

Emergence of Scale Free Networks from Ideal Free Distributions

Jorge Finke, Nicanor Quijano, and Kevin M. Passino *

February 7, 2007

Abstract

This paper presents a class of network optimization processes that account for the emergence of scale-free network structures. We introduce a mathematical framework that captures the connectivity and growth dynamics of a network with an arbitrary initial topology. We show how selection via differential node fitness affects the proportion of connections a node makes to other nodes, and how a heavy-tailed connectivity behavior manifests itself from consecutive achievements of ideal free distributions (IFDs). Finally, we present simulation results that show how this class of networks may emerge even when consecutive IFDs are not perfectly reached.

1 Introduction

There is growing interest in understanding the emergence of a class of real-world networks called “scale-free networks” (e.g., computer networks such as the World Wide Web [1], some protein-protein interaction networks [2], and networks created by the formation of sexual partnerships [3]). In this context, the number of edges (connections) that is most commonly found in a network (graph) indicates the *scale* of its connectivity distribution (e.g., the peak in a Poisson or bell-shaped distribution). Broadly speaking, the most notable feature of a scale-free network is its heavy-tailed (power-law), rather than a Poisson or bell-shaped, connectivity distribution. In particular, power-law distributions indicate that the probability $P(k)$ that a node connects to k other nodes is proportional to $k^{-\beta}$ for some positive constant β , implying that the number of edges (the degree) of the nodes of the network comprises different orders of magnitude (i.e., with a few nodes having a high degree, many having only a low one, and without a peak in the distribution). Although a prominent feature of scale-free networks, power-law distributions alone do not entirely describe the topological structure of these networks.

Another important feature that characterizes scale-free networks is the fact that highly connected nodes are inter-connected. Such nodes are called “hubs” and they represent the core structure of the network. They are both the strength and weakness of scale-free networks. If failures occur at *randomly* selected nodes, the vast majority of nodes are those with a small degree (i.e., not hubs), and the likelihood that a hub will be affected is almost negligible. By keeping its hubs intact, the network will not lose its connectedness. However, for *targeted* attacks (e.g., if we choose a few major hubs and try to take them out of the network) the likelihood of hub failures increases which consequently would cause the network to collapse. This property has come to be known as the “Achilles’ heel” of scale-free networks.

Although both of the above topological features may emerge in networks without growth (e.g., [4],[5]), most real-world networks describe interconnected systems that grow by the continuous addition of new nodes. For growing networks, “preferential attachment” (a concept introduced in [6]) is perhaps the best known mechanism that accounts for the emergence of scale-free networks (variations of this mechanism can be found in [7]-[9]). Preferential attachment assumes that the probability that a new node will connect to another node is proportional to the degree of that node. While preferential attachment has been proven to produce scale-free networks [10], many different mechanisms that lead to these types of networks have also been discovered (e.g., initial attractiveness [11], accelerating growth [12], and gradual aging [13]).

*This work was partially supported by the AFRL/VA and AFOSR Collaborative Center of Control Science (Grant F33615-01-2-3154).

In this paper, we do not focus on a particular mechanism that leads to scale-free networks. Instead, we introduce a generic set of dynamical equations that captures the broad tendencies of the connectivity of the nodes of a network. We concentrate on a particular state of the network where each node has the same “suitability” level (assumed to be a correlate of the relative fitness of a node), and show how scale-free network may emerge from consecutive achievements of this desired state. In particular, we use a concept from ecology known as the “ideal free distribution” (IFD) [14] to define the desired state. The IFD concept allows us to capture the dynamic *coupling* that results between the different nodes (i.e., it allows us to characterize how an edge connecting any two nodes affects another node in the network). In this context, the word “ideal” refers to the assumption that the connectivity dynamics of the network are *equally* affected by every node (i.e., the impact of a node on the connectivity dynamics of the network depends on the state of that node, but not on its location within the network). “Free” indicates that the nodes of the network can connect instantaneously and at no cost to any other node. Although our framework assumes that both assumptions will hold, there is an extensive literature that shows how an IFD may still be achieved even when one or both assumptions are relaxed (e.g., see [15]-[22]). It is, however, the general idea behind the IFD concept and the emergence of scale-free networks that is the most important and novel contribution of this paper.

It is important to understand how scale-free networks may emerge from IFD distributions for several reasons. First, from a game theoretic perspective, the IFD is a Nash equilibrium. Therefore, when a network reaches the IFD none of its nodes has anything to gain by unilaterally re-establishing edges to other nodes. Moreover, the IFD has been shown to be an evolutionarily stable strategy (ESS) [22]-[24], which broadly speaking means that whenever the IFD is achieved, there does not exist any other network topology (with the same edge capacity) that could “take over” the existing IFD distribution. In other words, the evolution of scale-free networks from IFDs shows that these networks may emerge from nodes adopting an ESS strategy, which is in fact a pattern regularly found in a large class of adaptive dynamics [25]. Second, the IFD is a concept that characterizes the outcome of a large class of attachment rules without explicitly describing the particular mechanism that leads to the emerging networks. Our approach borrows concepts from evolutionary ecology and control theory, and uses existing methods from both fields to relate these concepts to the scale-free literature. Finally, the emergence of scale-free networks from IFDs seems to be a robust process in the sense that even when perfect IFDs may not be achieved (e.g., under suboptimal conditions where the nodes cannot connect to any other node) scale-free networks will still emerge.

