

A Fitness-Based Model for Complex Networks

Zhou Fan '10[†]
Harvard University
Cambridge, MA 02138
zhoufan@fas.harvard.edu

Abstract

Complex networks such as the World Wide Web and social relationship networks are prevalent in the real world, and many exhibit similar structural properties. In this paper, a fitness-based model is developed for these complex networks. This model employs a purely “better-get-richer” method of network construction that is believed to realistically simulate the growth process of most real-world networks. Both computer-simulated results and theoretical analysis show that the degree distribution of networks created with this model depends on the distribution of vertex fitnesses; a power-law fitness distribution results in the commonly observed scale-free network structure. In addition, results indicate a small average path length and large clustering coefficient, in accordance with real-world phenomena. It is proposed that this model may serve as a possible explanation of the prevalence of scale-free networks in the real world.[‡]

4.1 Introduction

There are many examples of complex networks in the world, from the more common World Wide Web and social relationship networks to the more obscure power grid of the Western United States and network of scientific paper citations. Over the past decades, researchers have noted that many such real-world networks exhibit similar properties in structure and have studied and modeled them together under the term **complex networks**. A greater understanding of the structure of these abstract complex networks will undoubtedly heighten our understanding of the behavior of their real-world counterparts. Indeed, the study of complex networks has already led to advances in areas such as immunization and Internet simulation [BB1]. In this paper, we will provide a model of network growth similar to an existing model, but we will incorporate a fitness concept, and we will examine the structural properties of our model in comparison to real-world phenomena.

4.2 Background

In the field of complex networks, the individual network components are represented by vertices of a graph and the connections between them are represented by the edges. For instance, the vertices of a network representing the World Wide Web would be the web pages, with two vertices connected by an

[†]Zhou Fan, Harvard '10, is a prospective concentrator in mathematics or applied mathematics. He was born in Hangzhou, China and grew up in Parsippany, New Jersey, where he graduated from Parsippany Hills High School.

[‡]Part of the research for this paper was conducted at the 2005 Research Science Institute under the guidance of King Y. Yick, sponsored by a grant from the Center for Excellence in Education.

edge when there is a link from one page to the other. (For the purposes of this paper, we consider only undirected and unweighted edges.) It has been observed that the vertex degrees of a large majority of complex networks satisfy a power-law distribution, and such networks are called **scale-free** [AB].

The Barabási-Albert model (BA model), one of the most basic and widely-accepted models of complex networks, captures their scale-free structure [AB]. The BA model constructs networks based on the two ideas of network growth and preferential attachment: more popular vertices of a network attract more new vertices. In addition to being scale-free, networks constructed using this model have a small average path length between vertices and display a relatively high tendency for a vertex’s neighbors to connect to each other; this tendency is known as **clustering**. Both of these properties are also observed in real-world networks [AB]. One should note, however, that the BA model always predicts a power-law degree distribution where the probability density function of the vertex degrees, k , scales according to k^{-3} , while the degree distributions of real-world networks have varying powers of k . Also, a few real-world networks have an exponential degree distribution [St].

4.3 Fitness

The BA model relies on preferential attachment, the idea that a more popular website or scientific paper will attract more links or citations. A fundamentally different concept is that a more helpful, useful, ingenious, or simply “better” vertex will attract more such edges. This second concept is **fitness**-based, and the “better” vertices are deemed to be more **fit**. A weakness of the BA model is that it does not address fitness; for example, it does not allow a newer but very good scientific paper to become more frequently cited than an older but less significant one. Thus, a modification of the BA model has been developed that uses both preferential attachment and fitness [BB2]. This modified model, in essence, assumes that preferential attachment and fitness are separate and parallel causes of network structure.

In our paper, we examine whether a model based on fitness alone, without preferential attachment, can produce results similar to those produced by the BA model. This is intuitively reasonable; for example, a popular scientific paper probably becomes more frequently cited because it is better than other papers. We thus hypothesize that a model based solely on the fitness concept may produce results similar to those of the BA model. It should be noted that a network model based solely on the fitness concept has already been developed by Caldarelli et al., but it uses an approach to network construction different from that used in the BA model [CCRM, SC]. In this study, we instead examine a network construction algorithm based on the BA construction algorithm, but we employ the fitness concept instead of preferential attachment.

