Fitness-Weighted Preferential Attachment with Varying Number of New Connections

Juan Romero(✉), Jorge Finke, and Andrés Salazar

Pontificia Universidad Javeriana Cali, Santiago de Cali, Colombia
j.romero@javerianacali.edu.co

Abstract. Preferential attachment models are used to explain the emergence of power laws in the degree distributions of networks. These models assume that a new node attaches to a network by establishing edges to a fixed number of nodes. Nonetheless, for many empirical networks the number of new edges varies as more nodes become part of the network. This paper extends the linear preferential attachment model by considering that the number of new edges is characterized by a random variable that obeys a power law probability function. While most new nodes connect to a few nodes, some nodes connect to a larger number. We characterize the dynamics of growth of the degrees of the nodes and the degree distribution of the network.

Keywords: Preferential attachment · Node fitness · Harmonic number · Riemann zeta function

1 Introduction

The power law degree distribution of a number of networks can be explained by the principle of preferential attachment [3]. Linear preferential attachment, in particular, assumes that (i) new nodes attach to a network by establishing a fixed number of edges; and (ii) the probability of connecting to a target node is proportional to the degree of that node [1]. The model introduced in [1] generates networks in which the nodes that have been part of the network the longest gain over time the highest number of edges. However, the growth of the number of edges may not necessarily depend on the extent to which nodes have been part of the network. Nodes that recently joined network may quickly acquire a larger number of connections (e.g., in the context of citation networks, papers written by prestigious authors may earn a large number of citations in a short time). To allow new nodes to quickly become highly connected, the work in [2] introduces the concept of *node fitness* which considers that nodes have different abilities to compete for new edges. The fitness model assigns to each node a quality of attracting new edges and assumes that this fitness value does not varying over time.

H. Cherifi et al. (Eds.): COMPLEX NETWORKS 2019, SCI 881, pp. 612–620, 2020.
https://doi.org/10.1007/978-3-030-36687-2_51

This paper introduces a fitness-weighted preferential attachment mechanism. As in [2], each node has a fitness that remains unchanged over time. Unlike [2], the number of new edges established by the new node depends on the number nodes that are part of the network. Specifically, the probability that a new node connects to m nodes is proportional to m^{-s} for $s > 0$. While most new nodes connect to a few nodes, some new nodes connect to a larger number of nodes.

The contributions of our work are twofold. First, we describe the evolution of the degree of the nodes when the probability of connecting to a node obeys fitness-weighted preferential attachment. Second, we prove that the probability distribution of the degree of the nodes converges.

The remaining sections are organized as follows. Section 1 presents the formation mechanisms of the proposed model and derives the asymptotic behavior of the expected number of new edges. Section 2 characterizes the degree dynamics of a node. Sections 3 and 4 derive the degree distribution over all nodes and analyzes the asymptotic behavior of the expected value of the average network degree. Finally, Sect. 6 draws some conclusions and future research directions.

2 Attachment with Power Law Growth

Consider an undirected graph $G_t(V_t, E_t)$ with a set of nodes V_t and a set of edges E_t. For $t = 0$, let $G_0(V_0, E_0)$ be the initial graph with $|V_0| = n_0$ and $|E_0| = l_0$. Let $g(t, i)$ describe the degree of node i at time t. Moreover, let M_t be a random variable that describes the number of edges established by a new node attaching to the network. The evolution of G_t is based on the following mechanisms:

(i) Node growth: For each t, a new node is added to the set of nodes V_{t-1}.
(ii) Edge growth: For each t, M_t follows a probability function

$$f_t(m) = P[M_t = m] = C(t)m^{-s}, \tag{1}$$

where $s \in \mathbb{R}^+$ represents the scaling coefficient and $C(t)$ represents the proportionality constant of the distribution at time t.
(iii) Preferential attachment with fitness: The new node connects to node $i \in V_{t-1}$ with probability

$$\pi(i) = \frac{\eta_i g(t-1, i)}{\sum\limits_{j \in V_{t-1}} \eta_j g(t-1, j)} \tag{2}$$

where η follows a distribution $\rho(\eta)$ and each η_i represents fitness parameter of node i. The value of η remains fixed over time.