The remainder of this paper is organized as follows: In Section 2 we introduce a model that captures the connectivity and growth dynamics of a continuously growing network. We present sufficient conditions that guarantee that for a fixed number of nodes the IFD is an asymptotically stable state of the network as long as enough edges between the nodes can be established (Theorem 2.1). We then prove that consecutive achievements of IFDs lead to connectivity distributions that are scale-free (Theorem 2.2). Finally, in Section 3 we consider information-constrained cases of our model and present simulation results that show that scale-free networks may still emerge even when consecutive IFDs are not perfectly reached.

2 A Model of Network Evolution

Let the graph \mathcal{G}_0 represent a network at generation $n = 0$ with N_0 nodes, each with $M_0 \in \mathbb{R}_+$ edges, $\mathbb{R}_+ = (0, \infty)$, that are randomly connected to other nodes. For every generation $n \in \mathbb{N}$, $n > 0$, a new node with M edges, $\infty > M \geq M_0$, is added to the network (i.e., the new node may connect several edges to the same node in \mathcal{G}_n). After n generations the graph \mathcal{G}_n has a total of $N_n = n + N_0$ nodes. Let $\mathcal{H}(n) = \{1, \dots, N_n\}$ represent all the nodes of the network at generation n , so that this set includes the node that has been added at that generation. Furthermore, assume that the rate at which the newly added node and the existing nodes establish new edges is considerably faster than the rate at which \mathcal{G}_n grows. We will justify this assumption below. Next, in Sections 2.1 and 2.2 we develop a two-time-scale model that captures the particular connectivity and growth dynamics of such a network.

2.1 Connectivity Dynamics via Network Competition

For a fixed generation n , each node $i \in \mathcal{H}(n)$ is associated with a “suitability function” that characterizes the benefit for that node from creating new edges to other nodes. We assume an inverse relationship between the suitability level of a node and the number of new edges it establishes (e.g., in the internet since the server performance can degrade if more users are added, or in biology since it becomes more difficult to establish metabolic links if there are already too many present). In particular, let the suitability function of node i during generation n be defined by

$$s_i(x_i(t)) = \frac{n_i^{-\beta}}{k_i(n) + x_i(t)} \quad (1)$$

where $n_i \in \mathbb{N}$ is the generation step at which node $i \in \mathcal{H}(n)$ was added to the network (let $n_i = 1$ for all nodes in \mathcal{G}_0 so that s_i is well defined), $\beta \in (0, 1)$ is a predefined constant (e.g., $\beta = 1/2$ for networks with linear growth and linear preferential attachment [6]), $k_i(n) \in \mathbb{R}_+$ is the number of edges of node i between the end of generation $n - 1$ and start of generation n (i.e., $k_i(n)$ does not take into account *new* edges that are created during generation n and $k_i(0) = M_0$ for all nodes in \mathcal{G}_0), and $x_i(t) \in \mathbb{R}_+$ is the number of edges that are established during that generation with node i at some time t , $t \in \mathbb{R}_+$. We assume that edges that were connecting any two nodes during previous generations (those characterized by $k_i(n)$) cannot be removed. On the other hand, a new edge connecting two nodes may be established and removed (i.e., we allow for rewiring of edges that have been established within the same generation). Moreover, note that Eq. (1) assumes an inverse relationship between the suitability level of a node and the generation step when it is added to the network. If we let node i^* denote the node that has been added to the network during generation $n > 0$, then note that since a constant number of edges are added by node i^* it always satisfies $n_{i^*} = n$, $k_{i^*}(n_{i^*}) = 0$, and $x_{i^*}(t) = M$ for all $t > 0$.

Let $t = 0$ be the time instant at which the new node is added to the graph \mathcal{G}_n (i.e., the start of generation n), and assume that only when $t \rightarrow \infty$ will \mathcal{G}_n evolve into the next generation $n + 1$ (i.e., when another node will be added to the network). Let $x(t) = [x_1(t), \dots, x_{N_n}(t)]^\top$ represent the state of network at generation n and time t (i.e., the degree of every one of its nodes). We now focus on the dynamics of $x(t) = [x_1(t), \dots, x_{N_n}(t)]^\top$ for $t > 0$. In particular, let us assume that for a fixed generation n , the number of edges of node $i \in \mathcal{H}(n)$ evolves with respect to t according to

$$\frac{\dot{x}_i}{x_i(t)} = \alpha_n \left(\frac{n_i^{-\beta}}{k_i(n) + x_i(t)} - \sum_{j \in \mathcal{H}(n)} \frac{x_j(t)}{C_n} \frac{n_j^{-\beta}}{k_j(n) + x_j(t)} \right) \quad (2)$$

where C_n is a constant that we define below, and $\alpha_n > 0$ is a constant that characterizes the rate at which new connections are established in the network. Note that the left-hand-side of Eq. (2) is the normalized rate of change of the newly added edges of node i , and the right-hand-side indicates that if the benefit of linking a new edge to node i is greater (less) than the (weighted) average, then the rate of change of added edges of node i will increase (decrease). In other words, Eq. (2) characterizes tendencies of the connectivity of the nodes of a network that may be driven by different attachment mechanisms.

