2017 IEEE 56th Annual Conference on Decision and Control (CDC)
December 12-15, 2017, Melbourne, Australia

Preferential attachment with power law growth in the number of new edges

Juan Romero[1], Andrés Salazar[1] and Jorge Finke[2]

Abstract— **The Barabasi-Albert model is used to explain the formation of power laws in the degree distributions of networks. The model assumes that the principle of preferential attachment underlies the growth of networks, that is, new nodes connects to a fixed number of nodes with a probability that is proportional to their degrees. Yet, for empirical networks the number of new edges is often not constant, but varies as more nodes become part of the network. This paper considers an extension to the original Barabasi-Albert model, in which the number of edges established by a new node follows a power law distribution with support in the total number of nodes. While most new nodes connect to a few nodes, some new nodes connect to a larger number. We first characterize the dynamics of growth of the degree of the nodes. Second, we identify sufficient conditions under which the expected value of the average degree of the network is asymptotically stable. Finally, we show how the dynamics of the model resemble the evolution of protein interaction networks, Twitter, and Facebook.**

Index Terms— **Preferential attachment, Harmonic number, Riemann Zeta function, Lyapunov stability.**

I. INTRODUCTION

The behavior of the degree distribution of a number of empirical networks can be explained by the principle of preferential attachment. Linear preferential attachment is the basis for the formation mechanisms underlying the Barabasi-Albert model [1], which assumes that new nodes attach to a network by establishing a fixed number of new edges. Moreover, the probability of connecting to a particular node is directly proportional to the degree of that node. Past work has focused on extending linear attachment to other linkage mechanims, including sublinear and superlinear attachment [6], [5]. All of these models, however, operate under the assumption that as the network grows, the number of new edges remains constant.

A more realistic scenario would consider that the rate of new edges may vary over time. The work in [3] introduces a model, in which the number of new edges follows a distribution with support on a bounded set. Similarly, the work in [4] proposes a model and characterizes the degree distribution when the number of the new edges follows a Poisson distribution. This work introduces a model in which the probability that a new node connects to m nodes is proportional to m^{-s} for $s>0$. The proposed mechanism assumes that the number of new edges established at any time step depends on the set nodes that make up the network.

Our theoretical contributions are twofold. First, we identify conditions under which the probability distribution of the degree of the nodes converges. In particular, we show that the distribution converges if and only if $s>2$. Second, we characterize the stability properties of the expected average degree of the network.

The remaining sections are organized as follows. Section 1 presents the formation mechanisms of the model and derives the asymptotic behavior of the expected number of new edges. Section 2 characterizes the degree dynamics of a node. Section 3 characterizes the degree distribution. Section 4 analyzes the asymptotic behavior of the expected value of the average degree. Section 5 provides sufficient conditions that guarantee the stability of the average degree. Section 6 draws some conclusions and future research directions.

II. ATTACHMENT WITH POWER LAW GROWTH

Consider a series $\{G_t\}_{t\in\mathbb{N}}$, where each undirected graph $G_t(V_t,E_t)$ consist of a set of nodes V_t and 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 when attaching to the network. The evolution of G_t is based on the following mechanisms:

(i) *Growth:* A new node is added to the set of nodes V_{t-1}.

(ii) *New edges:* For $s\in\mathbb{R}^+$, M_t follows a probability function

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

where $C(t)$ represents the proportionality constant of the distribution of new edges at time t.

(iii) *Preferential attachment:* For $\alpha\in\mathbb{R}^+$, the new node connects to node $i\in V_{t-1}$ with probability

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

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

[1]Department of Natural Science and Mathematics,
j.romero@javerianacali.edu.co,
andresmsalazar@javerianacali.edu.co,
[2]Department of Electrical Engineering and Computer Science,
jfinke@javerianacali.edu.co,
Pontificia Universidad Javeriana Cali, Colombia.