Figure 1 illustrates the dynamics of a network based on the above mechanisms. According to mechanism (i), the set of nodes grows by the continuous addition of a node at every time step, so $n(t) = |V_t| = n_0 + t$. According to mechanism (ii), the growth of the set of edges follows a power law. The support of the probability function in (1) is V_{t-1}, so all instances of M_t are less than or

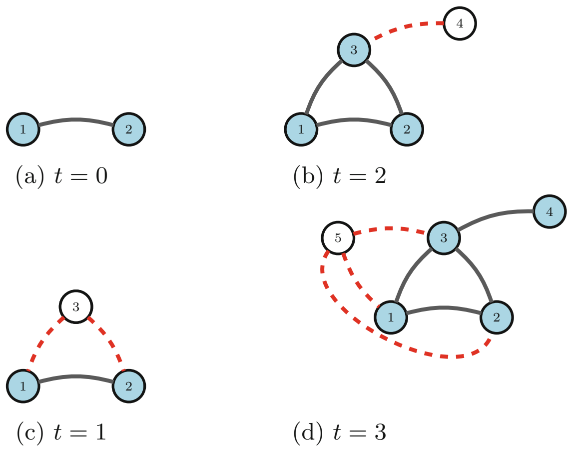

Fig. 1. Network evolution when the number of new edges (depicted by the dashed edges) follows a power law probability function.

equal to $n(t-1)$. Moreover, at any time t, the proportionality constant of $f_t(m)$ satisfies

$$1 = \sum_{m=1}^{n(t)-1} C(t)m^{-s} = C(t) \sum_{m=1}^{n(t)-1} m^{-s} = C(t)H_t(s), \tag{3}$$

where

$$H_t(s) = \sum_{m=1}^{n(t)-1} m^{-s} \tag{4}$$

represents the generalized harmonic number of order s. Note that $C(t) = 1/H_t(s)$. Note also that $H_t(s)$ depends of the initial number of nodes n_0. As t tends to infinity, $H_t(s)$ exists if $s > 2$, in which case the limit of $H_t(s)$ is represented by the Zeta-Riemann function.

$$\lim_{t\to\infty} H_t(s) = \sum_{m=1}^{\infty} m^{-s} = \zeta(s), \tag{5}$$

The cumulative distribution function of M_t is given by

$$F_m = P[M_t \leqslant m] = C(t)H_m(s),$$

where $1 \leqslant m \leqslant n(t) - 1$.

At any time t, the expected number of edges that a new node establishes is

$$\theta(t) = E[M_t] = \frac{1}{H_t(s)} \sum_{m=1}^{n(t)-1} m^{-(s-1)} = \frac{H_t(s-1)}{H_t(s)}, \tag{6}$$

where $\theta(0) = 0$. Note also that $E[M_t^2] = \frac{H_t(s-2)}{H_t(s)}$, which implies that

$$\sigma(t) = \mathrm{Var}[M_t] = E[M_t^2] - E[M_t]^2 = \frac{H_t(s)H_t(s-2) - H_t(s-1)^2}{H_t(s)^2} \tag{7}$$

According to lemma 1 in [3], $\theta(t)$ and $\sigma(t)$ satisfies

(a) $\lim_{t\to\infty} \theta(t) = \frac{\zeta(s-1)}{\zeta(s)}$
(b) $\lim_{t\to\infty} \sigma(t) = \frac{\zeta(s)\zeta(s-2) - \zeta(s-1)^2}{\zeta(s)^2}$; and
(c) $\theta(t)$ is a strictly increasing function.

To characterize the average of all instances of $\theta(t)$ in the limit of t, according to lemma 2 in [3], if $s > 2$, the asymptotic behavior of the average of the instances of $\theta(t)$ satisfies

$$\lim_{t\to\infty} \hat{\theta}_t = \frac{\zeta(s-1)}{\zeta(s)}, \tag{8}$$

where $\hat{\theta}_t = \sum_{i=1}^{t} \frac{\theta(i)}{t}$.

3 Degree Dynamics

We want to derive a functional form of the evolution of the degrees of the nodes as the network grows. Consider the following assumption.

A1: The number of nodes of the network grows at a constant rate.