Next, let $C_n \in \mathbb{R}_+$, $C_n \geq 2M$, be the total number of new *connections* that are established during generation n and let

$$\Delta_n = \left\{ x \in \mathbb{R}_+^{N_n} : \sum_{j \in \mathcal{H}(n)} x_j = C_n \right\}$$

represent the simplex over which these C_n new connections may be distributed. Note that since a given edge connects two particular nodes, C_n new connections represent the endpoints of $C_n/2$ new edges. Furthermore, let

$$\mathcal{X}_I(n) = \left\{ x \in \mathbb{R}_+^{N_n} : s_i(x_i) = \frac{n^{-\beta}}{M} \text{ for all } i \in \mathcal{H}(n) \right\} \quad (3)$$

define an invariant set that characterizes the distribution of interest of new connections across the entire network (recall that $M \geq M_0$ is the number of edges established with the new node that is added at each generation). Note that for any generation n , $\bar{x} \in \mathcal{X}_I(n)$ represents an IFD in the sense that edges are distributed in such a way that all nodes in $\mathcal{H}(n)$ have the same suitability levels [14].

Theorem 2.1 *For any fixed network generation n , there exists a positive constant C_n , such that if C_n new connections are established in the network, and the degree of every node of \mathcal{G}_n satisfies the dynamics described by Eq. (2), then $\mathcal{X}_I(n)$ (i.e., the IFD defined by Eq. (3)) is asymptotically stable with region of attraction Δ_n .*

Proof: Fix a generation $n > 0$. We begin by characterizing any edge distribution $\bar{x} = [\bar{x}_1, \dots, \bar{x}_{N_n}]^\top$ such that $\bar{x} \in \mathcal{X}_I(n)$. In particular, note that if $\bar{x} \in \mathcal{X}_I(n)$ then $\frac{ds_i(\bar{x}_i)}{dt} = 0$, and according to Eq. (1), $\frac{d\bar{x}_i}{dt} = 0$ for all $i \in \mathcal{H}(n)$. Moreover, since $\bar{x} > 0$, if the degree of every node of \mathcal{G}_n satisfies the dynamics described by Eq. (2) then $\bar{x} \in \mathcal{X}_I(n)$ satisfies

$$\frac{n_i^{-\beta}}{k_i(n) + \bar{x}_i} = \frac{1}{C_n} \sum_{j \in \mathcal{H}(n)} \bar{x}_j \frac{n_j^{-\beta}}{k_j(n) + \bar{x}_j}$$

for all $i \in \mathcal{H}(n)$, which implies that

$$\bar{x}_i = \frac{n_i^{-\beta} \sum_{j \in \mathcal{H}(n)} k_j(n) + n_i^{-\beta} \sum_{j \in \mathcal{H}(n)} \bar{x}_j(n) - k_i(n) \sum_{j \in \mathcal{H}(n)} n_j^{-\beta}}{\sum_{j \in \mathcal{H}(n)} n_j^{-\beta}}$$

Furthermore, since $\sum_{j \in \mathcal{H}(n)} \bar{x}_j = C_n$, we know that

$$\bar{x}_i = \frac{n_i^{-\beta} \sum_{j \in \mathcal{H}(n)} k_j(n) + n_i^{-\beta} C_n - k_i(n) \sum_{j \in \mathcal{H}(n)} n_j^{-\beta}}{\sum_{j \in \mathcal{H}(n)} n_j^{-\beta}} \quad (4)$$

By substituting Eq. (4) into Eq. (1) we get

$$\begin{aligned} s_i(\bar{x}_i) &= \frac{n_i^{-\beta} \sum_{j \in \mathcal{H}(n)} n_j^{-\beta}}{n_i^{-\beta} \sum_{j \in \mathcal{H}(n)} k_j(n) + n_i^{-\beta} C_n - k_i(n) \sum_{j \in \mathcal{H}(n)} n_j^{-\beta} + k_i(n) \sum_{j \in \mathcal{H}(n)} n_j^{-\beta}} \\ &= \frac{\sum_{j \in \mathcal{H}(n)} n_j^{-\beta}}{\sum_{j \in \mathcal{H}(n)} k_j(n) + C_n} \end{aligned} \quad (5)$$

Note that for all $i \in \mathcal{H}(n)$ such that $\bar{x} \in \mathcal{X}_I(n)$ each node i satisfies $s_i(\bar{x}_i) = \frac{n^{-\beta}}{M}$. Furthermore, using Eq. (5) we know that there exists a single constant

$$C_n = M \sum_{j \in \mathcal{H}(n)} \left(\frac{n_j}{n}\right)^{-\beta} - \sum_{j \in \mathcal{H}(n)} k_j(n) \quad (6)$$

and a *unique* distribution \bar{x} on Δ_n that belongs to the invariant set $\mathcal{X}_I(n)$. Next, we show that $\bar{x} \in \mathcal{X}_I(n)$ is asymptotically stable with region of convergence Δ_n .

Let V_n be defined as

$$V_n = - \sum_{i \in \mathcal{H}(n)} \frac{\bar{x}_i}{C_n} \ln \left(\frac{x_i}{\bar{x}_i} \right)$$

which is known as the Kullback-Leibler distance or the relative entropy function [26],[27]. It has been shown in [24] that this is a valid Lyapunov function candidate (i.e., it is positive definite and radially unbounded).