978-1-5090-2872-6/17/$31.00 ©2017 IEEE

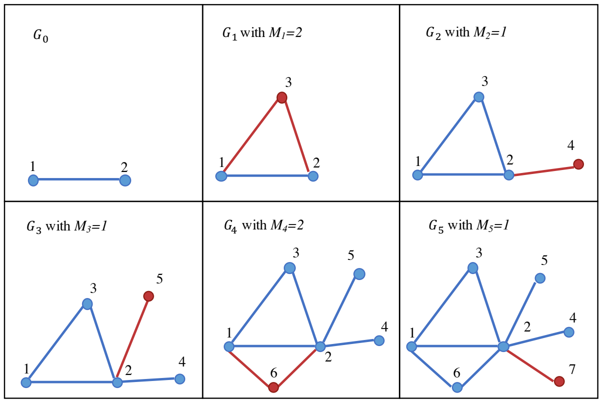

Fig. 1. Evolution of the network based on mechanisms $(i)-(iii)$.

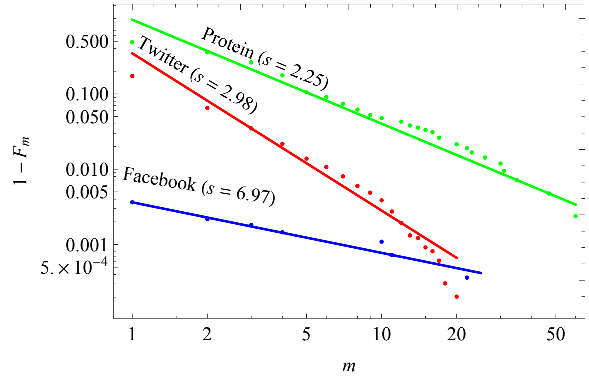

Fig. 2. Cumulative distribution of m for a protein interaction network, Twitter, and Facebook.

power law. The support of the probability function defined in (1) is the set V_{t-1}, so all realizations of M_t are less than or equal to $n(t-1)$. First, note that $C(t)$ satisfies

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

where

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

represents the generalized harmonic number of order s. Second, note that $H_t(s)$ depends of the initial number of nodes n_0. As t tends to infinity $H_t(s)$ exists if $s>2$. Finally, note that $C(t)=1/H_t(s)$ and

$$\lim_{t\to\infty} H_t(s)=\zeta(s), \quad (5)$$

where $\zeta(s)=\sum_{m=1}^{\infty} m^{-s}$ is the Zeta-Riemann function.

The cumulative distribution function of M_t is given by

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

where $1\leqslant m\leqslant n(t)-1$. Figure 2 shows the complementary cumulative distribution of new edges for three empirical networks and the estimated value of s. For the protein interactions, the network represents a proteome-scale map of human protein-protein interactions. For the Twitter network, edges represents follower-followee relationship between users. And for Facebook, edges represent friendships [7].

The expected number of edges that the new node establishes at time t is given by

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

where $\theta(0)=0$. The following lemma characterizes the asymptotic behavior of $\theta(t)$.

Lemma 1: If $s>2$, then the expected number of new edges satisfies:

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

Proof:

a) The result is an immediate consequence of (5) and (6).
b) Note that $H_t(s)=H_{t-1}(s)+(n_0+t-1)^{-s}$. Hence,

$$\begin{aligned}&\theta(t)-\theta(t-1)\\ &=\frac{H_{t-1}(s-1)+(n_0+t-1)^{1-s}}{H_{t-1}(s)+(n_0+t-1)^{-s}}-\frac{H_{t-1}(s-1)}{H_{t-1}(s)}\\ &=\frac{(n_0+t-1)^{-s}((n_0+t-1)H_{t-1}(s)-H_{t-1}(s-1))}{H_{t-1}^2(s)+H_{t-1}(s)(n_0+t-1)^{-s}}.\end{aligned}$$