Specifically, our algorithm is as follows: Fix a probability distribution of fitnesses, $\rho(\eta)$, and the number of edges m with which a newly formed vertex starts. When the network grows sufficiently large so that the initial vertices do not matter, m becomes the average vertex degree. Begin with N_0 vertices, where N_0 is small. Randomly assign to each vertex a fitness value η from the fitness distribution $\rho(\eta)$, where a high value of η corresponds to a vertex that is more fit. Once a fitness value is assigned to a vertex, it does not change. At each time step $t = 1, 2, 3, \dots$, add one vertex to the network, connect it to m existing vertices, and assign to it a fitness value based on $\rho(\eta)$. For each of these m new connections, the probability of connecting to an existing vertex i with fitness η_i is proportional to η_i , i.e.,

$$P = \frac{\eta_i}{\sum_{j=1}^N \eta_j}$$

with N being the size of the network prior to the addition of this new point. We connect the m edges so that no two edges connect to the same vertex.

4.4 Degree Distribution

The degree distribution of networks created using this fitness-based algorithm can be examined using the **continuum theory**, a method developed by Barbási and Albert in which network growth is treated as a continuous process to allow simplification of the model using calculus [AB]. Such an approximation should match closely with discrete network growth, provided that we consider networks of sufficiently large scale, i.e., networks that undergo a large number of timesteps. Consider a vertex V with fitness η , and assume that its degree k_V is a continuous function of time. Because during each unit of time m new edges are formed, we expect that

$$\frac{dk_V}{dt} \approx m \frac{\eta}{\sum_{j=1}^N \eta_j}.$$

For large enough N we can make the approximation

$$\sum_{j=1}^N \eta_j \approx N\bar{\eta} = (N_0 + t)\bar{\eta},$$

where $\bar{\eta}$ is the expected value of η . So

$$\frac{dk_V}{dt} = m \frac{\eta}{(N_0 + t)\bar{\eta}}.$$

Integration yields

$$k_V = \int \frac{m\eta}{(N_0 + t)\bar{\eta}} dt = \frac{m\eta}{\bar{\eta}} \ln(N_0 + t) + C.$$

Let t_0 be the time that this vertex was added to the network. Since $k_V = m$ at time $t = t_0$,

$$C = m - \frac{m\eta}{\bar{\eta}} \ln(N_0 + t_0)$$

$$k_V = m + \frac{m\eta}{\bar{\eta}} \ln \frac{N_0 + t}{N_0 + t_0}.$$

We can now calculate the cumulative distribution function (CDF) of k_V as

$$\begin{aligned} \mathbb{P}(k_V \leq k) &= \mathbb{P}\left(m + \frac{m\eta}{\bar{\eta}} \ln \frac{N_0 + t}{N_0 + t_0} \leq k\right) \\ &= \mathbb{P}\left(\ln \frac{N_0 + t}{N_0 + t_0} \leq \frac{\bar{\eta}(k - m)}{m\eta}\right) \\ &= \mathbb{P}\left(\frac{N_0 + t}{e^{\frac{\bar{\eta}(k - m)}{m\eta}}} - N_0 \leq t_0\right). \end{aligned}$$

There are $N_0 + t$ total vertices in the network, so for any particular τ , $1 \leq \tau \leq t$, the probability that $t_0 = \tau$ is $\frac{1}{N_0 + 1}$ and the probability that $t_0 = 0$ (the vertex is a starting vertex) is $\frac{N_0}{N_0 + t}$. Thus in the continuous analogue, $\mathbb{P}(t' \leq t_0) = \frac{N_0 + t'}{N_0 + t}$, so So

$$\begin{aligned} \mathbb{P}(k_V \leq k) &= \mathbb{P}\left(\frac{N_0 + t}{e^{\frac{\bar{\eta}(k - m)}{m\eta}}} - N_0 \leq t_0\right) \\ &= \frac{1}{N_0 + t} \left[t - \left(\frac{N_0 + t}{e^{\frac{\bar{\eta}(k - m)}{m\eta}}} - N_0\right) \right] \\ &= 1 - e^{-\frac{\bar{\eta}(m - k)}{m\eta}}. \end{aligned}$$