As a direct consequence of A1, the rate of change of the degree of any node is proportional to the probability that a new node establishes an edge to that node. That is, for node i

$$\frac{dg(t,i)}{dt} = \theta(t)\,\pi(i) \tag{9}$$

$$= \frac{\theta(t)\,\eta_i g(t,i)}{\sum_{j\in V_t} \eta_j g(t,j)}. \tag{10}$$

Note that if $\rho(\eta) = \delta(\eta - 1)$, i.e., all fitness are equal and (9) reduces to the scale-free model with varying number of new edges [3]. For that particular case, $g(t,i) \approx \theta(t)\left(\frac{t}{t_i}\right)^{1/2}$ where t_i denotes the time at which node i attaches to the network. To solve (9), consider an evolution of g that follows a power law, with exponent β. That is

$$g(t,i) \approx \theta(t)\left(\frac{t}{t_i}\right)^{\beta(\eta_i)}. \tag{11}$$

Note that the exponent β depends on η_i, which represents multiscaling in the dynamics of the degree of the nodes. Since nodes can only increase the number of edges over time and at most one edge is established to any particular node at any time t, the value of β is bounded [2]. Moreover, since a node can only increase the number of edges over time, $0 < \beta(\eta) < 1$.

To specify $\beta(\eta)$, we need to derive the expected value of $\sum_j \eta_j g(t,j)$. Since a new node attaches the network at $t = t_i$, the sum of the degree-weighted over all nodes of the network is

$$\left\langle \sum_j \eta_j g(t,j) \right\rangle = \int \int_1^t \eta\rho(\eta) g(t,i) dt_i d\eta \tag{12}$$

$$= \int \eta\rho(\eta) d\eta \int_1^t \theta(t) \left(\frac{t}{t_i}\right)^{\beta(\eta_i)} dt_i \tag{13}$$

$$= \int \eta\rho(\eta) d\eta \theta(t) \frac{t - t^{\beta(\eta)}}{1 - \beta(\eta)}. \tag{14}$$

Since $\beta(\eta) < 1$, in the large t limit, $t^{\beta(\eta)}$ is negligible compared to t. Therefore

$$\left\langle \sum_j \eta_j g(t,j) \right\rangle \approx \xi\theta(t)t, \tag{15}$$

where

$$\xi = \int \frac{\eta\rho(\eta)}{1 - \beta(\eta)} d\eta. \tag{16}$$

According to (15), we can write (9) as

$$\frac{dg(t,i)}{dt} \approx \frac{\theta(t)\, \eta_i g(t,i)}{\xi\theta(t)t}, \tag{17}$$

In the large t limit, we obtain

$$\frac{dg(t,i)}{dt} \approx \frac{\theta^*\, \eta_i g(t,i)}{\xi\theta^* t} \tag{18}$$

$$\approx \frac{\eta_i g(t,i)}{\xi t}, \tag{19}$$

where $\lim_{t\to\infty} \theta(t) = \theta^*$. Equation 18 has a solution of the form defined in (11), if

$$\beta(\eta) = \frac{\eta}{\xi}, \tag{20}$$

To determine the value of ξ, note that substituting (20) in (16) we obtain

$$1 = \int_0^{\eta_{\max}} \frac{1}{\frac{\xi}{\eta} - 1} \rho(\eta) d\eta. \tag{21}$$

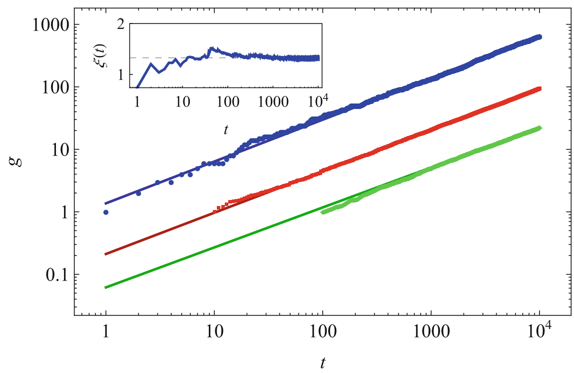

Fig. 2. Degree dynamics for nodes that attach to the network at $t = 10$, $t = 100$ and $t = 1000$ when η follows a uniform distribution. The solid lines represents the theoretical predictions. The curves represent averages over 100 runs. Inset: Asymptotic convergence of $\xi(t) = \sum_{j=1}^{t} \eta_j g(t,j)/t$.