Note that for $t>m$

$$\frac{t}{m^s}>\frac{m}{m^s}=\frac{1}{m^{s-1}}.$$

Moreover, according to (4), $(n_0+t-1)H_{t-1}(s)-H_{t-1}(s-1)>0$.
Therefore,

$$\theta(t)-\theta(t-1)>0.$$

■

Next, we want to characterize the average of all instances of $\theta(t)$ in the large t limit. Based on Lemma 1, we want to guarantee that the dynamics of the averages of $\theta(t)$ converge. Consider the sequence γ_t of instances of $\theta(t)$, given by

$$\gamma_t=\sum_{i=1}^{t}\frac{\theta(i)}{t}. \quad (7)$$

Lemma 2: If $s>2$, the asymptotic behavior of the average of the instances of $\theta(t)$ satisfies

$$\lim_{t\to\infty}\gamma_t=\frac{\zeta(s-1)}{\zeta(s)}. \quad (8)$$

Proof: The proof is based on Theorem 8.48 in [8]. Consider $T_t = \theta(t) - \frac{\zeta(s-1)}{\zeta(s)}$ and $\rho_t = \gamma_t - \frac{\zeta(s-1)}{\zeta(s)}$. We want to show that $\rho_t \to 0$ as $t \to \infty$.

On the one side, note that

$$T_1 + T_2 + \ldots + T_t = \theta(1) + \cdots + \theta(t) - t\frac{\zeta(s-1)}{\zeta(s)}, \tag{9}$$

and

$$\rho_t = \frac{T_1 + T_2 + \ldots + T_t}{t}. \tag{10}$$

Using Lemma 1, $T_t \to 0$ as $t \to \infty$, so for each $\varepsilon > 0$ we can take $N > 0$ such that $|T_t| < \varepsilon$ for all $t > N$. On the other side, note that there exists a constant $\delta > 0$ such that $|T_t| < \delta$ for all t. Hence

$$\begin{aligned}|\rho_t| \leqslant & \frac{|T_1| + |T_2| + \cdots + |T_N|}{t} \\ & + \frac{|T_{N+1}| + |T_{N+2}| + \cdots + |T_t|}{t} < \frac{N\delta}{t} + \varepsilon,\end{aligned}$$

which implies that $\limsup_{t\to\infty} |\rho_t| < \varepsilon$. Therefore, $\gamma_t \to \frac{\zeta(s-1)}{\zeta(s)}$. ■

III. DEGREE DYNAMICS

We now calculate the functional form of the evolution of the degree of a node.

Assumption 1: The number of nodes of the network grows at a constant rate.

As a direct consequence of Assumption 1, the rate of change of the degree of a node is proportional to probability that a new node establishes an edge to that node. That is, for any node i

$$\frac{dg(t,i)}{dt} = \theta(t)\,\pi(i) \tag{11}$$

$$= \frac{\theta(t)\,g^{\alpha}(t,i)}{\sum_{j\in V_t} g^{\alpha}(t,j)}. \tag{12}$$

Note that the proportionality constant $\theta(t)$ represents the expected number of edges established by a new node. Note also that for $\alpha = 1$, the sum in (12) takes into account all nodes, so the rate of change of the number of edges of node i can be written as

$$\frac{dg(t,i)}{dt} = \frac{\theta(t)\,g(t,i)}{2l_0 + 2\sum_{j=1}^{t}\theta(j)}. \tag{13}$$