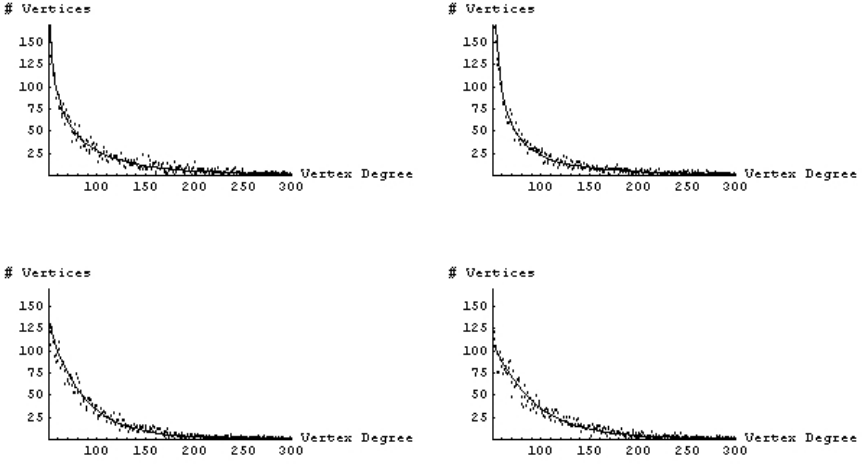


Figure 4.1: Predicted and simulated degree distributions. Solid lines represent predictions of the continuum theory and scatter plots represent simulated results for (a) uniform $\rho(\eta)$, $0 < \eta < 1$; (b) exponential $\rho(\eta) = e^{-\eta}$, $0 < \eta < \infty$; (c) power-law $\rho(\eta) \sim \eta^{-3}$, $1 < \eta < \infty$; (d) power-law $\rho(\eta) \sim \eta^{-4}$, $1 < \eta < \infty$.

We obtain the probability density function (PDF) of the vertex degree by differentiating the CDF with respect to k :

$$\frac{d}{dk} \mathbb{P}(k_V \leq k) = \frac{\bar{\eta}}{m\eta} e^{\frac{\bar{\eta}(m-k)}{m\eta}}.$$

This is the PDF for the degree of a vertex of fitness η , which we will denote as $P(k_\eta)$. To obtain the overall PDF, we take a weighted average of these fitness-based PDFs with the weights being the probabilities of having a fitness η . In other words,

$$P(k) = \int_{\eta_{\min}}^{\eta_{\max}} \rho(\eta) P(k_\eta) d\eta,$$

or

$$P(k) = \int_{\eta_{\min}}^{\eta_{\max}} \rho(\eta) \frac{\bar{\eta}}{m\eta} e^{\frac{\bar{\eta}(m-k)}{m\eta}} d\eta. \quad (4.1)$$

In this overall PDF, k is the continuous random variable for vertex degree, m is the constant for the average vertex degree, $\rho(\eta)$ is the PDF of fitnesses η , $\bar{\eta}$ is the expected value of η as determined by $\rho(\eta)$, and η_{\min} and η_{\max} are the bounds of the fitness values. It is important to note that this PDF does not depend on the present time t or the network size N . That is, as long as the size of the network is large enough so that the initial approximations are true, the PDF for the vertex degrees is constant over time as new vertices are added to the network.

We can scale equation (4.1) by multiplying by the total number of vertices N to predict the degree distribution of the network. To verify the predictions of the continuum theory, we numerically simulated this network construction algorithm for $m = 50$, $N = 5000$ and a variety of distributions $\rho(\eta)$ and calculated the degree distributions. The data from the simulations matches our theoretical result (Figure 4.1).

We also note that, as in previously developed fitness-based modifications of the BA model, $P(k)$ depends on the fitness distribution $\rho(\eta)$, and that for our model $P(k)$ is very versatile and varies greatly

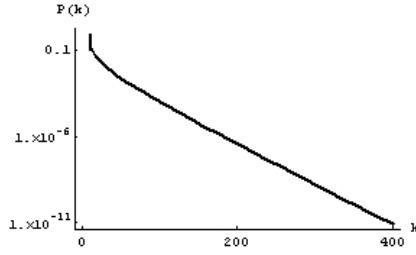


Figure 4.2: Semilog plot of $P(k)$ for uniform $\rho(\eta)$, $0 < \eta < 1$ and $m = 10$.

