Chapter 1

FITNESS-BASED GENERATIVE MODELS FOR POWER-LAW NETWORKS

Khanh Nguyen, Duc A. Tran
Department of Computer Science
University of Massachusets, Boston
knguyen,duc@cs.umb.edu

Abstract Many real-world complex networks exhibit a power-law degree distribution. A dominant concept traditionally believed to underlie the emergence of this phenomenon is the mechanism of preferential attachment which originally states that in a growing network a node with higher degree is more likely to be connected by joining nodes. However, a line of research towards a naturally comprehensible explanation for the formation of power-law networks has argued that degree is not the only key factor influencing the network growth. Instead, it is conjectured that each node has a "fitness" representing its propensity to attract links. The concept of fitness is more general than degree; the former may be some factor that is not degree, or may be degree in combination with other factors. This chapter presents a survey of representative models for generating power-law networks, that belong to this approach.

Keywords: Power-law networks, scale-free networks, random networks, fitness

1. Introduction

The last decade has seen much interest in studying complex real-world networks and attempting to find theoretical models that elucidate their structure. Although empirical networks have been studied for some time, a surge in activity is often seen as having started with Watts and Strogatz's paper on "small world networks" [Watts and Strogatz, 1998]. More recently, the major focus of research has moved from small-world networks to "scale-free" networks, which are characterized by having power-law degree distributions [Barabási and Albert, 1999]; that is, if $p(k)$ is the fraction of nodes in the network having degree k (i.e. having

Table 1.1. Real-world networks: power-law exponent (λ)

Case	λ
Mathematics coauthorship	2.2
Film actor collaborations	2.1-2.3
WWW	2.1
Internet backbone	2.15-2.2
Protein interactions	2 - 2.4
ER graph	NA
BA graph	3

k connections to other nodes) then (for suitably large k)

$$p(k) = ck^{-\lambda} \tag{1.1}$$

where $c = (\lambda - 1)m^{\lambda-1}$ is a normalization factor and m is the minimal degree in the network. This distribution is observed in many real-world networks, including the WWW [Albert et al., 1999], the Internet [Faloutros et al., 1999], metabolic networks [Jeong et al., 2000], protein networks [Jeong et al., 2001], co-authorship networks [Neuman, 2001], and sexual contact networks [Liljeros et al., 2001]. In these networks, there are a few nodes with high degree and many other nodes with small degree, a property not found in standard Erdós-Rényi random graphs [Erdós and Rényi, 1959].

The near ubiquity of heavy-tailed degree distributions such as the power-law (1.1) for real-world complex networks, together with the inadequacy of the Erdós-Rényi random graphs as a theoretical model for such networks, brings into sharp relief the fundamental problem of obtaining a satisfactory theoretical explanation for how heavy-tailed degree distribution can naturally arise in complex networks.

A dominant concept traditionally believed to underlie the emergence of the power-law phenomenon is the mechanism of preferential attachment, proposed by Barabási and Albert [Barabási and Albert, 1999]: the higher degree a node has, the more likely it is to be connected by new nodes. This model, hereafter referred to as the BA model, leads to a growing random network which simulations and analytic arguments show has a power-law degree distribution with exponent $\lambda = 3$. Despite its elegance and simplicity, a deficiency of this mechanism is due to its fixed power-law exponent. As real-world networks exhibit a wide range of exponents, typically between 2 and 3 (see Table 1.1 for examples), the BA model may only explain a small subset of complex networks.

Consequently, other mechanisms have been proposed. Some, e.g., [Dangalchev, 2004], are merely formulaic without a natural interpretation. Others use different connectivity information, not merely degree, of each node to influence the formation of a network, such as the mechanism in [Kumar et al., 2000] and extensions of the BA model in [Bianconi and Barabsi, 2001; Bedogne and Rodgers, 2006; Kong et al., 2009]. Still, a universally accepted explanation that works for not just one network but also others remains to be found. For example, if we use the BA-based models to explain the sexual contact network studied in [Liljeros et al., 2001], which is known to be power-law, a new individual will prefer to have sexual contact with those individuals who already have a large number of sexual contacts, while the explanation according to [Kumar et al., 2000] will infer that a new individual will have sexual contact with some existing partners of a randomly chosen individual. These explanations seem bizarre for human sexual behavior.

