same as (a) but for exponential fitness distribution. To decrease statistical fluctuations we show the cumulative distribution  $P(x > k)$  [13], which follows a stretched exponential. The inset shows  $\xi(\eta)$  as determined from fig. 1b using eq. (14) and demonstrate that  $\xi(\eta)$  is linear in  $\eta$ , in line with the theoretical predictions.

Assuming that  $\sum_i \eta_i k_i(t) / D \ln(t) \rightarrow Dm$  in the  $t \rightarrow \infty$  limit, using (2) we obtain that

$$k(t) = k(t_0) \left( \frac{\ln(t)}{\ln(t_0)} \right)^{\xi(\eta)}, \quad (14)$$

where  $\xi(\eta) = D/\eta$ . As fig. 1b shows, we find that indeed  $k(t)$  scales as a power of  $\ln(t)$ , while fig. 2b (inset) shows that the power depends linearly on  $\eta$ . Interestingly, numerical simulations indicate that in this case  $P(k)$  follows a stretched exponential (fig. 2b).

*Discussion* – The fitness model investigated in this paper reflects the basic properties of many real systems in which the nodes compete for links with other nodes, thus a node can acquire links only at the expense of the other nodes. The competitive nature of the model is guaranteed by the fact that nodes that are already in the system have to compete with a linearly increasing number of other nodes for a link. We find that allowing for different fitness, multiscaling emerges and the time dependence of a node’s connectivity depends on the fitness parameter,  $\eta$ . This allows nodes with a higher fitness to enter the system at a later time and overcome nodes that have been in the system for a much longer timeframe. Our results indicate, however, that not all  $\rho(\eta)$  distributions will result in a power law time

dependence and connectivity distribution. If  $\rho(\eta)$  decays exponentially, we find that  $P(k)$  follows a stretched exponential and  $k(t)$  follows a complex combination of logarithmic and power law behavior. This indicates that  $P(k)$  is not robust against changes in the functional form of the fitness distribution: with an appropriate choice of  $\rho(\eta)$  one can obtain a non-power-law distribution. As many real networks display  $P(k)$  that are best approximated with a power law, this implies that there are some restrictions regarding the nature of the  $\rho(\eta)$  distribution. For example, an exponential is clearly not appropriate. Understanding the restrictions on the classes of  $\rho(\eta)$  which support a power law  $P(k)$  is a formidable challenge for further work.

\* \* \*

We wish to acknowledge useful discussions with R. ALBERT, I. DERÉNYI, H. JEONG, E. SZATHMÁRY and T. VICSEK. This research was partially supported by NSF Career Award DMR-9710998.

#### REFERENCES

- [1] ERDŐS P. and RÉNYI A., *Publ. Math. Inst. Acad. Sci.*, **4** (1960) 17.
- [2] BOLLOBÁS B., *Random Graphs* (Academic Press, London) 1985.
- [3] ALBERT R., JEONG H. and BARABÁSI A.-L., *Nature*, **401** (1999) 130.
- [4] NEWMAN M. E. J., *J. Stat. Phys.*, **101** (2000) 819.
- [5] CALDARELLI G., MARCHETTI R. and PIETRONERO L., *Europhys. Lett.*, **52** (2000) 386.
- [6] KLEINBERG J., KUMAR S. R., RAGHAVAN P., RAJAGOPALAN S. and TOMKINS A., *International Conference on Combinatorics and Computing* (1999).
- [7] HUBERMAN B. A. and ADAMIC L. A., *Nature*, **401** (1999) 131.
- [8] BARABÁSI A.-L. and ALBERT R., *Science*, **286** (1999) 509.
- [9] FALOUTSOS M., FALOUTSOS P. and FALOUTSOS C., *Comput. Commun. Rev.*, **29** (1999) 251.
- [10] REDNER S., *Eur. Phys. J. B*, **4** (1998) 131.
- [11] ALBERT R. and BARABÁSI A.-L., *Phys. Rev. Lett.*, **85** (2000) 5234.
- [12] DOROGOVTSSEV S. N. and MENDES J. F. F., *Phys. Rev. E*, **62** (2000) 1842.
- [13] AMARAL L. A. N., SCALA A., BARTHÉLÈMY M. and STANLEY H. E., *Proc. Natl. Acad. Sci. USA*, **97** (2000) 11149.
- [14] WATTS D. J. and STROGATZ S. H., *Nature*, **393** (1998) 440.
- [15] BARTHÉLÈMY M. and AMARAL L. A. N., *Phys. Rev. Lett.*, **82** (1999) 3180.
- [16] BANAVAR J. R. and MARITAN A. and RINALDO A., *Nature*, **399** (1999) 130.
- [17] ADAMIC L. A. and HUBERMAN B. A., *Science*, **287** (2000) 2115.
- [18] Note that in some real systems the fitness can change with time, for example a research field can slowly close down or an actor can suspend acting diminishing the ability of the corresponding nodes to compete for links [12,13].
- [19] Equation (9) can also be derived from the normalization condition  $2k_0t = \sum_{j \in N(t)} k_j$ , a “mass conservation” law, giving the total number of links in the network at time  $t$ .
- [20] JEONG H., TOMBOR B., ALBERT R., OLTVAI Z. and BARABÁSI, A.-L., *Nature*, **407** (2000) 651.