Competition-Induced Preferential Attachment

N. Berger[1], C. Borgs[1], J. T. Chayes[1], R. M. D'Souza[1], and R. D. Kleinberg[2]

[1] Microsoft Research, One Microsoft Way, Redmond WA 98052, USA

[2] M.I.T. CSAIL, 77 Massachusetts Ave, Cambridge MA 02139, USA. Supported by a Fannie and John Hertz Foundation Fellowship.

Abstract. Models based on preferential attachment have had much success in reproducing the power law degree distributions which seem ubiquitous in both natural and engineered systems. Here, rather than assuming preferential attachment, we give an explanation of how it can arise from a more basic underlying mechanism of competition between opposing forces.

We introduce a family of one-dimensional geometric growth models, constructed iteratively by locally optimizing the tradeoffs between two competing metrics. This family admits an equivalent description as a graph process with no reference to the underlying geometry. Moreover, the resulting graph process is shown to be preferential attachment with an upper cutoff. We rigorously determine the degree distribution for the family of random graph models, showing that it obeys a power law up to a finite threshold and decays exponentially above this threshold.

We also introduce and rigorously analyze a generalized version of our graph process, with two natural parameters, one corresponding to the cutoff and the other a "fertility" parameter. Limiting cases of this process include the standard preferential attachment model (introduced by Price and by Barabási-Albert) and the uniform attachment model. In the general case, we prove that the process has a power law degree distribution up to a cutoff, and establish monotonicity of the power as a function of the two parameters.

1 Introduction

1.1 Network Growth Models

There is currently tremendous interest in understanding the mathematical structure of networks – especially as we discover how pervasive network structures are in natural and engineered systems. Much recent theoretical work has been motivated by measurements of real-world networks, indicating they have certain "scale-free" properties, such as a power-law distribution of degree sequences. For the Internet graph, in particular, both the graph of routers and the graph of autonomous systems (AS) seem to obey power laws [14, 15]. However, these observed power laws hold only for a limited range of degrees, presumably due to physical constraints and the finite size of the Internet.

Many random network growth models have been proposed which give rise to power law degree distributions. Most of these models rely on a small number of basic mechanisms, mainly preferential attachment[3]

[3] As Aldous [3] points out, proportional attachment may be a more appropriate name, stressing the linear dependence of the attractiveness on the degree.

[19, 4] or copying [17], extending ideas known for many years [12, 20, 22, 21] to a network context. Variants of the basic preferential attachment mechanism have also been proposed, and some of these lead to changes in the values of the exponents in the resulting power laws. For extensive reviews of work in this area, see Albert and Barabási [2], Dorogovtsev and Mendes [11], and Newman [18]; for a survey of the rather limited amount of mathematical work see [7]. Most of this work concerns network models without reference to an underlying geometric space. Nor do most of these models allow for heterogeneity of nodes, or address physical constraints on the capacity of the nodes. Thus, while such models may be quite appropriate for geometry-free networks, such as the web graph, they do not seem to be ideally suited to the description of other observed networks, *e.g.*, the Internet graph.

In this paper, instead of assuming preferential attachment, we show that it can arise from a more basic underlying process, namely competition between opposing forces. The idea that power laws can arise from competing effects, modeled as the solution of optimization problems with complex objectives, was proposed originally by Carlson and Doyle [9]. Their "highly optimized tolerance" (HOT) framework has reliable design as a primary objective. Fabrikant, Koutsoupias and Papadimitriou (FKP) [13] introduce an elegant network growth model with such a mechanism, which they called "heuristically optimized trade-offs". As in many growth models, the FKP network is grown one node at a time, with each new node choosing a previous node to which it connects. However, in contrast to the standard preferential attachment types of models, a key feature of the FKP model is the underlying geometry. The nodes are points chosen uniformly at random from some region, for example a unit square in the plane. The trade-off is between the geometric consideration that it is desirable to connect to a nearby point, and a networking consideration, that it is desirable to connect to a node that is "central" in the network as a graph. Centrality is measured by using, for example, the graph distance to the initial node. The model has a tunable, but fixed, parameter, which determines the relative weights given to the geometric distance and the graph distance.