In searching for a more natural explanation for the formation of power-law networks in the real world, there is a line of research, e.g., [Bianconi and Barabsi, 2001; Bedogne and Rodgers, 2006; Caldarelli et al., 2002; Vito et al., 2004; Ghadge et al., 2010], that is founded based on the conjecture that in many complex networks each node will have associated to it a "fitness" representing the propensity of the node to attract links. Using the fitness concept to explain the sexual contact network above, we could say that it is the fitness of an individual that attracts other individuals; an individual wants to have sexual contact with another individual because of the latter's fitness, not the latter's connectivity. The fact that a node has a high number of contacts may just be a consequence of its high fitness. The key challenge in the design of fitness-based models is how fitness is defined; for example, what is fitness? what is it made of? what is its influence? Fitness may be just degree, or something not, or a combination of many factors, explicit or implicit. In the following sections, three representative fitness-based models are described, which differ fundamentally in the approach to addressing these questions.

2. Fitness-based Model by [Bianconi and Barabsi, 2001]

An early fitness-based model for constructing power-law networks was proposed in [Bianconi and Barabsi, 2001]. This model assumes that the evolution of a network is driven by two factors associated with each node, its node degree and its fitness, which jointly determine the rate at which new links are added to the node. While node degree represents an ability to attract links that is increasing over time, node fitness represents something attractive about the node, that is constant. For example, in the WWW network, there can be two webpages published at the same time (thus, same degree) but later one might be much more popular than the other; this might be because of something intrinsic about one page, for example, its content, that makes it more attractive than the other page.

In the proposed model, each node i has a fitness Φ_i which is chosen according to some distribution $\rho(\Phi_i)$. The network construction algorithm generalizes the BA algorithm as follows:

- Parameters
  - n_0: the size of the initial network which can be any graph
  - $m \leq n_0$: the number of nodes a new node connects to when it joins the network
  - n: number of nodes in the final network
- Procedure
  1 Initially, start with the initial network of n_0 nodes, each assigned a random fitness according to distribution ρ

  2 Each time a new node is added,
    - Assign a random fitness to the new node according to distribution ρ
    - Add m edges linking the new node to m distinct existing nodes such that the probability Π_i for connecting to an existing node i is taken to be proportional to its fitness Φ_i:

$$\Pi_i = \frac{k_i \Phi_i}{\sum_j k_j \Phi_j} \tag{1.2}$$

  3 Stop after the n^{th} node is added

If every node has the same fitness, this model is identical to the orignal BA model, resulting in power-law networks of exponent $\lambda = 3$. In the case that fitness is chosen uniformly in the interval [0, 1], using the continutum theory one can infer that $p(k)$, the fraction of nodes in the network having degree k, follows a generalized power-law, with an inverse logarithmic correction.

$$p(k) \propto \frac{k^{-2.255}}{\ln k} \tag{1.3}$$

This prediction of the continuum theory has been confirmed by numerical simulations [Bianconi and Barabsi, 2001]. It is found that the degree of a node i over time, $k_i(t)$, follows a power-law for all fitness, and the scaling exponent for this power-law is $\Phi_i/1.255$, thus being larger for nodes with larger fitness. This allows a node with a higher fitness to enter the network late but still become more popular than nodes that have stayed in the network for a much longer period. A limitation of this model, however, is that it results in only power-law distributions with fixed exponents. Next, we will present a model that can generate a wider range of exponents.

3. Fitness-Based Model by [Ghadge et al., 2010]

In a citation network such as [Neuman, 2001] the different nodes (i.e. papers) will have different propensities to attract links (i.e. citations). The various factors that contribute to the likelihood of a paper being cited could include the prominence of the author(s), the importance of the journal in which it is published, the apparent scientific merit of the work, the timeliness of the ideas contained in the paper, etc. Moreover, it is plausible that the overall quantity that determines the propensity of a paper to be cited depends essentially multiplicatively on such various factors. The multiplicative nature is likely in this case since if one or two of the factors happen to be very small then the overall likelihood of a paper being cited is often also small, even when other factors are not small; e.g., an unknown author and an obscure journal were enough to bury a fundamentally important scientific paper.