Let $\bar{s} = \sum_{i \in \mathcal{H}(n)} \frac{x_i}{C_n} s_i$ denote the weighted average of the suitability levels of all nodes, and note that by taking the derivative along trajectories we obtain

$$\begin{aligned}
\dot{V}_n &= \frac{\partial V}{\partial x} \frac{\partial x}{\partial t} \\
&= -\frac{1}{C_n} \sum_{i \in \mathcal{H}(n)} \bar{x}_i \frac{1}{x_i} \dot{x}_i \\
&= -\frac{1}{C_n} \sum_{i \in \mathcal{H}(n)} \bar{x}_i \alpha_n (s_i - \bar{s}) \\
&= -\frac{\alpha_n}{C_n} \sum_{i \in \mathcal{H}(n)} \bar{x}_i s_i + \frac{\alpha_n \bar{s}}{C_n} \sum_{i \in \mathcal{H}(n)} \bar{x}_i \\
&= -\frac{\alpha_n}{C_n} \sum_{i \in \mathcal{H}(n)} \bar{x}_i s_i + \frac{\alpha_n \bar{s} C_n}{C_n} \\
&= -\frac{\alpha_n}{C_n} \sum_{i \in \mathcal{H}(n)} \bar{x}_i s_i + \alpha_n \sum_{i \in \mathcal{H}(n)} \frac{x_i}{C_n} s_i \\
&= \frac{\alpha_n}{C_n} \sum_{i \in \mathcal{H}(n)} s_i (x_i - \bar{x}_i)
\end{aligned}$$

In order to prove that $\dot{V}_n < 0$, we will show that the maximum of \dot{V}_n is zero, and that this maximum is only reached when $x \in \mathcal{X}_I(n)$. We use Lagrange multiplier theory [28], and define the following optimization problem

$$\begin{aligned}
\max J &= \frac{\alpha_n}{C_n} \sum_{i \in \mathcal{H}(n)} \frac{n_i^{-\beta}}{k_i(n) + x_i} (x_i - \bar{x}_i) \\
\text{subject to} & \quad x \in \Delta_n
\end{aligned}$$

First, we need to define the Jacobian of the cost function J which using Eq. (4) is given by

$$\begin{aligned}
\frac{\partial J}{\partial x_i} &= \frac{\alpha_n n_i^{-\beta} (k_i(n) + \bar{x}_i)}{C_n (k_i(n) + x_i)^2} \\
&= \frac{\alpha_n n_i^{-\beta} \left(n_i^{-\beta} \sum_{j \in \mathcal{H}(n)} k_j(n) + n_i^{-\beta} C_n \right)}{C_n (k_i(n) + x_i)^2 \sum_{j \in \mathcal{H}(n)} n_j^{-\beta}}
\end{aligned}$$

Using Lagrange multipliers for any $i \in \mathcal{H}(n)$, we obtain that

$$\frac{\alpha_n \left(n_i^{-\beta} \right)^2 \left(\sum_{j \in \mathcal{H}(n)} k_j(n) + C_n \right)}{C_n (k_i(n) + x_i^*)^2 \sum_{j \in \mathcal{H}(n)} n_j^{-\beta}} + \lambda^* = 0$$

where x^* is the regular point for the optimization problem, and λ^* is the Lagrange multiplier. For any $i, j \in \mathcal{H}(n)$ this is equivalent to

$$\frac{\alpha_n \left(n_i^{-\beta} \right)^2 \left(\sum_{j \in \mathcal{H}(n)} k_j(n) + C_n \right)}{C_n (k_i(n) + x_i^*)^2 \sum_{j \in \mathcal{H}(n)} n_j^{-\beta}} = \frac{\alpha_n \left(n_j^{-\beta} \right)^2 \left(\sum_{j \in \mathcal{H}(n)} k_j(n) + C_n \right)}{C_n (k_j(n) + x_j^*)^2 \sum_{j \in \mathcal{H}(n)} n_j^{-\beta}} \quad (7)$$

Moreover, since all the terms are positive, Eq. (7) implies

$$n_i^{-\beta} (k_j(n) + x_j^*) = n_j^{-\beta} (k_i(n) + x_i^*)$$

Summing over all terms in $\mathcal{H}(n)$, we obtain that x_i^* is the same as the one defined in Eq. (4).

Now, in order to prove that x_i^* is indeed a global optimum point, we need to show that the Hessian is negative definite. In this case, note that

$$\frac{\partial^2 J}{\partial x_i^2} = -\frac{2\alpha_n n_i^{-\beta} (\bar{x}_i + k_i)}{C_n (k_i(n) + x_i)^3} < 0$$

and

$$\frac{\partial^2 J}{\partial x_i \partial x_j} = 0$$

Hence, $\nabla^2 J$ is negative definite and x^* is a global maximum for the constrained optimization problem. Moreover, the maximum value of J is given by

$$\max J = \frac{\alpha_n}{C_n} \sum_{i \in \mathcal{H}(n)} \frac{n_i^{-\beta}}{k_i(n) + x_i^*} (x_i^* - \bar{x}_i) = 0$$

Since $\dot{V}_n = \max J$, $\dot{V}_n < 0$ for all $x_i \neq \bar{x}_i$ and $\dot{V}_n = 0$ if $x_i = \bar{x}_i$ the equilibrium point in Eq. (4) is asymptotically stable. In other words, \dot{V}_n is negative, except when it is equal to the equilibrium point. Therefore, the region of asymptotic stability is defined by the set Δ_n . ■

Remarks:

- The proposed connectivity dynamics imply that, given the same number of connecting edges, nodes that have been added to the network at earlier generation steps will have higher suitability levels than those added at later generations (i.e., by the way suitability functions are defined in Eq. (1)). They relate broadly to the basic idea behind preferential attachment [6] where older nodes establish more edges because they have been part of the network for more generations, and had higher chances of establishing them. Note, however, that unlike preferential attachment our model assumes that several edges may be established between any two nodes.
- There are different models that lead to the IFD (or an approximate version of it) even when the ideal and free assumptions are relaxed to some extent. Such models may allow for suboptimal conditions that are far from the ones assumed in this section. In particular, nodes may establish random connections between them or create new edges with nodes of lower suitabilities, the nodes' rate of change of new edges may not be proportional to the difference in suitabilities, but to the highest suitability level of the entire network, etc. (for a brief overview of several mechanisms leading to IFDs see [22]).
- The IFD reached at each generation step can also be seen as an “ideal dominance distribution” (IDD) [14]. In behavioral ecology, the IDD is interpreted via sequential settling of species of animals at IFDs (i.e., in this context, sequential edge attachments). At each generation step when a new node enters the system, the network will try to distribute its connections with respect to those nodes that are “better” (i.e., have higher suitability levels). However, not all nodes can be connected to the best ones since this becomes harder as the number of connections increases (and its suitability levels decrease).
- For a given network generation n , note that according to Eq. (6) C_n increases as M increases. This implies that for the IFD to be achieved, as more connections are made by a new node, the total number of new connections across the entire network must also increase. Moreover, as β increases, the difference in suitability of two consecutively added nodes (with the same number edges) also increases, and the total number of new connections must again increase for the IFD to be reached.

Next, we show how consecutive achievements of IFDs may impact the connectivity distribution of the network as it grows.

2.2 Network Evolution

Given \mathcal{G}_n , let the degree of node i at generation $n + 1$ be given by

$$k_i(n + 1) = k_i(n) + \bar{x}_i \quad (8)$$

where $\bar{x} \in \mathcal{X}_I(n)$ holds the number of edges ultimately established by all nodes during generation n . Theorem 2.2 characterizes the evolution of $k_i(n)$ as $n \rightarrow \infty$.

Theorem 2.2 *Given any initial graph \mathcal{G}_0 with N_0 nodes, each with M_0 randomly distributed edges, if a new node with $M \geq M_0$ edges is added to the network and the degree of each node evolves according to Eq. (8), then as $n \rightarrow \infty$ the graph \mathcal{G}_n attains a scale-free connectivity distribution. Moreover, the total number of edges created during each generation $C_n \rightarrow \frac{M}{1-\beta}$ as $n \rightarrow \infty$.*

Proof: Note that for any generation n and for all nodes to achieve the same suitability the IFD must be reached at the suitability level of the node that has been added during that generation, i.e., node i^* , since it establishes a constant number of new connections. In particular, since $n_{i^*} = n$, $k_{i^*}(n) = 0$, and $x_{i^*} = M$ we know that

$$s_{i^*}(x_{i^*}) = \frac{n^{-\beta}}{M} \quad (9)$$

Moreover, if $\bar{x} \in \mathcal{X}_I(n)$ for each n (i.e., if the IFD is achieved at every generation step n), then according to Eqs. (1), (3), and (9)

$$\frac{n^{-\beta}}{M} = \frac{n_i^{-\beta}}{k_i(n) + \bar{x}_i}$$

for all $i \in \mathcal{H}(n)$. Moreover, according to Eq. (8) we know that

$$k_i(n + 1) = M \left(\frac{n}{n_i} \right)^\beta$$

which indicates that the degree of any node i follows a power law. Following the same ideas as in [29], one can then write the probability that a node has degree $k_i(n + 1)$ smaller than k as,

$$\begin{aligned} P[k_i(n + 1) < k] &= P \left[M \left(\frac{n}{n_i} \right)^\beta < k \right] \\ &= P \left[n_i > n \left(\frac{M}{k} \right)^{1/\beta} \right] \\ &= 1 - \frac{M^{1/\beta} n}{k^{1/\beta} (n + N_0)} \end{aligned}$$

since nodes are added at equal time intervals to the network. Therefore, the distribution of edges is given by

$$P(k) = \frac{\partial P[k_i(n + 1) < k]}{\partial k} \sim 2M^{1/\beta} k^{-\gamma}$$

as $n \rightarrow \infty$ with $\gamma = \frac{1}{\beta} + 1$, which leads to a scale-free connectivity distribution.

Finally, we show that C_n converges as n tends to infinity. In particular, note that the the total number of established connections at a generation n can be rewritten as

$$\sum_{j \in \mathcal{H}(n)} k_j(n) = \sum_{j=1}^{n-1} C_j + M_0 N_0$$

so that according to Eq. (6) we can write the total number of connections that have been added up to (and including) generation n as

$$\sum_{j=1}^n C_j = M \sum_{j \in \mathcal{H}(n)} \left(\frac{n_j}{n}\right)^{-\beta} - M_0 N_0 \quad (10)$$