The suggestion that competition between two metrics could be an alternative to preferential attachment for generating power law degree distributions represents an important paradigm shift. Though FKP introduced this paradigm for network growth, and FKP networks have many interesting properties, the resulting distribution is not a power law in the standard sense [5]. Instead the overwhelming majority of the nodes are leaves (degree one), and a second substantial fraction, heavily connected "stars" (hubs), producing a node degree distribution which has clear bimodal features.[4]

Here, instead of directly producing power laws as a consequence of competition between metrics, we show that such competition can give rise to a preferential attachment mechanism, which in turn gives rise to power laws. Moreover, the power laws we generate have an upper cutoff, which is more realistic in the context of many applications.

---

[4] In simulations of the FKP model, this can be clearly discerned by examining the probability distribution function (pdf); for the system sizes amenable to simulations, it is less prominent in the cumulative distribution function (cdf).

1.2 Overview of Competition-Induced Preferential Attachment

We begin by formulating a general competition model for network growth. Let $x_0, x_1, \ldots, x_t$ be a sequence of random variables with values in some space Λ. We think of the points $x_0, x_1, \ldots, x_t$ arriving one at a time according to some stochastic process. For example, we typically take Λ to be a compact subset of $\mathbb{R}^d$, x_0 to be a given point, say the origin, and $x_1, \ldots, x_t$ to be i.i.d. uniform on Λ. The network at time t will be represented by a graph, $G(t)$, on $t+1$ vertices, labeled $0, 1, \ldots, t$, and at each time step, the new node attaches to one or several nodes in the existing network. For simplicity, here we assume that each new node connects to a single node, resulting in $G(t)$ being a tree.

Given $G(t-1)$, the new node, labeled t, attaches to that node j in the existing network that minimizes a certain cost function representing the trade-off of two competing effects, namely connection or startup cost, and routing or performance cost. The connection cost is represented by a metric, $g_{ij}(t)$, on $\{0, \ldots, t\}$ which depends on $x_0, \ldots, x_t$, but not on the current graph $G(t-1)$, while the routing cost is represented by a function, $h_j(t-1)$, on the nodes which depends on the current graph, but not on the physical locations $x_0, \ldots, x_t$ of the nodes $0, \ldots, t$. This leads to the cost function

$$c_t = \min_j \left[\alpha g_{tj}(t) + h_j(t-1) \right], \tag{1}$$

where α is a constant which determines the relative weighting between connection and routing costs. We think of the function $h_j(t-1)$ as measuring the centrality of the node j; for simplicity, we take it to be the hop distance along the graph $G(t-1)$ from j to the root 0.

To simplify the analysis of the random graph process, we will assume that nodes always choose to connect to a point which is closer to the root, i.e. they minimize the cost function

$$\tilde{c}_t = \min_{j : \|x_j\| < \|x_t\|} \left[\alpha g_{tj}(t) + h_j(t-1) \right], \tag{2}$$

where $\| \cdot \|$ is an appropriate norm.

In the original FKP model, Λ is a compact subset of $\mathbb{R}^2$, say the unit square, and the points x_i are independently uniformly distributed on Λ. The cost function is of the form (1), with $g_{ij} = d_{ij}$, the Euclidean metric (modeling the cost of building the physical transmission line), and $h_j(t)$ is the hop distance along the existing network $G(t)$ from j to the root. A rigorous analysis of the degree distribution of this two-dimensional model was given in [5], and the analogous one-dimensional problem was treated in [16].

Our model is defined as follows.