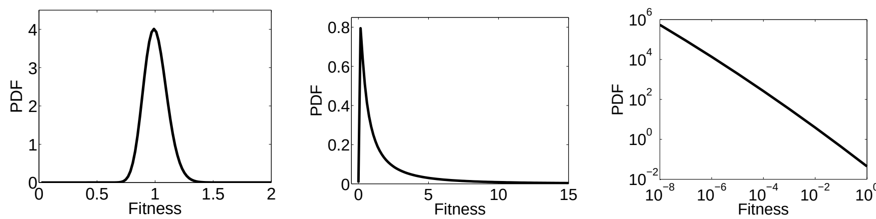
The lognormal fitness attachment model (LNFA), proposed in [Ghadge et al., 2010], was motivated by the observation above. In this model, the fitness Φ_i representing the property of each node i to attract links is formed multiplicatively from a number of factors $\{\phi_1,\ \phi_2,\ ...,\ \phi_L\}$ as follows:

$$\Phi_i = \prod_{l=1}^{L} \phi_l \tag{1.4}$$

where each factor ϕ_l is represented as a real non-negative value. Since there may be many factors contributing to the a node's attractiveness, explicit or implicit, we assume that the number of factors ϕ_i is reasonably large and that they are statistically independent. The fitness Φ_i will therefore be lognormally distributed, irrespective of the manner in which the individual factors are distributed. Indeed, we have

$$\ln \Phi_i = \sum_{l=1}^{L} \ln \phi_l \tag{1.5}$$

and the Central Limit Theorem implies that this sum will converge to a normal distribution. Therefore, $\ln \Phi_i$ will be normally distributed. Since a random variable X has a lognormal distribution if the random variable



(a) $\sigma = 0.1$ (b) $\sigma = 1.5$ (c) $\sigma = 9$

Figure 1.1. Lognormal fitness distribution for three representative values of σ. The extreme cases are when σ is small (0.1) or large (9). In most cases, the distribution will have the shape similar to the case $\sigma = 1.5$. Figure 1.1(c) is plotted in the log-log scale to emphasize that all the nodes except a few have fitness zero (or extremely close) while the rest (very few) have high fitness.

$Y = \ln X$ has a normal distribution, Φ_i will be lognormally distributed. The density function of the normal distribution is

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/(2\sigma^2)}, \tag{1.6}$$

where μ is the mean and σ is the standard derivation (i.e. σ^2 is the variance). The range of the normal distribution is $y \in (-\infty, \infty)$. It follows from the logarithmic relation $Y = \ln X$ that the density function of the lognormal distribution is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\ln x-\mu)^2/2\sigma^2}. \tag{1.7}$$

It is conventional to say that the lognormal distribution has parameters μ and σ when the associated normal distribution has mean μ and standard deviation σ. The range of the lognormal distribution is $x \in (0, \infty)$. The lognormal distribution is skewed with mean $e^{\mu+\sigma^2/2}$ and variance $(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$.

Thus, the basic hypothesis that each of the nodes has associated to it a fitness of the form (1.4) entails that under quite general conditions this fitness will be lognormally distributed. In the LNFA model, without loss of generality, one can assume that $\mu = 0$; hence, the fitness distribution is characterized by only a single parameter σ. This lognormal distribution has mean $e^{\sigma^2/2}$ and variance $(e^{\sigma^2} - 1)e^{\sigma^2}$; examples are shown in Figure 1.1.