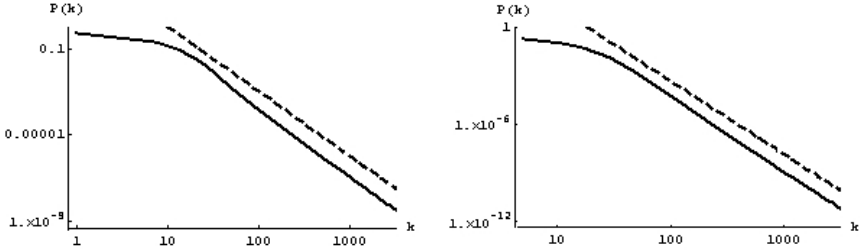


Figure 4.3: Log-log plot of $P(k)$ for $m = 10$ and (a) $\rho(\eta) \sim \eta^{-3}$, (b) $\rho(\eta) \sim \eta^{-4.5}$. Solid lines are plots of $P(k)$; dashed lines are plots of k^{-3} and $k^{-4.5}$ for (a) and (b) respectively.

with different fitness distributions. Evaluating this integral for a uniform fitness distribution over varying bounds and average vertex degree m results in varying exponential-tailed distributions for $P(k)$; one distribution is shown in Figure 4.2. Evaluating this integral for a power-law fitness distribution $\rho(\eta) \propto \eta^{-b}$ over varying m yields distributions of $P(k)$ with power-law tails of the same power $-b$; two such distributions are shown in Figure 4.3. Thus, with different power-law fitness distributions, we can obtain scale-free networks with degree distributions of various powers.

4.5 Path Length

Two other empirical properties observed in real-world networks are a small average path length between vertices and a high tendency for small clusters of highly connected nodes to form. We examined path length and clustering of networks produced by our model using computer simulations, and we draw comparisons both to empirical data and to results of the BA model. All data for path length and clustering coefficients are average values over 50 network constructions. We find through our simulations that our fitness-based algorithm does generate networks with small average path lengths. Using a power-law distribution with power $-b$, we find that for fixed values of N and m , the average path length of a network quickly increases to a low asymptotic limit as b increases (Figure 4.4a). Fixing m and b , we observe that the average path length increases logarithmically with N , a phenomenon also observed both in the original BA model and in random graphs (Figure 4.4b) [AB]. However, as in the BA model, the path lengths of our networks are of the same order of magnitude but consistently lower than those of real-world networks of the same size and average vertex degree, indicating that our algorithm may be overly effective, as compared to real-world processes, in bringing the vertices of the network closer together.

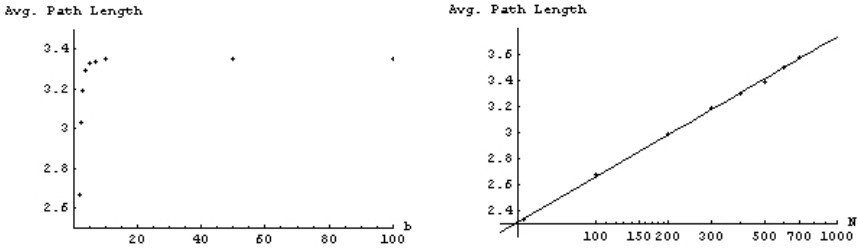


Figure 4.4: (a) Linear plot of path length versus b for $N = 300$ and $m = 3$. (b) Log-linear plot of path length versus N for $b = 3$ and $m = 3$. Solid line is the exponential regression curve.

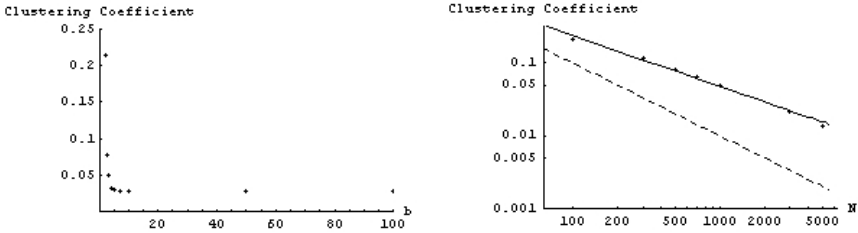


Figure 4.5: (a) Linear plot of clustering coefficient C versus b for $N = 1000$ and $m = 10$. (b) Log-log plot of C versus N for $b = 3$ and $m = 10$. Solid line corresponds to $N^{-0.7}$, and dashed line corresponds to C for a random graph with $m = 10$.

4.6 Clustering Coefficient

To quantify the concept of clustering, we use the clustering coefficient C developed by D. J. Watts [Wa]. C is the average of $\frac{2E_i}{k_i(k_i-1)}$ for all vertices i in the network, where k_i is the degree of vertex i and E_i is the number of edges in the subgraph of its k_i neighboring vertices. As in the case path length, if we fix the network size N and the average vertex degree m , then the clustering coefficient rapidly decreases to an asymptotic limit as b increases (Figure 4.5a). To obtain an idea of how large or small these clustering coefficients are, we fix m and b and compare the clustering coefficients of our networks to those of random graphs for different values of N (Figure 4.5b). We first note that the clustering coefficients of our networks are consistently higher than those of random graphs of the same size (whose clustering coefficients are given by $\frac{m}{N}$), and this difference increases with the size of the network. Secondly, C decreases with N as a power-law, as is observed for both random graphs and BA networks. Finally, the power of this relationship between C and N is -1 for random graphs, -0.75 for BA networks, and -0.70 for our fitness-based networks, while for real-world networks, this power is 0 and network size does not seem to affect the value of C [AB].