Definition 1 (Border Toll Optimization Process) *Let $x_0 = 0$, and let $x_1, x_2, \ldots$ be i.i.d., uniformly at random in the unit interval $\Lambda = [0, 1]$, and let $G(t)$ be the following process: At $t = 0$, $G(t)$ consists of a single vertex 0, the root. Let $h_j(t)$ be the hop distance to 0 along $G(t)$, and let $g_{ij}(t) = n_{ij}(t)$ be the number of existing nodes between x_i and x_j at time t, which we refer to as the* jump cost *of i connecting to j. Given $G(t-1)$ at time $t-1$, a new vertex, labeled t, attaches to the node j which minimizes the cost function (2). Furthermore, if there are several nodes j that minimize this cost function and satisfy the constraint, we choose the one whose position x_j is nearest to x_t. The process so defined is called the* border toll optimization process *(BTOP).*

As in the FKP model, the routing cost is just the hop distance to the root along the existing network. However, in our model the connection cost metric measures the number of "borders" between two nodes: hence the name BTOP. Note the correspondence to the Internet, where the principal connection cost is related to the number of AS domains crossed – representing, *e.g.*, the overhead associated with BGP, monetary costs of peering agreements, etc. In order to facilitate a rigorous analysis of our model, we took the simpler cost function (2), so that the new node always attaches to a node to its left.

It is interesting to note that the ratio of the BTOP connection cost metric to that of the one-dimensional FKP model is just the local density of nodes: $n_{ij}/d_{ij} = \rho_{ij}$. Thus the transformation between the two models is equivalent to replacing the constant parameter α in the FKP model with a variable parameter $\alpha_{ij} = \alpha\rho_{ij}$ which changes as the network evolves in time. That α_{ij} is proportional to the local density of nodes in the network reflects a model with an increase in cost for local resources that are scarce or in high demand. Alternatively, it can be thought of as reflecting the economic advantages of being first to market.

Somewhat surprisingly, the BTOP is equivalent to a special case of the following process, which closely parallels the preferential attachment model and makes no reference to any underlying geometry.

Definition 2 (Generalized Preferential Attachment with Fertility and Aging) *Let A_1, A_2 be two positive integer-valued parameters. Let $G(t)$ be the following Markov process, whose states are finite rooted trees in which each node is labeled either* fertile *or* infertile. *At time $t = 0$, $G(t)$ consists of a single fertile vertex. Given the graph at time t, the new graph is formed in two steps: first, a new vertex, labeled $t+1$ and initialized as infertile, connects to an old vertex j with probability zero if j is infertile, and with probability*

$$Pr(t+1 \to j) = \frac{\min\{d_j(t), A_2\}}{W(t)} \tag{3}$$

if j is fertile. Here, $d_j(t)$ is equal to 1 plus the out-degree of j, and $W(t) = \sum_j' \min\{d_j(t), A_2\}$ with the sum running over fertile vertices only. We refer to vertex $t+1$ as a child of j. If after the first step, j has more than $A_1 - 1$ infertile children, one of them, chosen uniformly at random, becomes fertile. The process so defined is called a generalized preferential attachment process with fertility threshold A_1 and aging threshold A_2.

The special case $A_1 = A_2$ is called the competition-induced preferential attachment process *with parameter A_1.*

The last definition is motivated by the following theorem, to be proved in Section 2. To state the theorem, we define a graph process as a random sequence of graphs $G(0), G(1), G(2), \ldots$ on the vertex sets $\{0\}, \{0,1\}, \{0,1,2\}, \ldots$, respectively.

Theorem 1 *As a graph process, the border toll optimization process has the same distribution as the competition-induced preferential attachment process with parameter $A = \lceil \alpha^{-1} \rceil$.*

Certain other limiting cases of the generalized preferential attachment process are worth noting. If $A_1 = 1$ and $A_2 = \infty$, we recover the standard model of preferential attachment [19, 4]. If $A_1 = 1$ and A_2 is finite, the model is equivalent to the standard model of preferential attachment with a cutoff. On the other hand, if $A_1 = A_2 = 1$, we get a uniform attachment model.

The degree distribution of our random trees is characterized by the following theorem, which asserts that almost surely (a.s.) the fraction of vertices having degree k converges to a specified limit q_k, and moreover that this limit obeys a power law for $k < A_2$, and decays exponentially above A_2.