The network construction algorithm works as follows:

- Parameters

- σ: parameter for the lognormal fitness distribution
- n_0: the size of the initial network which can be any graph
- $m \leq n_0$: the number of nodes a new node connects to when it joins the network
- n: number of nodes in the final network

- Procedure

1 Initially, start with the initial network of n_0 nodes, each assigned a random fitness according to the lognormal distribution

2 Each time a new node is added,

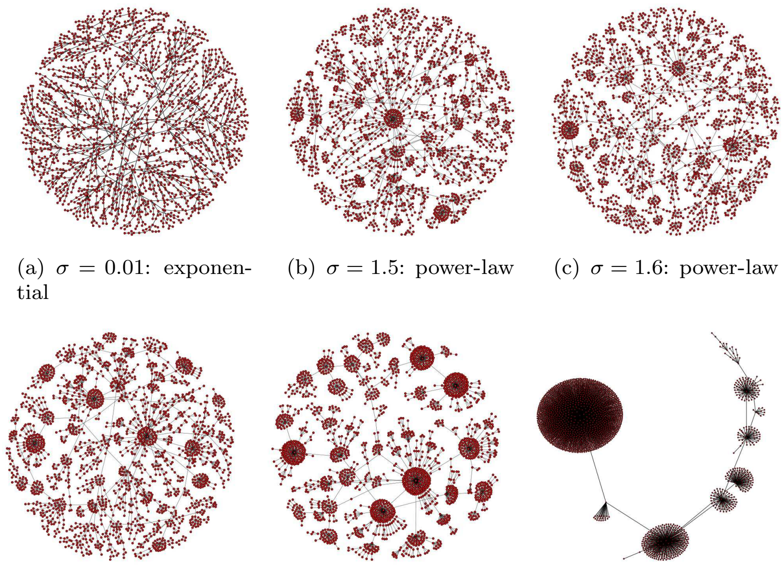
- Assign a random fitness to the new node according to the lognormal distribution
- Add m edges linking the new node to m distinct existing nodes such that the probability Π_i for connecting to an existing node i is taken to be proportional to its fitness Φ_i:

$$\Pi_i = \frac{\Phi_i}{\sum_j \Phi_j} \tag{1.8}$$

3 Stop after the n^{th} node is added

LNFA is almost identical to the BA model, the difference being that fitness information is used in place of degree information. Although this difference seems to be minor, it makes a fundamental shift in how the network is formed. To make this point, recall that in the BA protocol, the degree of a new node at the time it joins the network is small (m) and so it takes this node a long time before it may become a preferential choice for future new nodes to attach to. In LNFA, the new node may have a large fitness at the time it joins the network, making itself a preferential choice immediately. This is naturally reasonable because the attractiveness of a node may not result from how many nodes it is connected to; it may instead result from the “inner self” factors such as the personality of a person in a friendship network and his or her age.

As demonstrated in Figure 1.2, LNFA can be used to generate a large spectrum of networks as seen in the real world and it is possible to do so by varying the parameter σ. Consider two extreme cases of this parameter. In the first case, if σ is zero, nodes have exact same fitnesses and so, basing on Formula 1.8, each time a new node joins the network it chooses an existing node as neighbor with equal chance. This construction is simply the random graph model of [Callaway et al., 2001] which



(a) $\sigma = 0.01$: exponential (b) $\sigma = 1.5$: power-law (c) $\sigma = 1.6$: power-law

(d) $\sigma = 3.0$: power-law (e) $\sigma = 4.0$: power-law (f) $\sigma = 9$: winner-take-all

Figure 1.2. 2000-node networks resulting from increasing σ. Transitions from exponential to power-law to winner-take-all graphs are observed

yields a network with exponential degree distribution. On the other extreme, if σ is increased to reach a certain threshold, few nodes will stand out having very large fitnesses while all the other nodes will have very low fitnesses (zero or near zero; see Figure 1.1(c)). Consequently, an extremely high number of connections will be made to just a single node, resulting in a monopolistic network; this "winner-take-all" degree pattern is also observed in the real world [Barabási, 2001]. Between these two extreme cases (exponential and monopolistic) we find a spectrum of scale-free networks.