It is of interest to consider the behavior of the network when nodes with different fitness compete for new edges. In particular, when η is chosen uniformly from the interval $[0, 1]$. Using (21) the constant ξ is defined as

$$\exp(-2/\xi) = 1 - \frac{1}{\xi} \tag{22}$$

which has solution $\xi^* = 1.255$. Figure 2 shows the behavior of $g(t, i)$ for a simulated network with $s = 4$.

4 Degree Distribution

Next, we calculate the degree distribution $Q_t(k)$, that is, the probability that a node has degree k at time t, when $t \to \infty$ we denote by $Q_\infty(k) = Q(k)$. [4] shows how to find the functional form of $Q(k)$ for fixed t, based on [4] we characterizes the asymptotic behavior of $Q_t(k)$. Let ℓ be the minimum degree of the network at time t, the probability that a node has degree $k = \ell$ at time t is

$$Q_t(\ell) = \frac{C(t)\ell^{-s}\xi(t)t}{\xi(t)t + n(t)\bar{\eta}_\ell \ell} \tag{23}$$

where $\bar{\eta_\ell} = \sum_{j \in A_\ell} \frac{\eta_j}{|A_\ell|}$. If $k > \ell$ the probability that a node has degree k at time t is

$$Q_t(k) = \frac{C(t)k^{-s}\xi(t)t + n(t)\bar{\eta}_{k-1}(k-1)Q_t(k-1)}{\xi(t)t + n(t)\bar{\eta}_k k} \tag{24}$$

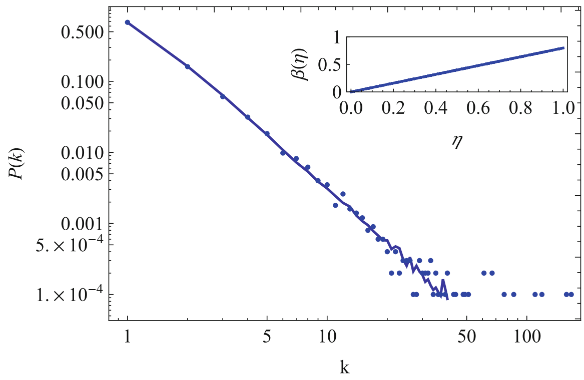

Fig. 3. Degree distribution for a simulated network (dotted curve) and the theoretical prediction (solid curve). Inset: The dependence of $\beta(\eta)$ on the fitness parameter η when $\rho(\eta)$ follows a uniform distribution in $[0,1]$.

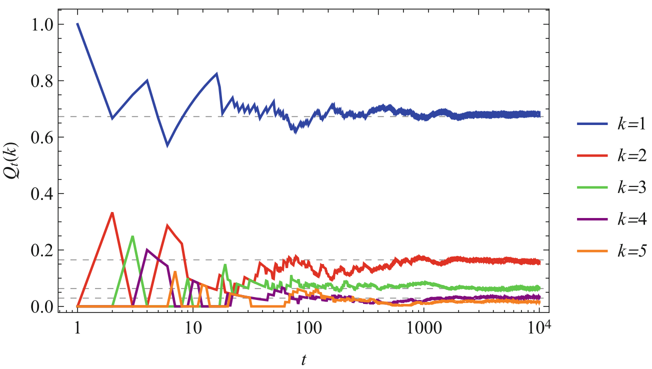

Fig. 4. Asymptotic degree distribution for a simulated network and the theoretical prediction for $k = 1, 2, 3, 4$.

Therefore, the asymptotic degree distribution, if $k = \ell$, is

$$Q(\ell) = \frac{C^* \ell^{-s} \xi^*}{\xi^* + \bar{\eta}_\ell \ell}, \tag{25}$$

and if $k > \ell$

$$Q(k) = \frac{C^* k^{-s} \xi^* + \bar{\eta}_{k-1}(k-1)Q(k-1)}{\xi^* + \bar{\eta}_k k}. \tag{26}$$

Figures 3 and 4 show the degree distribution for the mechanism with $s = 4$ and the asymptotic behavior of the degree distribution.