Theorem 2 *Let A_1, A_2 be positive integers and let $\tilde{G}(t)$ be the generalized preferential attachment process with fertility parameter A_1 and aging parameter A_2. Let $N_0(t)$ be the number of infertile vertices at time t, and let $N_k(t)$ be the number of fertile vertices with $k-1$ children at time t, $k \geq 1$. Then:*

1. *There are numbers $q_k \in [0,1]$ such that, for all $k \geq 0$*

$$\frac{N_k(t)}{t+1} \to q_k \quad \text{a.s., as } t \to \infty. \tag{4}$$

2. *There exists a number $w = w(A_1, A_2) \in [0,2]$ such that the q_k are determined by the following equations:*

$$q_i = \left(\prod_{k=2}^{i} \frac{k-1}{k+w} \right) q_1 \quad \text{if} \quad i \leq A_2, \tag{5}$$

$$q_i = \left(\frac{A_2}{A_2+w} \right)^{i-A_2} q_{A_2} \quad \text{if} \quad i > A_2 \tag{6}$$

$$1 = \sum_{i=0}^{\infty} q_i, \qquad \text{and} \qquad q_0 = \sum_{i=1}^{\infty} q_i \min\{i-1, A_1-1\}.$$

3. *There are positive constants c_1 and C_1, independent of A_1 and A_2, such that*

$$c_1 k^{-(w+1)} < q_k/q_1 < C_1 k^{-(w+1)} \tag{7}$$

   for $1 \leq k \leq A_2$.

4. *If $A_1 = A_2$, the parameter w is equal to 1, and for general A_1 and A_2, it decreases with increasing A_1, and increases with increasing A_2.*

Equation (7) clearly defines a power law degree distribution with exponent $\gamma = w + 1$ for $k \leq A_2$. Note that for measurements of the Internet the value of the exponent for the power law is $\gamma \approx 2$. In our border toll optimization model, where $A_1 = A_2$, we recover $\gamma = 2$.

The convergence claim of Theorem 2 is proved using a novel method which we believe is one of the main technical contributions of this work. For preferential attachment models which have been analyzed in the past [1, 6, 8, 10], the convergence was established using the Azuma-Hoeffding martingale inequality. To establish the bounded-differences hypothesis required by that inequality, those proofs employed a clever coupling of the random decisions made by the various edges, such that the decisions made by an edge e only influence the decisions of subsequent edges which choose to imitate e's choices. A consequence of this coupling is that if e made a different decision, it would alter the degrees of only finitely many vertices. This in turn allows the required bounded-differences hypothesis to be established. No such approach is available for our models, because the coupling fails. The random decisions made by an edge e may influence the time at which some node v crosses the fertility or aging threshold, which thereby exerts a subtle influence on the decisions of *every* future edge, not only those which choose to imitate e.

Instead we introduce a new approach based on the second moment method. The argument establishing the requisite second-moment upper bound is quite subtle; it depends on a computation involving the eigenvalues of a matrix describing the evolution of the degree sequence in a continuous-time version of the model. The details are presented in the full version of this paper. Here we consider only the evolution of the *expected* degree sequence, see Sec. 3.

2 Equivalence of the two models

2.1 Basic properties of the border toll optimization process

In this section we will turn to the BTOP defined in the introduction, establishing some basic properties which will enable us to prove that it is equivalent to the competition-induced preferential attachment model. In order to avoid complications we exclude the case that some of the x_i's are identical, an event that has probability zero. We say that $j \in \{0, 1 \ldots, t\}$ lies to the right of $i \in \{0, 1 \ldots, t\}$ if $x_i < x_j$, and we say that j lies directly to the right of i if $x_i < x_j$ but there is no $k \in \{1, \ldots, t\}$ such that $x_i < x_k < x_j$. In a similar way, we say that j is the first vertex with a certain property to the right of i if j has that property and there exists no $k \in \{1, \ldots, t\}$ such that $x_i < x_k < x_j$ and k has the property in question. Similar notions apply with "left" in place of "right". The following definition and lemma are illustrated in Fig. 1.