4. Fitness-based Model by [Caldarelli et al., 2002]

It is argued in that in many cases of interest, the power-law degree behavior is neither related to dynamical properties nor to preferential attachment. Also, the concept of having the likelihood that a new node creates a link to an existing node depend solely on the latter's fitness might be applicable only for certain networks. The model proposed in [Caldarelli et al., 2002] is suited for complex networks where it is the mutual benefit that makes two nodes link to each other. This model puts the emphasis on fitness itself without using the preferential attachment rule. Further, it is a static model building a network by growing links instead of growing nodes.

Specifically, the network construction algorithm starts with a set of n isolated nodes, where n is the size of the network to be built. Similar to the models discussed in the previous sections, each node i has a fitness Φ_i drawn from some distribution ρ. Then, for every pair of nodes, i and j, a link is drawn with a probability $f(\Phi_i, \Phi_j)$ which is some joint function of Φ_i and Φ_j. This model can be considered a generalization of the Erdós-Rényi (ER) random graph model [Erdós and Rényi, 1959]. Rather than using an identical link probability for every pair of nodes as in the ER model, here two nodes are linked with a likelihood depending jointly on their fitnesses. A general expression for $p(k)$ can be easily derived. Indeed, the mean degree of a node of fitness η is

$$k(\Phi) = n \int_0^\infty f(\Phi, x)\rho(x)\, dx = nF(\Phi) \tag{1.9}$$

Assuming F to be a monotonous function, and for large enough n, the following form can be obtained for $p(k)$:

$$p(k) = \rho\left[F^{-1}\left(\frac{k}{n}\right)\right] \frac{d}{dk} F^{-1}\left(\frac{k}{n}\right) \tag{1.10}$$

Thus, one can choose an appropriate formula for ρ and f to achieve a given distribution for $p(k)$. It is shown that a power law for $p(k)$ will emerge if fitness follows a power law and the linking probability for two nodes is proportional to the product of their fitnesses. For example, one can choose $f(\Phi_i, \Phi_j) = (\Phi_i \Phi_j)/\Phi_M^2$, where Φ_M is the maximal fitness in the network, and $\rho(\Phi) \propto \Phi^{-\beta}$ (corresponding to a Zipf's behavior with Zipf coefficient $\alpha = 1/(\beta - 1)$).

In the case that fitness does not follow a power law distribution, it is possible to find a linking function that will result in a power-law degree distribution. For example, considering an exponential fitness distribution, $\rho(\Phi) \propto e^{-\Phi}$ (representing a Poisson distribution), one can choose $f(\Phi_i, \Phi_j) = \theta[\Phi_i + \Phi_j - z(n)]$, where θ is the usual Heaviside step function and $z(n)$ is some threshold, meaning that two nodes are neighbors only if the sum of their fitness values is larger than the threshold $z(n)$. Using these rules, the degree distribution has a power law with exponent $\lambda = 2$. This is interesting because it shows that power-law networks can emerge even if fitness is not power-law. The same behavior also emerges if a more generic form of the linking function is used, $f(\Phi_i, \Phi_j) = \theta[\Phi_i^c + \Phi_j^c - z^c(n)]$ (where c is an integer number); however, the power-law degree distribution has logarithmic corrections in some cases.

Closely related to the above model is the work of [Vito et al., 2004] which also assumes the same concept that the linking probability is a joint function of the fitnesses of the end nodes. In this related work, it is concluded that for any given fitness distribution $\rho(\Phi)$ there exists a function $g(\Phi)$ such that the network generated by $\rho(\Phi)$ and $f(\Phi_i, \Phi_j) = g(\Phi_i)g(\Phi_j)$ is power-law with an arbitrary real exponent.

