

# Power-Law Distribution of the World Wide Web

Barabási and Albert (1) propose an improved version of the Erdős-Rényi (ER) theory of random networks to account for the scaling properties of a number of systems, including the link structure of the World Wide Web (WWW). The theory they present, however, is inconsistent with empirically observed properties of the Web link structure.

Barabási and Albert write that because “of the preferential attachment, a vertex that acquires more connections than another one will increase its connectivity at a higher rate; thus, an initial difference in the connectivity between two vertices will increase further as the network grows. . . . Thus older . . . vertices increase their connectivity at the expense of the younger . . . ones, leading over time to some vertices that are highly connected, a ‘rich-get-richer’ phenomenon” [figure 2C of (1)]. It is this prediction of the Barabási-Albert (BA) model, however, that renders it unable to account for the power-law distribution of links in the WWW [figure 1B of (1)].

We studied a crawl of 260,000 sites, each one representing a separate domain name. We counted how many links the sites received

from other sites, and found that the distribution of links followed a power law (Fig. 1A). Next, we queried the InterNIC database (using the WHOIS search tool at [www.networksolutions.com](http://www.networksolutions.com)) for the date on which the site was originally registered. Whereas the BA model predicts that older sites have more time to acquire links and gather links at a faster rate than newer sites, the results of our search (Fig. 1B) suggest no correlation between the age of a site and its number of links.

The absence of correlation between age and the number of links is hardly surprising; all sites are not created equal. An exciting site that appears in 1999 will soon have more links than a bland site created in 1993. The rate of acquisition of new links is probably proportional to the number of links the site already has, because the more links a site has, the more visible it becomes and the more new links it will get. (There should, however, be an additional proportionality factor, or growth rate, that varies from site to site.)

Our recently proposed theory (2), which accounts for the power-law distribution in the number of pages per site, can also be applied to the number of links a site receives. In this model, the number of new links a site receives at each time step is a random fraction of the number of links the site already has. New sites, each with a different growth rate, appear at an exponential rate. This model yields scatter plots similar to Fig. 1B, and can produce any power-law exponent  $\gamma > 1$ .

Lada A. Adamic  
 Bernardo A. Huberman  
 Xerox Palo Alto Research Center  
 3333 Coyote Hill Road  
 Palo Alto, CA 94304, USA  
 E-mail: [ladamic@parc.xerox.com](mailto:ladamic@parc.xerox.com)

**References**

1. A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
2. B. A. Huberman and L. A. Adamic, *Nature* **401**, 131 (1999).

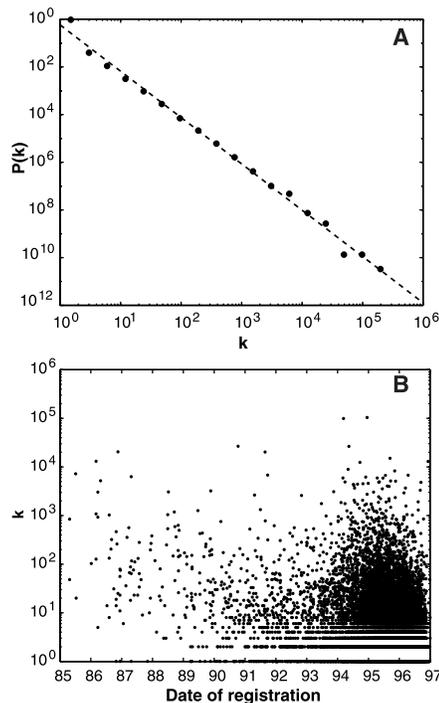
10 November 1999; accepted 4 February 2000

*Response:* Adamic and Huberman offer additional support for the evolutionary network model that we offered (1). The apparent mess in their fig. 1B is rooted in their choice not to average their data. We believe that taking the average over all points of the same age, and extracting the trends within those averages, would have unveiled the increasing tendency predicted by our model.