Definition 3 *A vertex i is called* fertile *at time t if a hypothetical new point arriving at time $t+1$ and landing directly to the right of x_i would attach itself to the node i. Otherwise i is called* infertile *at time t.*

Lemma 1. *Let $0 < \alpha < \infty$, let $A = \lceil \alpha^{-1} \rceil$, and let $0 < t < \infty$. Then*

i) The node 0 is fertile at time t.

ii) Let i be fertile at time t. If i is the rightmost fertile vertex at time t (case 1), let ℓ be the number of infertile vertices to the right of i. Otherwise (case 2), let j be the next fertile vertex to the right of i, and let $\ell = n_{ij}(t)$. Then $0 \leq \ell \leq A-1$, and the ℓ infertile vertices located directly to the right of i are children of i. In case 2, if $h_j > h_i$, then j is a fertile child of i and $\ell = A-1$. As a consequence, the hop count between two consecutive fertile vertices never increases by more than 1 as we move to the right, and if it increases by 1, there are $A-1$ infertile vertices between the two fertile ones.

iii) Assume that the new vertex at time $t+1$ lands between two consecutive fertile vertices i and j, and let $\ell = n_{ij}(t)$. Then $t+1$ becomes a child of i. If $\ell + 1 < A$, the new vertex is infertile at time $t+1$, and the fertility of all old vertices is unchanged. If $\ell + 1 = A$ and the new vertex lies directly to the left of j, the new vertex is fertile at time $t+1$ and the fertility of the old vertices is unchanged. If $\ell + 1 = A$ and the new vertex does not lie directly to the left of j, the new vertex is infertile at time $t+1$, the vertex directly to the left of j becomes fertile, and the fertility of all other vertices is unchanged.

iv) If $t+1$ lands to the right of the rightmost fertile vertex at time t, the statements in iii) hold with j replaced by the right endpoint of $[0,1]$, and $n_{ij}(t)$ replaced by the number of vertices to the right of i.

v) If i is fertile at time t, it is still fertile at time $t+1$.

vi) If i has k children at time t, the $\ell = \min\{A-1, k\}$ leftmost of them are infertile at time t, and any others are fertile.

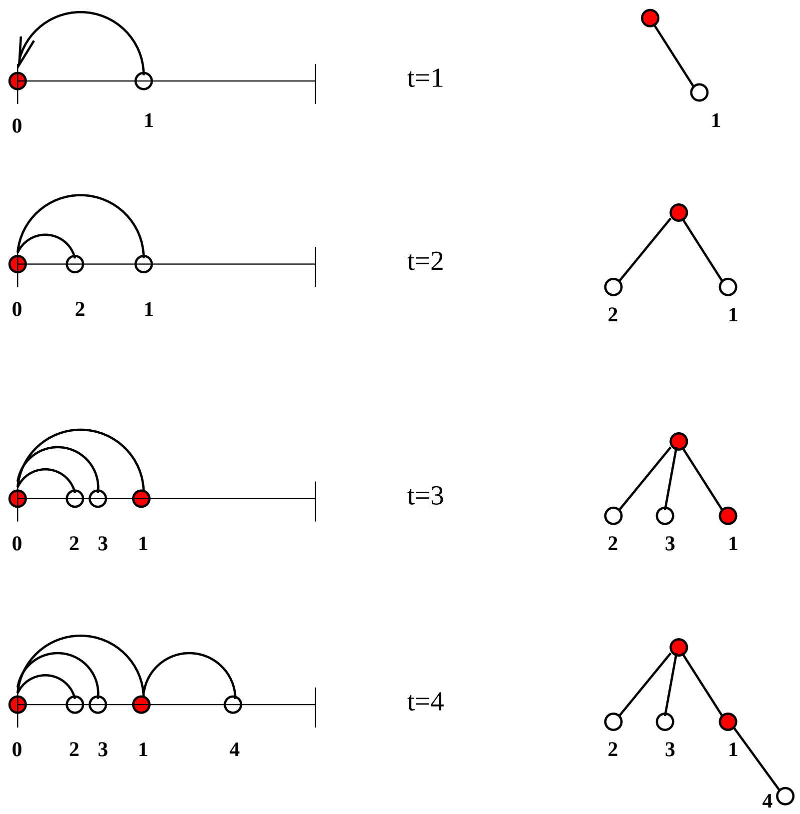

Fig. 1. A sample instance of BTOP for $A = 3$, showing the process on the unit interval (on the left), and the resulting tree (on the right). Fertile vertices are shaded, infertile ones are not. Note that vertex 1 became fertile at $t = 3$.