According to Lemma 1, for a large enough t, the terms in the denominator can be neglected and

$$\frac{dg(t,i)}{dt} \approx \frac{\theta^* g(t,i)}{2t\theta^* + 2l_0} \tag{14}$$

$$\approx \frac{1}{2}\frac{g(t,i)}{t}, \tag{15}$$

where $\theta^* = \frac{\zeta(s-1)}{\zeta(s)}$. By integrating (15) and using the initial condition $g(t_i,i) = \theta(t_i)$, we obtain

$$g(t,i) = \theta(t_i)\left(\frac{t}{t_i}\right)^{1/2} \tag{16}$$

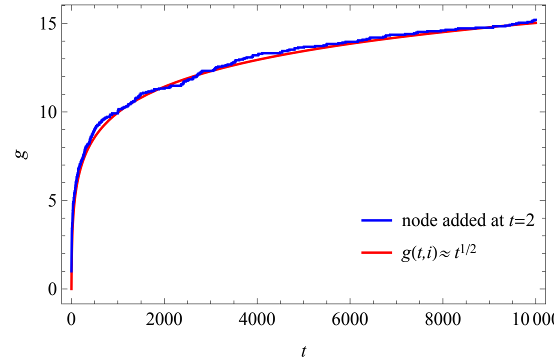

Fig. 3. Degree dynamics for a node that attaches to a simulated network at $t = 1$ and the theoretical prediction for $\alpha = 0.3$.

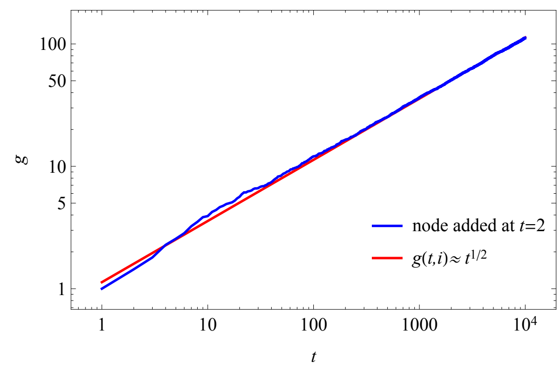

Fig. 4. Degree dynamics for a node that attaches to a simulated network at $t = 1$ and the theoretical prediction for $\alpha = 1$.

Similarly to case with $\alpha = 1$, when $0 < \alpha < 1$ the denominator at (12) can be approximated by $\left(\sum_{j\in V_t} g(t,j)\right)^{\alpha}$ and when $\alpha > 1$ by $\alpha\left(\sum_{j\in V_t} g(t,j)\right)$. Next, we specify approximations of the functional forms of g.

- For $0 < \alpha < 1$ (i.e., for *sublinear preferential attachment*)
$$g(t,i) \approx (\ln(t))^{\frac{1}{1-\alpha}} \tag{17}$$
- For $\alpha = 1$ (i.e., for *linear preferential attachment*)
$$g(t,i) \approx t^{1/2} \tag{18}$$
- For $\alpha > 1$ (i.e., for *superlinear preferential attachment*)
$$g(t,i) \approx t \tag{19}$$

Figures 3 and 4 show the behavior of $g(t,i)$ for a network generated by the model with sublinear preferential attachment ($\alpha = 0.3$) and linear preferential attachment. Both attachment mechanims satisfy $s = 4$, but only the latter represents a scale free network [2].

IV. DEGREE DISTRIBUTION

Next, we want to specify the degree distribution of the network p_k. Let A_k be the set of nodes with degree k. The probability that a new node connects to node $j \in A_k$ is given by

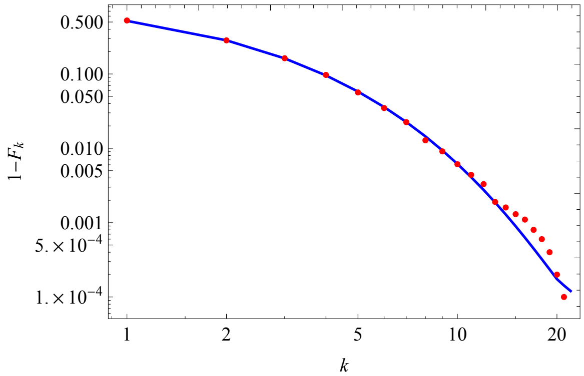

Fig. 5. Complementary cumulative degree distribution for a simulated network (dotted curve) and the theoretical prediction (solid curve) for $\alpha = 0.3$.