Applying Eq. (10) to generations $n-1$ and n we obtain the the total number of edges added during generation n expressed as

$$\begin{aligned} C_n &= M \left(\sum_{j \in \mathcal{H}(n)} \left(\frac{n_j}{n}\right)^{-\beta} - \sum_{j \in \mathcal{H}(n-1)} \left(\frac{n_j}{n-1}\right)^{-\beta} \right) \\ &= M \left(\sum_{j \in \mathcal{H}(n)} \left(\frac{n_j}{n}\right)^{-\beta} - \sum_{j \in \mathcal{H}(n-1)} \left(\frac{n_j}{n-1}\right)^{-\beta} - \left(\frac{n}{n-1}\right)^{-\beta} + \left(\frac{n}{n-1}\right)^{-\beta} \right) \\ &= M \left(\frac{n}{n-1}\right)^{-\beta} + M \left(\sum_{j \in \mathcal{H}(n)} \left(\frac{n_j}{n}\right)^{-\beta} - \sum_{j \in \mathcal{H}(n)} \left(\frac{n_j}{n-1}\right)^{-\beta} \right) \\ &= M \left(\frac{n}{n-1}\right)^{-\beta} + M \left(\sum_{j \in \mathcal{H}(n)} \frac{n^\beta - (n-1)^\beta}{n_j^\beta} \right) \end{aligned}$$

Since $\lim_{n \rightarrow \infty} \left(\frac{n}{n-1}\right)^{-\beta} = 1$ and $\lim_{n \rightarrow \infty} \sum_{j \in \mathcal{H}(n)} \frac{n^\beta - (n-1)^\beta}{n_j^\beta} = \frac{\beta}{1-\beta}$ we conclude that

$$\begin{aligned} \lim_{n \rightarrow \infty} C_n &= M \left(1 + \frac{\beta}{1-\beta} \right) \\ &= \frac{M}{1-\beta} \end{aligned}$$

■

Remarks:

- Note that our model assumes that the edge connectivity process which is represented by a fast time scale does not interfere with the network's growing process which is represented by a slow one (i.e., the two time scales are isolated). Although our results in Theorem 2.2 require that the time indices of both scales tend to infinity, this is a valid approximation that is commonly used in evolutionary ecology when two processes are known to evolve at disparate rates.
- While the hubs in a scale-free network play the key role in determining the robustness properties against random node failures, the emergence of a scale-free network from consecutive achievements of IFDs suggests that its connectivity distribution evolves from a particular node strategy, which if adopted by every node in the network cannot be invaded by any competing alternative strategy (i.e., since an IFD strategy where the suitability of every node is equal is in fact an ESS). In a game theory context, this means that if we view the new node that is added at every generation step as a mutant strategy, then such a strategy will never invade an incumbent IFD strategy (adopted by all other nodes) by playing against the field. It implies that natural selection would be the driving force that selects against using strategies with lower payoffs and ensures that all nodes adopt an IFD strategy which would then lead to a scale-free connectivity distribution. Broadly speaking, our results relate to the emergence of scaling in software architecture graphs from well-defined local optimization processes [30].

3 Simulations

To gain insight into the connectivity dynamics of the network let us first illustrate some concepts introduced in Section 2.1. Let $\beta = \frac{1}{2}$, $M = 2$, $M_0 = 1$, and $N_0 = 1$, and let us study the evolution of the network for

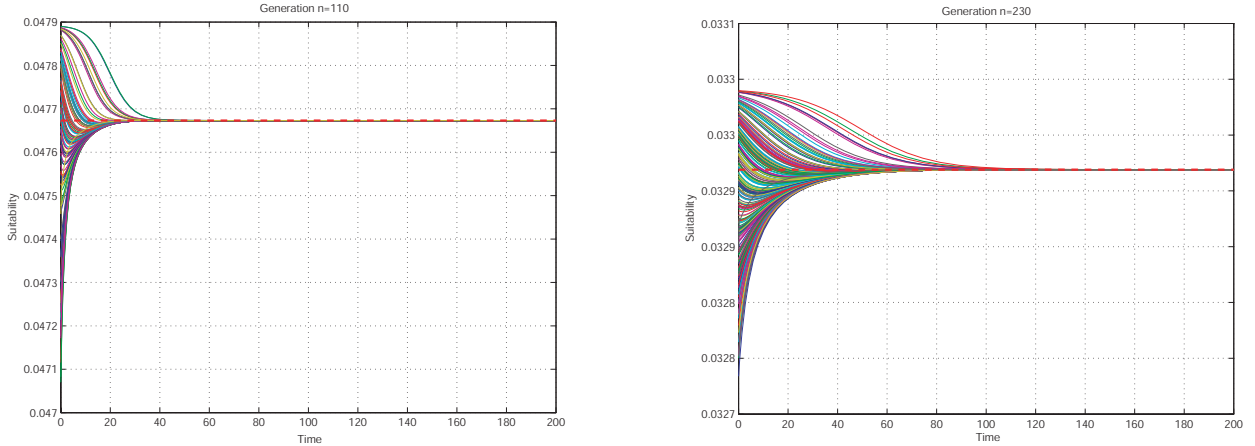


Figure 1: The suitability functions defined by Eq. (1) are plotted for different generations; when $n = 110$ (left), $n = 230$ (right). In both cases $\beta = \frac{1}{2}$, $M = 2$, $M_0 = 1$, and $N_0 = 1$.

