Scale-Free Networks Using Local Information for Preferential Linking

Aldridge, C.

Department of Information Science, University of Otago, P. O. Box 56, Dunedin 9001, New Zealand,
E-Mail: caldridge@infoscience.otago.ac.nz

Keywords: *MODSIM; scale-free network*

EXTENDED ABSTRACT

Scale-free networks are a recently developed approach to modeling the interactions found in complex natural and man-made systems. Such networks exhibit a power-law distribution of node link (degree) frequencies $P(k)$ in which a small number of highly connected nodes predominate over a much greater number of sparsely connected ones.

The importance of scale-free networks is emphasized by the number of real networks now identified as exhibiting power-law distributions. A recently identified, but now classic example of a scale-free network is the World Wide Web: web pages are nodes, which are connected by hyperlinks. Other examples of such networks traverse disparate fields: scientific paper citations, communications networks and power grids, neural networks, and protein-protein interactions.

The now classic Albert-Barabási constructive algorithm for constructing scale-free networks centers on the concept of preferential attachment in which the probability of a new node linking to an existing node is proportional to its relative number of links. This paper contends that when a new node is appended, the global knowledge of node degrees required by Albert-Barabási approach is unrealistic. Instead we propose a locally-derived linking criterion in which only a small part of the total network is considered each time. The Albert-Barabási constructive algorithm then becomes a limiting case of the proposed local algorithm. This paper investigates the some of properties of the resulting networks.

As shown in Figure 1, the degree distributions of networks constructed using local preferential linking exhibit an unusually fat tail. That is, the network has a small number of extremely well-connected nodes which, in effect, comprise 'super hubs'.

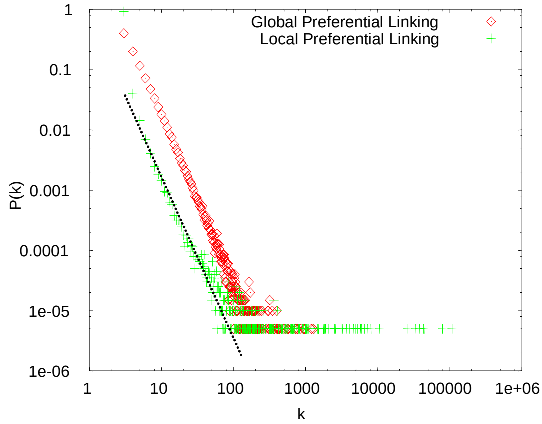

Figure 1. Node degree distributions $P(k)$ for 200,000 node networks each with linkage rate $m = 3$ created using global preferential linking (Barabási-Albert) versus local preferential linking with neighborhood distance $l = 1$ (this research). The trend-line is an eyeballed estimate.

The cumulative probability distribution of node degrees proves to be a sensitive means of evaluating subtle differences between generated distributions. Figure 2 highlights the existence of 'super hubs' in the extended tail of the connectivity distribution produced by from the local preferential linking approach.

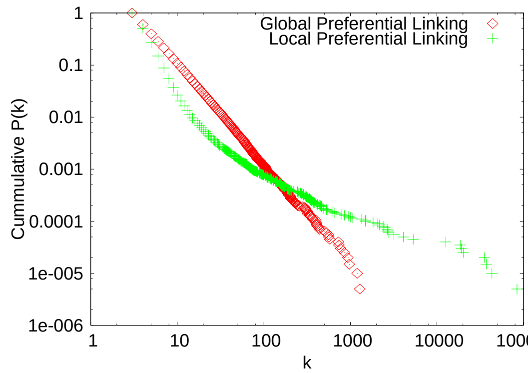

Figure 2. Cumulative node degree distributions $P(k)$ for 200,000 node networks each with linkage rate $m = 3$ created using global preferential linking (Barabási-Albert) versus local preferential linking with neighborhood distance $l = 1$ (this research).

1. INTRODUCTION

1.1. Random Networks

Networks have been used to model the structure of many systems. Objects are represented as nodes,[1] relationships between objects as links. Links may be directed from one object to another, or non-directed. The number of links for a given node is the degree *k* of the node. [2] The degree distribution of a network with *N* nodes *P*(*k, N*) is a key statistic in describing the network.

"Classical" random networks as characterized by Erdös and Rényi (1959, 1961) comprise a fixed number of nodes *N*. In constructing a network, links between nodes are made randomly with equal probability. Such networks have a Poisson distribution in which the degree distribution decreases rapidly; highly connected nodes are vanishingly unlikely, as in Figure 3, "Static network, random linking". These networks were the focus of attention for network science for almost four decades.