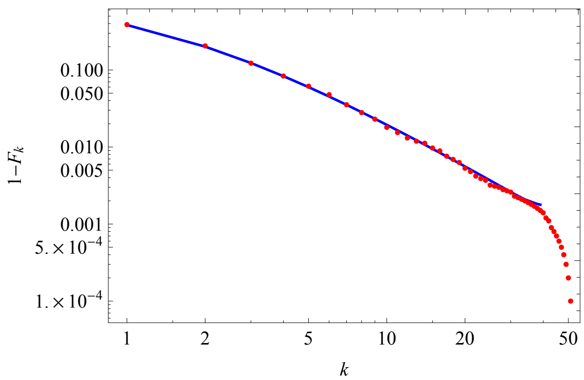

Fig. 6. Complementary cumulative degree distribution for a simulated network (dotted curve) and the theoretical prediction (solid curve) for $\alpha = 1$.

$$
\begin{aligned}
q_k = \sum_{i \in A_k} \pi(i) &= \frac{\sum_{i \in A_k} g^{\alpha}(t,i)}{\sum_{j \in V_t} g^{\alpha}(t,j)} \\
&= \frac{k^{\alpha}|A_k|}{\sum_{j \geqslant 1} j^{\alpha}|A_j|} \\
&= \frac{k^{\alpha}|A_k|}{n(t)\langle g^{\alpha}\rangle}
\end{aligned}
$$

where $\langle g^{\alpha}\rangle = \sum_{j \geqslant 1} j^{\alpha}|A_j|/n(t)$. The probability that a node has degree k at time t is given by [3],

$$
p_k = \begin{cases} \frac{\langle g^{\alpha}\rangle}{\langle g^{\alpha}\rangle + \theta(t)} & \text{if } k = 1 \\ \\ \frac{\theta(t)(k-1)^{\alpha}}{\langle g^{\alpha}\rangle + \theta(t)k^{\alpha}} p_{k-1} & \text{if } k > 1. \end{cases} \tag{20}
$$

And the cumulative degree distribution is

$$
F_k = \sum_{y=1}^{k} p_y. \tag{21}
$$

Figures 5 and 6 show the complementary cumulative distribution for a network with sublinear ($\alpha = 0.3$) and linear attachment. Again, for both mechanisms $s = 4$.

Next, we turn our attention to the behavior of the average degree of the network.

V. 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. We want to characterize the asymptotic behavior of the expected average degree of the network. Let L_t represents the number of new edges established at time t. For $t > 0$, the expected number of new edges is given by

$$
l(t) = E[L_t] = l_0 + \sum_{i=1}^{t} \theta(i). \tag{22}
$$

The first term of the right-hand of (22) corresponds to the initial number of edges of G_t; the second term to the contribution by new nodes that add M_t edges at time t. Now, let N_t denote the total degree at time t. The expected value of N_t is given by

$$
e(t) = E[N_t] = 2l(t) = 2\left(l_0 + \sum_{i=1}^{t} \theta(i)\right). \tag{23}
$$

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

$$
d(t) = E[D_t] = \frac{e(t)}{n(t)} = \frac{2l_0 + 2\sum_{i=1}^{t} \theta(i)}{n_0 + t}. \tag{24}
$$

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 the expected average degree $d(t)$ converges to

$$
\lim_{t \to \infty} d(t) = 2\frac{\zeta(s-1)}{\zeta(s)} \tag{25}
$$

Proof: According to (1), note that

$$
\lim_{t \to \infty} d(t) = 2 \lim_{t \to \infty} \frac{l_0}{n_0 + t} + 2 \lim_{t \to \infty} \frac{1}{n_0 + t} \sum_{i=1}^{t} \theta(i). \tag{26}
$$