We now turn our attention to understanding the behavior of the expected average network degree.

5 Expected Average Degree

The expected degree of a node, selected uniformly at random at time t, is equal to the expected average degree of the network at time t. This section characterizes the asymptotic behavior of the expected average degree of the network. Let N_t denote the total degree at time t. The expected value of N_t is

$$e(t) = E[N_t] = 2\left(l_0 + \sum_{i=0}^{t} \theta(t)\right). \tag{27}$$

Furthermore, let D_t denote a random variable that describes the average degree of the network. For $t > 0$ the expected value of D_t is

$$d(t) = E[D_t] = \frac{e(t)}{n(t)} = \frac{2l_0 + 2\sum_{i=0}^{t} \theta(i)}{n_0 + t}. \tag{28}$$

Since $\theta(0) = 0$, note that $d(0) = \frac{2l_0}{n_0}$. The following theorem characterizes the asymptotic behavior of $d(t)$.

Theorem 1. *The asymptotic behavior of expected average degree $d(t)$ converges to*

$$\lim_{t\to\infty} d(t) = 2\frac{\zeta(s-1)}{\zeta(s)} \tag{29}$$

Proof. Since

$$d(t) = \frac{2l_0 + 2\sum_{i=0}^{t} \theta(i)}{n_0 + t},$$

note that

$$\lim_{t\to\infty} d(t) = 2\lim_{t\to\infty} \frac{l_0}{n_0 + t} + 2\lim_{t\to\infty} \frac{1}{n_0 + t}\sum_{i=0}^{t} \theta(i). \tag{30}$$

Therefore, applying lemma 2 in [3] to (30), we obtain (29).

Figure 5 illustrates the asymptotic behavior of $d(t)$.

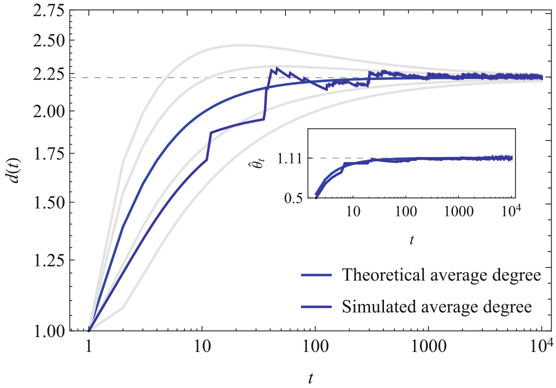

Fig. 5. Asymptotic behavior of the average degree $d(t)$ and simulated average degree with $s = 4$. Inset: Asymptotic behavior of $\hat{\theta}_t$ and simulated average of all instances of M_t.

6 Conclusions

This paper introduces a novel model that relaxes the original assumption of the Barabási-Albert model on how new edges are established. We characterize the dynamics of the growth of the degrees of the nodes and derive the asymptotic behavior of the resulting cumulative distribution. The resulting distribution approaches a stationary distribution if and only if the scaling exponent of the distribution of new edges is strictly greater than two. Finally, we show that the expected value of the average degree converges to an equilibrium. Understanding how other probability functions, of the number of edges established by the new nodes, impact the evolution of the network remains a future research direction.

Acknowledgments. This research was supported by the Center of Excellence and Appropriation in Big Data and Data Analytics (CAOBA), founded by the Ministry of Information Technologies and Telecommunications of Colombia (MinTIC) and the Colombian Administrative Department of Science, Technology and Innovation (COLCIENCIAS) under grant no. FP44842-anex46-2015.

References

1. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
2. Bianconi, G., Barabási, A.L.: Competition and multiscaling in evolving networks. EPL (Europhys. Lett.) **54**(4), 436 (2001)
3. Romero, J., Salazar, A., Finke, J.: Preferential attachment with power law growth in the number of new edges. In: 2017 IEEE 56th Annual Conference on Decision and Control (CDC), pp. 2680–2685. IEEE (2017)
4. Zadorozhnyi, V., Yudin, E.: Growing network: models following nonlinear preferential attachment rule. Phys. A **428**, 111–132 (2015)