Proof. Statement i) is trivial, statement v) follows immediately from iii) and iv), and vi) follows immediately from ii). So we are left with ii) — iv). We proceed by induction on t. If ii) holds at time t, and iii) and iv) hold for a new vertex arriving at time $t + 1$, ii) clearly also holds at time $t + 1$. We therefore only have to prove that ii) at time t implies iii) and iv) for a new vertex arriving at time $t + 1$. Using, in particular, the last statement of ii) as a key ingredient, the proof is straightforward but lengthy. This will appear in the full version of the paper. □

2.2 Proof of Theorem 1

In the BTOP, note that our cost function $\min_j [\alpha n_{tj}(t) + h_j(t-1)]$, and hence the graph $G(t)$, only depends on the order of the vertices $x_0, \ldots, x_t$, and not on their actual positions in the interval $[0, 1]$. Let $\boldsymbol{\pi}(t)$ be the permutation of $\{0, 1, \ldots, t\}$ which orders the vertices $x_0, \ldots, x_t$ from left to right, so that

$$x_0 = x_{\pi_0(t)} < x_{\pi_1(t)} < \cdots < x_{\pi_t(t)}. \tag{8}$$

(Recall that the vertices $x_0, x_1, \ldots, x_t$ are pairwise distinct with probability one.) We can consider a change of variables, from the x's to the length of the intervals between successive ordered vertices:

$$s_i(t) \equiv x_{\pi_{i+1}(t)} - x_{\pi_i(t)} \quad \text{if} \quad 0 \le i \le t-1 \quad \text{and} \quad s_t(t) = 1 - x_{\pi_t(t)}. \tag{9}$$

The lengths then obey the constraint: $\sum_{i=0}^{t} s_i = 1$. The set of interval lengths, $\boldsymbol{s}(t)$ together with the set of permutation labels $\boldsymbol{\pi}(t) = (\pi_0(t), \pi_1(t), \ldots, \pi_t(t))$ is an equivalent representation to the original set of position variables, $\boldsymbol{x}(t)$.

Let us consider the process $\{\boldsymbol{\pi}(t)\}_{t\geq 1}$. It is not hard to show that this process is a Markov process, with the initial permutation being the trivial permutation given by $\pi_i(1) = i$, and the permutation at time $t+1$ obtained from $\boldsymbol{\pi}(t)$ by inserting the new point $t+1$ into a uniformly random position. More explicitly, the permutation $\boldsymbol{\pi}(t+1)$ is obtained from $\boldsymbol{\pi}(t)$ by choosing $i_o \in \{1, \ldots, t+1\}$ uniformly at random, and setting

$$\pi_i(t+1) = \begin{cases} \pi_i(t) & \text{if} \quad i \leq i_0 \\ t+1 & \text{if} \quad i = i_0 \\ \pi_{i-1}(t) & \text{if} \quad i > i_0. \end{cases} \tag{10}$$

Indeed, let $I_k(t) = [x_{\pi_k(t)}, x_{\pi_{k+1}(t)}]$, and consider for a moment the process $(\boldsymbol{\pi}(t), \boldsymbol{s}(t))$. Then the conditional probability that the next point arrives in the k-th interval, I_k, depends only on the interval length at time t:

$$\begin{aligned} &Pr\left[x_{t+1} \in I_k \,|\boldsymbol{\pi}(t), \boldsymbol{s}(t), \boldsymbol{\pi}(t-1), \boldsymbol{s}(t-1), \ldots, \boldsymbol{\pi}(0), \boldsymbol{s}(0)\right] \\ &\qquad\qquad = Pr\left[x_{t+1} \in I_k \,|\boldsymbol{\pi}(t), \boldsymbol{s}(t)\right] = s_k(t). \end{aligned} \tag{11}$$

