RANDOMLY STOPPED LINKING GENERATES SCALE FREE NETWORKS

Josh Johnston and Tim Andersen

SIAM Workshop on Network Science 2020
July 9–10 · Toronto

Summary

Since the Barabási-Albert (BA) model was proposed twenty years ago, the power law degree distribution of scale-free networks is usually explained with preferential attachment. Popularly described as “the rich get richer”, preferential attachment ascribes inherent value to popularity. We present a randomly stopped linking model that forms scale free networks without the presence of preferential attachment.

Introduction

Though its applicability in network science is debated [4], the BA model states that degree distribution- the number of links per node- in a scale free network tends toward a power law due to growth and preferential attachment [2]. It suggests an irresistible first-mover advantage that has informed growth-at-all-cost strategies in many industries. Growth in the BA model means links are formed when new nodes join the network, so older nodes have more chances to collect links. Preferential attachment further disadvantages late arrivals because links form with probability proportional to the degree of the existing nodes. Many complex systems show latecomers overcoming this initial advantage, including Google’s search engine and Facebook’s social network platform displacing well-established players. This required tempering the BA model with a concept of intrinsic fitness [3].

Previous research demonstrates fitness alone is sufficient to generate scale free networks in special cases, controlling linking with a function of the fitnesses of the two vertices involved that is selected appropriately for the particular fitness distribution, requiring symmetry between the two linking nodes [7]. Our approach leads to scale free networks using a reparameterization of the Configuration Model (CM) [6] and intrinsic fitness, rather than a fitness linking function or preferential attachment and growth.

A Randomly Stopped Model

In the BA model, a node added to the network links to existing nodes, which are preferentially picked with probability proportional to the existing nodes’ degrees. Instead, our model considers that nodes begin with one link, and each link added is a discrete decision made in series. After the first link, there is a chance the node will gain another. If not, the process ends. If the node adds a second link, there is now a chance to add a third link, and so on. In the simplest approximation, we use a constant marginal probability to add each link.

$$p_X(k) = (1-p)^{k-1}p \tag{1}$$

$$\text{where } k \in \{1, 2, 3, ...\}$$

This randomly stopped process of adding links with the same probability is described by a geometric distribution (Equation 1), the number k Bernoulli trial failures before the first success. A single geometric distribution is not heavy-tailed and does not fit the high variance of a scale free network well. But if each node has a different fitness, the variance of parameter p over all nodes can also be high. In fact, a generalized central limit theorem for variables with infinite variance leads to heavy tails, even if the component distributions are not themselves heavy-tailed [8]. Specifically, a heavy tail results from mixing geometric distributions that have uniformly-distributed parameters [1].

$$p_k = \int_0^1 (1-p)^{k-1} p \, dp = \frac{1}{k(k+1)} \text{ for } k > 0 \tag{2}$$

We propose setting node degree according to a mixture of geometric distributions that together approximate a power law. We assign all nodes a link-stopping probability p pulled randomly from a uniform distribution between 0 to 1. As the number of nodes grows large, Equation 2 approximates the probability mass function (PMF) for the resulting mixture of geometric distributions. This is further approximated by a power law with $\gamma = 2$ because $p_k \sim k^{-2}$ as k gets larger.

$$p_k = \frac{1}{k(k+1)(k+2)} \text{ for } k > 0 \tag{3}$$

In comparison, the BA model has the exact degree distribution shown in Equation 3 and is approximated by a power law with $\gamma = 3$. In real networks, the scale free regime falls between these two models and is defined by $2 \leq \gamma \leq 3$.

We generated networks using the BA Model and the randomly stopped linking model each with 30,000 nodes. The BA Model adds 2 links with each new node, connecting to existing nodes with probability proportional to the degree of the existing nodes. Our randomly stopped linking model, however, does not use preferential attachment or growth. Instead, we form links following the CM [6]: Create nodes and allocate link stubs to each one, then select pairs of nodes randomly to connect until all stubs are linked. Our model is distinguished from other versions of the CM by how the stubs are allocated to nodes following a mixture of geometric distributions that approximates a power law. To determine the number of link stubs for each node, first assign a stopping probability from a random variable uniformly distributed between 0 and 1. This probability is the parameter p in Equation 1 and the resulting geometric distributions contribute to the overall mixture. Next, allocate each node a number of links (degree) from a geometrically-distributed random variable according to the PMF in Equation 1 and the node's fitness, p. Finally, connect these link stubs by selecting random pairs according to the CM.

Result

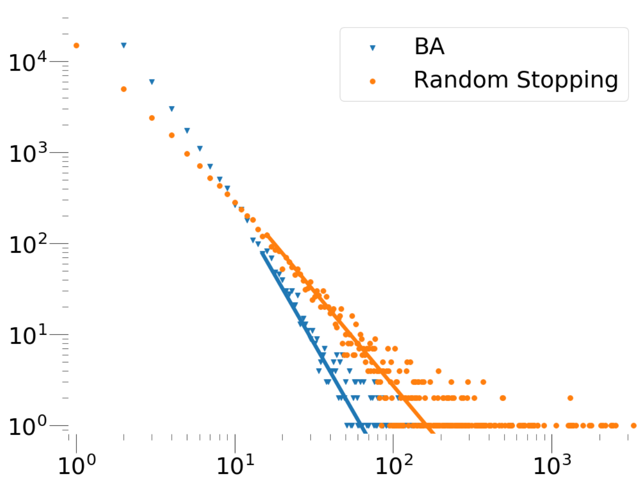

Figure 1: Degree distribution and power law fit for networks generated from the BA and random stopping models

	BA Model	Random Stopping Model
k_{min}	15	16
γ	3.03	2.08
p-value	0.18	0.95

Table 1: Power law fit to models

The degree distribution of both networks is shown as Figure 1. We fit power laws to each distribution using the method described in [5]. The results in Table 1 agree with the theoretically expected γ values for both distributions. The k_min is the degree after which the distribution behaves like a power law and is similar for both. We accept the power law as a plausible hypothesis when the p-value is greater than 0.1, which it is for both models. The p-value is substantially larger for the random stopping model, providing support for the hypothesis a mixture of geometric distributions results in a power law tail.

This randomly stopped model shows that a mixed geometric distribution with widely-varying intrinsic node fitness results in scale free networks where degree distribution approaches a power law. The model provides stronger agreement with a power law than the BA preferential attachment model without requiring any network effects or a fitness linking function. This alternative implies preferential attachment and growth may have lesser roles in forming scale free networks than generally assumed.

References

[1] E. P. Balanzario, R. M. Ramírez, and J. S. Ortiz. The randomly stopped geometric brownian motion. *Statistics and Probability Letters*, 90:85–92, 2014.

[2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[3] G. Bianconi and A.-L. Barabási. Competition and multiscaling in evolving networks. *Europhysics Letters (EPL)*, 54(4):436–442, 2001.

[4] A. D. Broido and A. Clauset. Scale-free networks are rare. *Nature Communications*, 10(1), 2019.

[5] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, Apr 2009.

[6] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[7] V. D. P. Servedio, G. Caldarelli, and P. Buttà. Vertex intrinsic fitness: How to produce arbitrary scale-free networks. *Physical Review E*, 70(5), 2004.

[8] W. Willinger, D. Alderson, J. Doyle, and L. Li. More "normal than normal: Scaling distributions and complex systems. *Proceedings of the 2004 Winter Simulation Conference, 2004.*, 2004.