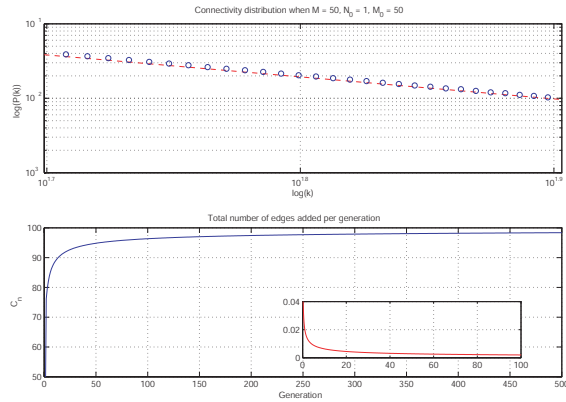
up to 250 generations. Using these parameters we simulate the dynamics described in Eq. (2) for each node at every generation step. Figure 1 shows the suitability for generations $n = 110$ and $n = 230$. As we proved above, an equilibrium point is reached and it is indeed the IFD described by Eq. (3). For instance, if we take the 50th network generation, the suitability function for each of node $i \in \mathcal{H}(n)$ converges to $\sqrt{2}/20$, which is equal to $\frac{n^{-\beta}}{M}$, when $n = 110$ and the values for β and M are the ones selected before (of course, the same holds for each generation step n). Note that the equilibrium point decreases as n increases as we expected from Eq. (3). Moreover, as the number of generation steps increases, the time it takes to converge increases also, but an IFD is still reached.

Next, to study the network's emerging connectivity we run simulations using 10000 generations. The top plot in Figure 2a shows the results when $\beta = \frac{1}{2}$, $M = 50$, $M_0 = 50$, and $N_0 = 1$. The dotted line represents the theoretical probability distribution, while the dots represent the results for our network model. As expected, the probability that a node has many connections is small and decreases as the number of connections increases. The exponent for the power law in this case is $\gamma = 3$. An important feature of our model is the fact that the number of edges added per generation does not grow unbounded. The bottom plot in Figure 2a illustrates the convergence of the total number of new connections C_n for the first 500 generation steps and the same parameters as before.

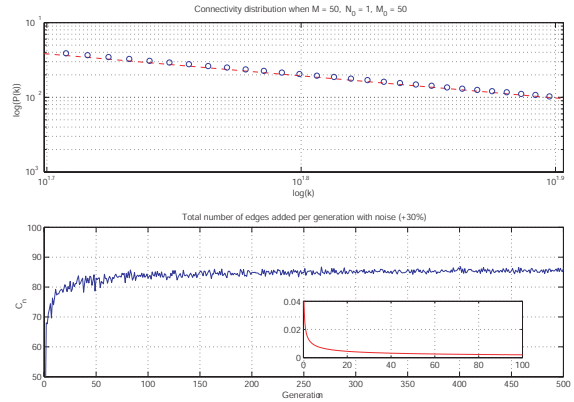
Finally, in order to study robustness properties of our model, we study the effect of not reaching the IFD at its desired level (i.e., at $\frac{n^{-\beta}}{M}$ for generation n). Figure 2b shows the connectivity distribution and the total number of edges created during each generation when we perturb the desired IFD state by adding random noise to C_n (i.e., when too many or too few edges are established). We modify the C_n value computed in Eq. (6) by adding to it a zero mean random variable with finite variance. While the top plot in Figure 2b shows that a scale-free distribution will still emerge, the bottom plot shows that C_n no longer converges to $\frac{M}{1-\beta}$, yet remains bounded.

Next, we allow for some random edges to be created between any two nodes, so that equal suitability levels at the IFD cannot be achieved (i.e., only an approximate IFD can be reached, but $x(t) \in \Delta_n$ for all t). Even with this type of perturbation (on the state of the network), a scale-free distribution again emerges, and the simulation results are similar to those presented in Figure 2a.

Note that Figure 2 represents information constrained cases of our model. For the first case (Figure 2b) we would have some type of irregularity in the system when determining the number of edges that need to be added per generation. Consequently, instead of having a perfect value for C_n we obtain just a noisy estimated value. In the second case, the information is altered when we distribute the number of edges that are ultimately established. In either case, this implies that the IFD is only reached approximately (i.e., not as defined by Eq. (3)), but a scale-free network may still emerge.



(a) Simulation results when consecutive IFDs are perfectly reached.



(b) Simulation results when approximate IFDs are reached.

Figure 2: Simulation results for 10000 generations. The parameters selected for these simulations are $\beta = \frac{1}{2}$, $M = 50$, $M_0 = 50$, and $N_0 = 1$. The top plots illustrate the connectivity distributions from the theory (dashed) [6] and from simulations (o). The bottom plots illustrate the total number of new connections that are added each generation. The small inset plots show the suitability at which the IFD is achieved for the first 100 generations.

4 Conclusions

In our model of network evolution, any two nodes establish new connections as a result of their difference in suitability levels. Due to maintenance costs, establishing a new edge punishes the connected nodes by degrading their suitability levels. We define a particular class of suitability functions where nodes that have been part of the network for longer pay a lower penalty to maintain the same amount of connections than nodes recently added. We then prove that for this class of suitability functions, a node connectivity where all nodes have equal suitability levels, namely an ideal free distribution (IFD), is an asymptotically stable equilibrium point.

When a new node is added to the network it attaches to a finite number of nodes, thereby forcing the suitability levels of these nodes to decrease. We show that under certain conditions a new IFD will be reached, so that the suitability of the new node eventually equals that of all other nodes. Furthermore, we prove that as the network grows and more connections are established, subsequent achievements of IFDs will lead to a connectivity distribution that is scale-free.

Finally, our simulation results suggest that not achieving perfect IFDs does not considerably affect the emergent distribution. Developing a mathematical model that takes into account how information flow constraints (e.g., delays in sensing suitability levels and establishing new connections) affect the achievable IFD and the emerging connectivity distributions remains a future research direction, as does considering a wider class of suitability functions.