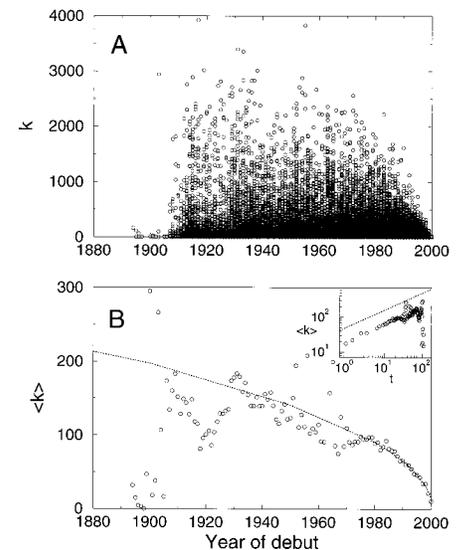
Although we do not have access to their

data, we can illustrate the same procedure for the network of movie actors that we discussed (1). When the connectivity of the individual actors is plotted as a function of the release year of their first movie (Fig. 1A), the results are very similar to those shown in fig. 1B of Adamic and Huberman’s comment. The only difference is that the movie industry had its boom not 4 years ago, as did the WWW, but rather at the beginning of the century; thus, the apparently structureless regime persists much longer. When the connectivity of the actors that debuted in the same year is averaged, however, the average connectivity in the last 60 years increases with the actor’s age, in line with the predictions of our theory, and the curve follows a power law for almost a hundred years (Fig. 1B). We expect that a similar increasing tendency would appear for the WWW data after averaging, but the length of the scaling interval would be limited by the Web’s comparatively brief history.

The fluctuations that lead to the apparent randomness of Fig. 1A are due to the individual differences in the rate at which nodes increase their connectivity. It is easy to include such differences in the model and continuum theory proposed by



**Fig. 1.** (A) The distribution function for the number of links,  $k$ , to Web sites (from crawl in spring 1997). The dashed line has slope  $\gamma = 1.94$ . (B) Scatter plot of the number of links,  $k$ , versus age for 120,000 sites. The correlation coefficient is 0.03.



**Fig. 1.** (A) Scatter plot of movie actor connectivity,  $k$  (the number of other actors with which he or she performed during his or her career), versus the year of debut. All actors from the Internet Movie Database were included;  $n = 392,340$ . (B) Average movie actor connectivity,  $\langle k \rangle$ , versus year of debut. To determine  $\langle k \rangle$ ,  $k$  is averaged over all actors that debuted in the same year. The curve shows a systematic increase in the average connectivity with the actor’s professional lifetime,  $t = (2000 - \text{year of debut})$ . The dotted line follows  $k(t) \sim t^\alpha$ , with  $\alpha = 0.49$ , very close to the prediction  $\alpha = 0.5$  of (1). Inset shows a log-log plot of  $\langle k \rangle$  as a function of  $t$ , which illustrates the presence of scaling in the last century. The dotted line has slope 0.5.

## TECHNICAL COMMENT

us ( $I$ ). Assigning intrinsic growth rates,  $\eta_i$ , randomly to each vertex  $i$ , mimicking their fitness to acquire links, modifies the growth equation to  $\partial_t k_i = \eta_i k_i / \sum_j \eta_j k_j$ . This can account for the different growth rates witnessed for different Web sites and different actors: the connectivity of a given node is predicted to increase as  $k(t) \sim t^{\beta(\eta)}$ , where  $\beta(\eta)$  can be determined analytically. Additional system-specific de-

tails, such as the inclusion of new internal links, rewiring, and aging, can be also added to the model and to the continuum theory (2-4).

**A.-L. Barabási**  
**R. Albert**  
**H. Jeong**  
**G. Bianconi**

*Department of Physics*  
*University of Notre Dame*

*Notre Dame, Indiana 46556, USA*  
*E-mail: alb@nd.edu*

### References

1. A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
2. ———, H. Jeong, *Physica* **272A**, 173 (1999).
3. S. N. Dorogovtsev and J. F. F. Mendes, <http://xxx.lanl.gov/abs/cond-mat/0001419>.
4. L. A. N. Amaral, A. Scala, M. Barthélémy, H. E. Stanley, <http://xxx.lanl.gov/abs/cond-mat/0001458>.

13 January 2000; accepted 4 February 2000