Therefore, applying Lemma 2 in (26), we obtain (25). ■

Figure 7 illustrates the asymptotic behavior of $d(t)$. Finally, note that

$$
d(t+1) - d(t) = \frac{2}{n(t+1)n(t)} \omega(t), \tag{27}
$$

where

$$
\omega(t) = \left(n_0\theta(t+1) - l_0 + t\theta(t+1) - \sum_{i=1}^{t} \theta(i)\right).
$$

The term $t\theta(t+1) - \sum_{i=1}^{t} \theta(i)$ can be expanded as

$$
(\theta(t+1) - \theta(1)) + \ldots + (\theta(t+1) - \theta(t)). \tag{28}
$$

Since θ is a strictly increasing function, then (28) is positive, which suggests that the function ω is positive for

$t > 0$ if the first term satisfies $n_0\theta(t+1) - l_0 > 0$. Taking the minimum value of $\theta(t)$, we obtain $n_0\theta(1) - l_0 > 0$. Therefore, if $n_0\theta(1) - l_0 > 0$, then d is strictly increasing. This implies that if $d(0) < 2n_0\theta(1)$ then d is a strictly increasing function.

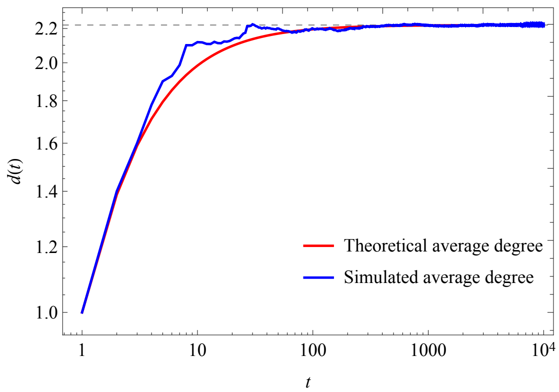

Fig. 7. Asymptotic behavior of the average degree $d(t)$ with $s = 4$.

VI. STABILITY OF THE EXPECTED AVERAGE DEGREE

Next, let $d^* = \frac{2\zeta(s-1)}{\zeta(s)}$ and consider the set $S \cup S^e$, where $S = d(\mathbb{N})$ is the direct image set of d and

$$S^e = \{(d^* + \varepsilon, d^* - \varepsilon) : \varepsilon \geqslant 0\}.$$

Clearly, $S \cup S^e \subseteq \mathbb{R}$.

Consider the function $V : S \cup S^e \to \mathbb{R}_0^+$, defined by

$$V(d(t)) = |d(t) - d^*|.$$

Note that $V(d(t)) = 0$ if and only if $d(t) = d^*$ and $V(d(t)) > 0$ for all t.

Proposition 1: The set S^e is a non-empty invariant set.

Proof: Since $d^* \in \mathbb{R}$, we can find a rational number $q \in S^e$ close enough to d^*. In particular, if we take $d(0) = q$ we guarantee that d remains close to S^e. According to Lemma 4.1 in [9], S^e is a non-empty invariant set. In particular, if $\varepsilon = 0$ then $d^* \in S^e$. ∎

The following proposition characterizes the monotonicity of the function V over time.

Proposition 2: The function $V : S \cup S^e \to \mathbb{R}_0^+$ is a decreasing function.

Proof: For the verification of the proposition we want to prove that $V(d(t+1)) - V(d(t)) < 0$. Consider the following cases:

(i) Suppose that $d(0) < 2n_0\theta(1)$, then $d(t)$ is an increasing function. Therefore,

$$V(d(t+1)) - V(d(t)) = d(t+1) - d(t) < 0.$$

(ii) Suppose that $d(0) > 2n_0\theta(1)$, then $d(t)$ is a decreasing function. Therefore,

$$V(d(t+1)) - V(d(t)) = -d(t+1) + d(t) < 0.$$

∎

Note that for all $\varepsilon_1 > 0$, there exists a $\delta_1 = \varepsilon_1 > 0$, such that all $d(t) \in S \cup S^e$. First, if $|d(t) - d^*| > \varepsilon_1$, then $V(d(t)) > \delta_1$. Second, if $|d(t) - d^*| < \varepsilon_1$, then $V(d(t)) \leqslant \delta_1$. Together with the Proposition 1 and 2, these bounds imply that d^* is stable [10]. Moreover, because $V(d(t)) \to 0$ as $t \to \infty$, d^* is globally asymptotically stable. Figure 8 shows the convergence of the Lyapunov functions, for various initial networks.

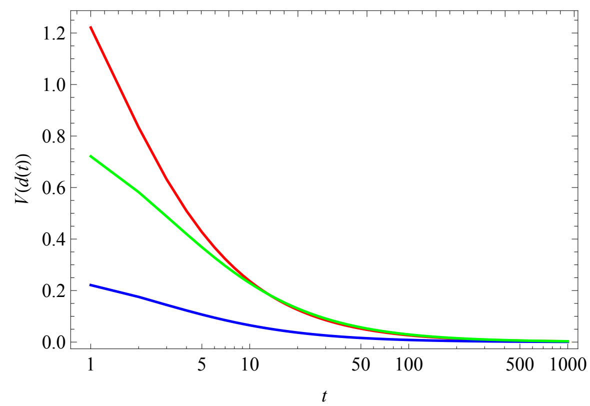

Fig. 8. Evolution of V for different initials networks.

VII. CONCLUSIONS

This paper introduces a model that relaxes the original assumption of the Barabasi-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. This distribution approaches a stationary distribution if and only if the scaling exponent of the distribution of new edges is strictly greater than two. We then show that the expected value of the average degree converges to an equilibrium. Finally, we prove that this equilibrium is globally asymptotically stable. Understanding how different types of growth in the number of new edges impact the evolution of the network remains a future research direction.

ACKNOWLEDGMENT

This research was supported in part by the Center of Excellence and Appropriation in Big Data and Data Analytics (CAOBA) at the Pontificia Universidad Javeriana, the Ministry of Information Technologies and Telecommunications of Colombia (MinTIC), and the Colombian Administrative Department of Science, Technology and Innovation (COLCIENCIAS), under grant no. FP44842-546-2015.

REFERENCES

[1] A. Barabási and R. Albert, “Emergence of scaling in random networks", Science, vol. 286, pp. 509-512, 1999.
[2] A. Barabási, “Linked: The new science of networksâĂİ. Perseus, 2002.
[3] V. Zadorozhny and E. Yudin, “Models following nonlinear preferential attachment rule", Physica A, vol. 248, pp. 111-132, 2015.
[4] P. Sheridan, Y. Yuichi , and S. Hidetoshi. “A preferential attachment model with Poisson growth for scale-free networks", Annals of the Institute of Statistical Mathematics vol. 60, no. 4, pp. 747-761. 2008.
[5] S. Dereich and P. Morters. “Random networks with sublinear preferential attachment: the giant component", The Annals of Probability, vol. 41, no. 1, pp. 329-384. 2013.
[6] W. Liang and S. Zhu, “Scale-free networks by super-linear preferential attachment rule", Physica A: Statistical Mechanics and its Applications, vol. 387, no. 14, pp. 3789-3795. 2008.
[7] KONECT, “The Koblenz Network Collection", 2016. [online] http://konect.uni-koblenz.de, [Accessed: 19- Mar- 2017].
[8] T. Apostol. *Mathematical analysis*, Pearson, Second ed., 1974.
[9] H. Khalil, *Nonlinear Systems*, Pearson, Third ed., 2001.
[10] K. Burgess and K. Passino, “Stability analysis of load balancing systems", International Journal of Control, vol. 61, no. 2, pp. 357-393, 1995.