References

- [1] R. Albert, H. Jeong, and A.-L. Barabási, “Internet: Diameter of the world-wide web,” *Nature*, vol. 401, pp. 130–131, 1999.
- [2] R. V. Solé, R. Pastor-Satorras, E. D. Smith, and T. Kepler, “A model of large-scale proteome evolution,” *Advances in Complex Systems*, vol. 5, pp. 43–54, 2002.
- [3] A. Schneeberger, C. H. Mercer, S. A. Gregson, N. M. Ferguson, C. A. Nyamukapa, R. M. Anderson, A. M. Johnson, and G. P. Garnett, “Scale-free networks and sexually transmitted diseases: a description of observed patterns of sexual contacts in Britain and Zimbabwe,” *Sexually transmitted diseases*, vol. 31, pp. 388–390, June 2004.

- [4] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, “Scale-free networks from varying vertex intrinsic fitness,” *Physical Review Letters*, vol. 89, December 2002.
- [5] V. C. Barbosa, R. Donangelo, and S. R. Souza, “Emergence of scale-free networks from local connectivity and communication trade-offs,” *Physical Review E*, vol. 74, pp. 016113 – 016117, 2006.
- [6] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, pp. 509–512, October 1999.
- [7] P. L. Krapivsky, S. Redner, and F. Leyvraz, “Connectivity of growing random networks,” *Phys. Rev. Lett.*, vol. 85, pp. 4629–4632, Nov 2000.
- [8] P. Holme and B. J. Kim, “Growing scale-free networks with tunable clustering,” *Phys. Rev. E*, vol. 65, p. 026107, Jan 2002.
- [9] C. Cooper and A. Frieze, “A general model of web graphs,” *Random Structures and Algorithms*, vol. 22, no. 3, pp. 311 – 335, 2003.
- [10] S. Dorogovtsev, J. Mendes, and A. Samukhin, “Structure of growing networks: Exact solution of the barabási-albert’s model,” *Physical Review Letters*, vol. 85, 2000.
- [11] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, “Structure of growing networks with preferential linking,” *Phys. Rev. Lett.*, vol. 85, pp. 4633–4636, Nov 2000.
- [12] A. Barabási, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek, “Evolution of the social network of scientific collaborations,” *Physica A*, vol. 311, pp. 590–614, 2002.
- [13] S. Dorogovtsev and J. Mendes, “Evolution of reference networks with aging,” *Physical Review E*, vol. 62, pp. 1842–1845, 2000.
- [14] S. D. Fretwell and H. L. Lucas, “On territorial behavior and other factors influencing distribution in birds,” *Acta Biotheoretica*, vol. 19, pp. 16–36, 1970.
- [15] M. Astrom, “Travel cost and the ideal free distribution,” *OIKOS*, vol. 69, pp. 516–519, 1994.
- [16] J. Finke and K. Passino, “Stable emergent heterogeneous agent distributions in noisy environments,” in *Proceeding of the American Control Conference*, (Minneapolis, MN), pp. 2130 – 2135, June 2006.
- [17] J. Finke and K. M. Passino, “Stable cooperative vehicle distributions for surveillance,” to appear in the *ASME Journal of Dynamic Systems, Measurement, and Control*, 2007.
- [18] R. Korona, “Travel cost and the ideal free distribution of ovipositing female flour beetles, *Tribolium confusum*,” *Animal Behavior*, vol. 40, pp. 186–187, 1990.
- [19] M. V. Abrahams, “Patch choice under perceptual constraints: A case for departures from an ideal free distribution,” *Behavioral Ecology and Sociobiology*, vol. 70, pp. 999–1007, 1986.
- [20] T. Tregenza, “Building on the ideal free distribution,” *Advances in Ecological Research*, vol. 26, pp. 253–307, 1995.
- [21] J. Finke and K. M. Passino, “Stable cooperative multiagent spatial distributions,” in *Proceedings of the IEEE Conference on Decision and Control and the European Control Conference*, (Seville, Spain), pp. 3566–3571, December 2005.
- [22] R. Cressmann and V. Krivan, “Migration dynamics for the ideal free distribution,” *The American Naturalist*, vol. 168, pp. 384–397, September 2006.
- [23] N. Quijano and K. Passino, “Optimality and stability of the ideal free distribution with application to temperature control,” in *Proceeding of the American Control Conference*, (Minneapolis, MN), pp. 4836 – 4841, June 2006.

- [24] N. Quijano and K. M. Passino, "The ideal free distribution: Theory and engineering application." To appear, *IEEE Transactions on Systems, Man, and Cybernetics Part B*, 2006.
- [25] J. Maynard-Smith, *Evolution and the Theory of Games*. Cambridge, UK: Cambridge: Cambridge University Press, 1982.
- [26] J. W. Weibull, *Evolutionary Game Theory*. London, England: The MIT Press, 1997.
- [27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: Wiley-Interscience, 1991.
- [28] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific Press, 1995.
- [29] R. Albert and A.-L. Barabási, "Scale-free networks," *Reviews of Modern Physics*, vol. 74, pp. 47–97, January 2002.
- [30] S. Valverde, R. F. Canchoand, and R. V. Solé, "Scale-free networks from optimal design," *Europhysics Letters*, vol. 60, no. 4, pp. 512–517, 2002.