By way of contrast, an empirical study of part of a real-world network—the World Wide Web (*www*)—identified a 'fat-tailed' degree distribution having many highly-connected nodes like that shown in Figure 3 for "Growing network, preferential linking" (Albert et al. 1999). A web crawler was used to compile a complete map of the Notre Dame University domain, *nd.edu*. Web pages were mapped as nodes; hyperlinks as links. The degree distribution of this network was found to follow a power law, $P(k) \propto k^{-\lambda}$. When log(*P*(*k*)) is plotted against log(*k*), this function is linear with a slope of -λ (Figure 3). For the domain investigated, outgoing hyperlinks had $\lambda_{out} = 2.45$ and incoming links, $\lambda_{in} = 1.1$. Significantly, the degree distribution is independent of the size of the network. It is scale-free.

Key properties of scale-free networks are that they are dominated by a relatively few, highly connected nodes ('hubs'), with the vast majority of nodes being poorly connected. Scale-free networks are therefore extremely resistant to disruption by random deletion of nodes. On the other hand, targeted deletion of hubs can rapidly destroy a large proportion of network connectivity.

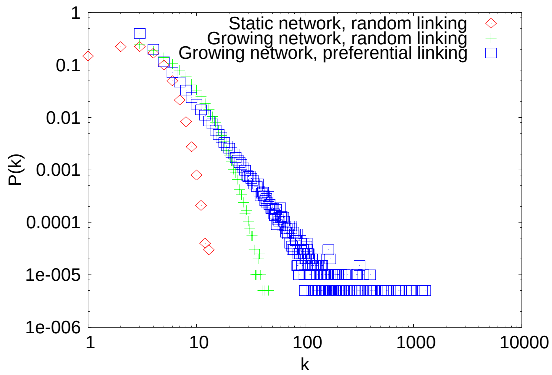

Figure 3. Empirical node degree distributions *P*(*k*) for networks that link to nodes with equal probability, compared with the Barabási-Albert network where the probability of linking is proportional to node degree, *k*. Each network has 200,000 nodes.

Scale-free networks also exhibit the 'small-world' effect (Milgram, 1967, Watts and Strogatz, 1988) in that the average shortest distance between randomly selected nodes is relatively small. For instance Barabási et al. (1999) estimated the average diameter of the *www* is $\langle d \rangle = 18.59$ links. In other words, any two pages of what was then estimated to be the 800 million web pages of the World Wide Web are, in principle, only separated by an average of about 19 clicks. (This, of course, assumes an intelligent search. A random search would take a many, many more clicks.)

Theoretical physicists, principally Dorogovtsev and Mendes (2001), have used a continuum approach to develop a theory of evolving networks. An extensive review of growing networks, their properties, theoretical models, and real-world examples, is presented in Dorogovtsev and Mendes 2003.

Having identified the *www* as a scale-free network, Barabási and Albert (1999) went on to describe a simple construction algorithm for a growing network based upon *preferential linking*. In the method of preferential linking used, the probability of establishing a link from a new node to an existing node is proportional to the relative number of its links (its degree *k*). The algorithm was shown to produce a power-law degree distribution analogous to that found for the *www*, with $\lambda \approx 2.9$.

In their 1999 paper, Barabási and Albert demonstrated that *both* the growth *and* the preferential linking features of their algorithm are required to generate a network with a power-law

[1] The nomenclature of networks adopted varies with context: Mathematicians refer to networks as "graphs" made of "vertices" and "edges"; Physicists use the terms 'nodes' or 'sites', and 'links' or 'bonds'.

[2] Also referred to as 'connectivity'.

degree distribution. First, they used a growing algorithm in which new nodes were linked randomly with equal probability to existing nodes. This eliminated the power-law distribution, and resulted in a rapidly increasing, exponential, degree distribution, $P(k) \approx e^{-\beta k}$ shown in Figure 3, "Growing network, random linking". In a second approach, the researchers took a fixed number of nodes N, selected nodes at random, and then preferentially linked from them. While initially the resulting networks exhibited power-law scaling, this relationship ultimately breaks down because, if construction continues long enough, all nodes become mutually linked.

The Barabási-Albert scale-free network model has been found to describe a wide variety of phenomena (Table 1).

Table 1. Examples of real-world scale-free networks (In part after a compilation by Dorogovtsev and Mendes, 2003.)