4.7 Conclusion

We have created a network model that parallels a simple and accepted existing model, the BA model, but that uses a “better-get-richer” instead of “richer-get-richer” growth algorithm. Our study indicates several important facts about our fitness-based network model. The first is that through a power-law fitness distribution, we can obtain scale-free networks. It may seem that having a power-law fitness

distribution is an arbitrary criterion, but in many real-world situations where individuals such as people or cities are ranked according to wealth or some other measure of “fitness,” these fitnesses fall under a power law distribution, as is stated in the empirical Zipf’s law [CCRM]. Thus, it may be a reasonable hypothesis that real-world networks have power-law fitness distributions. If this were true, our model would indicate that Zipf’s law and the ubiquitous nature of scale-free networks in the real world might be related phenomena. The varying powers of the degree distributions of real networks can be explained by varying powers of fitness distributions; the analysis of our model shows that these two powers are equal.

A second observation is that in our model, non-power-law distributions of fitness result in other network structures. Specifically, a uniform fitness distribution results in an exponential degree distribution. This may be related to certain real-world networks that are indeed not scale-free but follow such an exponential degree distribution. The Western United States power grid and the network of neurons in a human brain are notable instances of such exponential distributions [AB]. The structures of these two networks in particular are heavily influenced by the physical location of their vertices, and thus the vertex fitness values may be more indicative of the number of other vertices that physically surround them and thus may fall under a relatively more uniform probability distribution than the fitnesses of networks without this distance restriction.

A final observation is that our fitness-based networks with power-law fitness distributions very closely resemble the BA network, particularly with respect to how path length and clustering scale with network size. Along with a scale-free degree distribution, this is evidence that our models are very similar in structure to BA networks. Thus, we have shown that newly added vertices of a network do not need knowledge of the popularity of the current vertices in order to maintain a scale-free network structure, and that knowledge of the vertex popularity values (as in the BA model) does not alter three of the most significant structural properties. It should be noted, though, that this result is dependent on the hypothesis that fitness distributions are power laws.

Important work needs to be done in studying on a microscopic level the growth patterns of particular real-world networks to determine their underlying fitness distributions. Further work in this area can also be done by examining models with vertex fitnesses that vary over time, as well as by adding complications such as directed and weighted edges. Overall, we have shown that a fitness-based variation of the BA model can produce some of the important trends observed in the structure of real-world complex networks.

4.8 Acknowledgements

The bulk of this research was performed in the summer of 2005 under the mentorship of King Y. Yick, graduate student of mathematics at MIT. It was conducted as part of the Research Science Institute (RSI), sponsored by the Center for Excellence in Education. Staff of RSI 2005, in particular Dr. Jenny Sendova, contributed to the original drafting of this paper.

References

- [AB] Reka Albert and Albert-László Barabási: Statistical mechanics of complex networks. *Reviews of Modern Physics* **74** #1 (2002), 47–97. See also references therein.
- [BB1] Albert-László Barabási and Eric Bonabeau: Scale-free networks. *Scientific American* (May 2003), 50–59.
- [BB2] Ginestra Bianconi and Albert-László Barabási: Competition and multiscaling in evolving networks. *Europhysics Letters* **54** #4 (2001), 436–442.
- [CCRM] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz: Scale-free networks from varying vertex intrinsic fitness. *Physical Review Letters* **89** #25 (2002), 8702.
- [Ev] David Everitt: Generating random variables. Available at <http://www.it.usyd.edu.au/~deveritt/networksimulation/rv.pdf> (2005/08/01).

- [SC] Vito D. P. Servedio and Guido Caldarelli: Vertex intrinsic fitness: How to produce arbitrary scale-free networks. *Physical Review E* **70** (2004), 056126.
- [St] Steven H. Strogatz: Exploring complex networks. *Nature* **410** (2001), 268–276.
- [Wa] Duncan J. Watts: Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology* **105** #2 (1999), 493–527.