5. Summary

This chapter has reviewed representative models for constructing power-law complex networks, that are inspired by the idea that there is some intrinsic fitness associated with a node to drive its evolution in the network. This fitness might be causal to why a node has a high degree or a low degree, or it can be an independent factor which together with the node's degree affect the node's ability to compete for links. The models discussed differ in how they interpret fitness and its influence on growing the network. [Bianconi and Barabsi, 2001] argues for its model that both degree and fitness jointly determine the growth rate of node degree. This model may apply to complex networks such as a social network where a person's attractiveness is a combination of both his or her experience (represented by node degree) and talent (represented by fitness), or the WWW network where a webpage is popular because its long time staying online (represented by node degree) and quality of its content (represented by fitness). On the other hand, [Caldarelli et al., 2002; Vito et al., 2004; Ghadge et al., 2010] propose that all the attractiveness factors associated with a node can be combined into a single factor (fitness). While the models in [Caldarelli et al., 2002; Vito et al., 2004] are motivated by static networks where two nodes require mutual benefit in order to make a connection, the lognormal fitness model in [Ghadge et al., 2010] is suitable to explain growing networks in which a node wants to be a neighbor of another solely because of the latter's fitness, regardless of the former's fitness. Although none of the these models is one-size-fits-all, they do represent a vast population of complex networks. The current models, however, assume that fitness is an intrinsic factor that does not change over time. In practice, there are cases of networks where the overall attractiveness of a node might increase for one period of time and decrease for another. It is thus an interesting research problem to explore fitness models that allow fitness to have its own evolution. The future research should also pay great attention to (in)validating theoretical models with the data collected from real-world networks.

Acknowledgments

The authors wish to thank their UMass Boston colleagues, Shilpa Ghadge, Bala Sundaram, and Timothy Killingback, for interesting discussions regarding the lognormal fitness model presented in this chapter.

References

Albert, R., Jeong, H., and Barabási, A. L. (1999). Diameter of the world-wide-web. *Nature*, 400:107–110.

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

Barabási, A.L. (2001). The physics of the web. *Physics World*, 14-7:33–38.

Bedogne, C. and Rodgers, G. J. (2006). Complex growing networks with intrinsic vertex fitness. *Phys. Rev. E*, 74:046115.

Bianconi, G. and Barabsi, A. L. (2001). Competition and multiscaling in evolving networks. *Europhysics Letters*, 54:436–442.

Caldarelli, G., Capocci, A., Rios, P. De Los, and Munoz, M.A. (2002). Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.*, 89-25:258702.

Callaway, Duncan S., Hopcroft, John E., Kleinberg, Jon M., Newman, M. E. J., and Strogatz, Steven H. (2001). Are randomly grown graphs really random? *Physical Review E*, 64:041902.

Dangalchev, Ch. (2004). Generation models for scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 338:659–671.

Erdós, P. and Rényi, A. (1959). On random graphs. *Publicationes Mathematicae*, 6:290–297.

Faloutros, M., P.Faloutros, and Faloutros, C. (1999). On power-law relationship of the internet topology. *ACM SIGCOM 99, Comp. Comm. Rev.*, 29:251–260.

Ghadge, S., Killingback, T., Sundaram, B., and Tran, Duc A. (2010). A statistical construction of power-law networks. *International Journal on Parallel, Emergent, and Distributed Systems*, 25:223–235.

Jeong, H., Mason, S., Oltvai, R., and Barabási, A.L. (2001). Lethality and centrality in protein networks. *Nature*, 411:41.

Jeong, H., Tombor, B., Albert, R., Oltvai, N., and Barabási, A.L. (2000). The large-scale organization of metabolic networks. *Nature*, 607:651.

Kong, Joseph S., Sarshar, Nima, and Roychowdhury, Vwani P. (2009). Experience versus talent shapes the structure of the web. *CoRR*, abs/0901.0296.

Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., and Upfal, E. (2000). Stochastic models for the web graph. In *FOCS '00: Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, page 57, Washington, DC, USA. IEEE Computer Society.

Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., and Aberg, Y. (2001). The web of human sexual contacts. *Nature*, 411:907.

Neuman, M. (2001). The structure of scientific collaboration networks. *Proc. Nat. Acad. Sci. USA*, 98:404–409.

Vito, D.P.S., Caldarelli, G., and Butta, P. (2004). Vertex intrinsic fitness: How to produce arbitrary scale-free networks. *Phys. Rev. E*, 70:056126.

Watts, D. and Strogatz, S. (1998). Collective dynamics of "small-world" networks. *Nature*, 393:440–442.