Network	**No. of Nodes**	**No. of Links**	**λ**	**Source**
Map of nd.edu domain	326,000	1,470,000	λ_{in} = 2.1 λ_{out} = 2.45	Albert et al. 1999
ISI citations 1981 – June 1997	783,000	6,716,000	λ_{in} = 3.0	Redner 1988
Collaboration network of screen actors	212,000	61,086,000	λ = 2.3	Newman 2001e
Web of human sexual contacts	2810	—	λ = 3.4	Liljeros et al. 2001
Protein-protein interactions (yeast proteome)	1,870	2,240	λ ≈ 2.5	Jeong et al. 2001
Java development framework—classes and interactions	1,376	2,174	λ = 2.5	Valverde et al. 2002
Large digital electronic circuits	20,000	40,000	λ = 3.0	Ferrer I Cancho et al. 2001b
Energy landscape network for a 14-atom cluster	4196	87,219	λ = 3.0	Doye 2002
Coauthorships in the SPIRES e-archive	56,600	4,899,000	λ = 1.2	Newman 2001
E-mail net	59,192	—	λ = 1.8	Ebel et al. 2002
English words, as they are linked within sentences	470,000	17,000,000	λ = 1.5, 2.7	Ferrer I Cancho et al. 2001a
Air transportation: non-stop passenger flights between cities.	3883	531,574	2.0	Guimerà et al. 2003, 2005

In this paper, we take issue with the Albert – Barabási construction algorithm, noting that their approach to preferential linking requires at all times global knowledge of the degree distribution $P(k, N)$ of the evolving network. This is readily apparent when implementing the algorithm, as at each time step[3] s, when a new node is added, the

[3] Note that in the "classic" Albert -- Barabási algorithm, one new node is added at each time

probability of attaching that node to an existing one i requires calculation of the probability $p(k_i) = k_i / \sum_j k_j$. That is, not only must the degrees of individual nodes be tracked, but a total of all node degrees must be maintained. In terms of large real-world networks, this is unrealistic. For instance, when a web page author is creating hyperlinks they do not consider the World Wide Web in its entirety; they use their local knowledge. Such knowledge is inevitably a tiny proportion of the present-day World Wide Web.

The apparent need for global knowledge while constructing a scale-free network is somewhat puzzling. The scale-free property suggests that the degree distribution of the network is independent of its size. Therefore, it should be possible to use knowledge of the degree distribution of some lesser part of the network as a sufficient alternative. If this were possible, then there would be no need for global knowledge during network construction. Surprisingly, this inconsistency appears to have escaped the attention of other researchers in the field.

We suggest that the use of local knowledge in network construction could result in a more realistic model for the behavior of real-world networks.

The above observations lead to the question, "Can a modified Albert-Barabási algorithm using only local knowledge create a scale-free network with similar properties to those evolved using global knowledge?"

2. NETWORK CONSTRUCTION USING LOCAL KNOWLEDGE

The Barabási-Albert network construction algorithm is as follows (Barabási and Albert 1999):

- Start with a small number of nodes m_0.
- At each time step s add a new node with linkage rate $m(\leq m_0)$ links made preferentially to existing network nodes. The probability of linking to node i is $p(k_i) = k_i / \sum_{m_0+s} k_j$.
- Stop when the network has reached the required size $N = m_0 + s$.

step. Therefore, if the initial size of a network is m_0, the size of the network at time s is $s + m_0$.

We now propose a modified algorithm requiring only a local knowledge of network nodes when preferentially linking from a new node:

- Start with a small number of nodes m_0.
- At each time step s:
  - Select an existing node u_s at random.
  - Assemble a set L_s of l_s "local nodes" consisting of u_s and those nodes within distance d of u_s (i.e. its "neighbourhood").[4]
  - Add a new node v_s with $m(\leq m_0)$ links made preferentially to m nodes in L_s. The probability of linking to node i of the l local nodes is $p(k_i) = k_i / \sum_l k_j$.
- Stop when the network has reached the required size $N = m_0 + s$.

In the above algorithm the method of preferential linking is the same as in the Barabási-Albert approach, except that instead of using the degree information from *all* nodes, only that from the set of "local nodes" L is used.

3. EMPIRICAL COMPARISON OF "LOCAL" AND "GLOBAL" NETWORKS

Python scripts were written to implement constructive algorithms for the following networks:

- A static network with random linking (the "traditional" Erdös-Rényi network).
- A growing network in which each new node is randomly linked with equal probability to existing nodes.
- A growing network with global preferential random linking. (The Barabási-Albert "scale-free" network, described above).
- A growing network with local preferential random linking, as proposed in this research (above).

Because scale-free effects are only readily apparent in large networks, the scripts were used to create networks that were as large as possible given the hardware and time available. The Psycho run-time compiler available for Python was used to minimize the effect on execution time of using a scripted language. The networks

[4] The distance between two nodes is the shortest path length between them, measured in links.

created and analyzed ranged in size from about 12,000 to 200,000 nodes

Figure 3 shows the node degree distributions for the well-known networks already introduced. The Erdös-Rényi network has a Poisson degree distribution, and the growing, randomly linked network an exponential degree distribution. In both these distributions the degree probability rapidly decreases as node degree increases. Both networks have a natural scale of the order of the average degree (Dorogovtsev and Mendes, 2003). By way of contrast, the Barabási-Albert network is fat-tailed; it has a small number of very highly connected nodes. The linear plot is consistent with a power law, $P(k) \propto k^{-\lambda}$, with $\lambda \approx 2.9$. The distribution is therefore scale-free.

Note that the spread of points in the power-law plot at low probabilities is a size effect found in real distributions. That is, the uncertainty in the distribution becomes large where there are only a few nodes (1–10) of the same connectivity.

The plot of degree distribution for the network constructed with local preferential linking shows a significant superficial resemblance to the plot for the Barabási-Albert network (Figure 4). On closer examination, the plot can be divided into three parts. The central part is essentially linear. For this section, the slope λ is estimated at 2.8. This is only slightly less than that for the global linked network. However, for the first point of the local distribution, the probability of $k = m = 3$ is 0.92. That is, 92% of network nodes failed to gain additional links through later preferential linking. This contrasts with the global models in which only 40% of nodes are in this category. Finally, the tail of the local distribution extends to include a few rare nodes that are much more connected than any in the equivalent global network model.

Plotting the respective cumulative degree probability distributions (Figure 5) confirms the significant differences between the two distributions. Here, the plot for the global preferentially linked network is still clearly linear, with an expected cumulative power function slope of λ-1 (Dorogovtsev and Mendes 2003). On the other hand, the local preferentially linked network appears to comprise an initial linear component with $\lambda \approx 4.3$ and a final, perhaps linear, component with a λ of around 1.7. The region between these components could be interpreted as a transition between them. The unusually long tail of the locally derived distribution confirms the presence of the few extremely highly connected nodes—what might be termed 'super hubs'. This is dramatically illustrated by comparing the most connected node for each of the respective 200,000 node example networks. The most connected node in network with global preferential linking has 1,283 links, while the equivalent for the local network has 107,716 links.

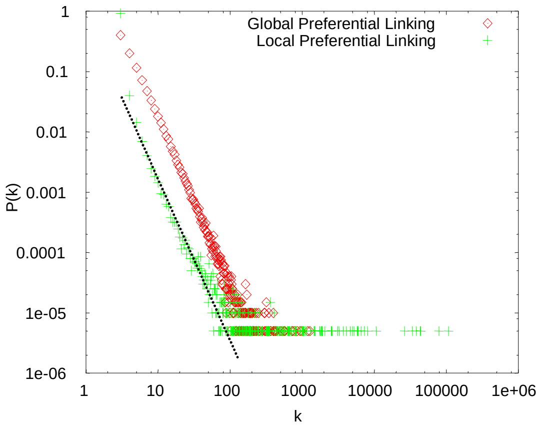

Figure 4. Node degree distributions $P(k)$ for 200,000 node networks each with linkage rate $m = 3$ created using global preferential linking (Barabási-Albert) versus local preferential linking with neighborhood distance $l = 1$ (this research). The trend-line is an eyeballed estimate.

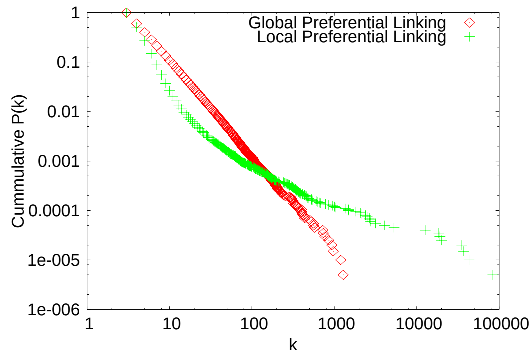

Figure 5. Cumulative node degree distributions $P(k)$ for 200,000 node networks each with linkage rate $m = 3$ created using global preferential linking (Barabási-Albert) versus local preferential linking with neighborhood distance $l = 1$ (this research).

This 'super hub' outcome for local linking appears to be a consequence of the constrained access to the network when assembling the set of local nodes L_s. For a network such as the above, where $m = 3$ and $d = 1$, most of the nodes randomly targeted within the network will be of degree $k = 3$. Such nodes are also likely to have neighbors with a connectivity of three. Therefore, most of the time the local neighborhood will also consist of three nodes, each with the same degree, $k = m = 3$. In such cases, the intended preferential linking degenerates to simple random linking. As noted above, this linking strategy in a growing network results in an exponential degree distribution, accounting for the initial rapidly decreasing node distribution in Figure 5.