Integrating out the dependence on the interval length from the above equation we get:

$$\begin{aligned} Pr\left[x_{t+1} \in I_k \,|\boldsymbol{\pi}(t)\right] &= \int Pr\left[x_{t+1} \in I_k \,|\boldsymbol{\pi}(t), \boldsymbol{s}(t)\right] dP(\boldsymbol{s}(t)) \\ &= \int s_k(t) dP(\boldsymbol{s}(t)) = \frac{1}{t+1}, \end{aligned} \tag{12}$$

since after the arrival of t points, there exist $(t+1)$ intervals, and by symmmetry they have equal expected length. Thus the probability that the next point arrives in the k-th interval is uniform over all the intervals, proving that $\boldsymbol{\pi}(t)$ is indeed a Markov chain with the transition probabilities described above.

With the help of Lemma 1, we now easily derive a description of the graph $G(t)$ which does not involve any optimization problem. To this end, let us consider a vertex i with ℓ infertile children at time t. If a new vertex falls into the interval directly to the right of i, or into one of the intervals directly to the right of an infertile child of i, it will connect to the vertex i. Since there is a total of $t+1$ intervals at time t, the probability that a vertex i with ℓ infertile children grows an offspring is $(\ell+1)/(t+1)$. By Lemma 1 (vi), this number is equal to $\min\{A, k_i\}/(t+1)$, where $k_i - 1$ is the number of children of i. Note that fertile children do not contribute to this probability, since vertices falling into an interval directly to the right of a fertile child will connect to the child, not the parent.

Assume now that i did get a new offspring, and that it had $A-1$ infertile children at time t. Then the new vertex is either born fertile, or makes one of its infertile siblings fertile. Using the principle of deferred decisions, we may assume that with probability $1/A$ the new vertex becomes fertile, and with probability $(A-1)/A$ an old one, chosen uniformly at random among the $A-1$ candidates, becomes fertile.

This finishes the proof of Theorem 1.

3 Convergence of the Degree Distribution

3.1 Overview

To characterize the behavior of the degree distribution, we derive a recursion which governs the evolution of the vector $\boldsymbol{N}(t)$, whose components are the number of vertices of each degree, at the time when there are t nodes in the network. The conditional expectation of $\boldsymbol{N}(t+1)$ is given by an evolution equation of the form

$$\mathbb{E}\left(\boldsymbol{N}(t+1) - \boldsymbol{N}(t) \mid \boldsymbol{N}(t)\right) = M(t)\boldsymbol{N}(t),$$

where $M(t)$ depends on t through the random variable $W(t)$ introduced in Definition 2. Due to the randomness of the coefficient matrix $M(t)$, the analysis of this evolution equation is not straightforward. We avoid this problem by introducing a continuous-time process, with time parameter τ, which is equivalent to the original discrete-time process up to a (random) reparametrization of the time coordinate. The evolution equation for the conditional expectations in the continuous-time process involves a coefficient matrix M that is not random and does not depend on τ. We will first prove that the *expected* degree distribution in the continuous-time model converges to a scalar multiple of the eigenvector $\hat{p}$ of M associated with the largest eigenvalue w. This is followed by the much more difficult proof that the *empirical* degree distribution converges a.s. to the same limit. Finally, we translate this continuous-time result into a rigorous convergence result for the original discrete-time system.

The key observation is that, in this continuous-time model, the number of vertices of degree k, $\widehat{N}_k(\tau)$, grows exponentially at a rate determined by the largest eigenvalue of this matrix, w, while the difference $q_j\widehat{N}_k(\tau) - q_k\widehat{N}_j(\tau)$ has an exponential growth rate which is at most the second eigenvalue; for the matrix in question this is strictly less than w. This guarantees that the ratio $\widehat{N}_k(\tau)/\widehat{N}_j(\tau)$ will converge almost surely to q_k/q_j, for all k and j. The convergence of the normalized degree sequence to the vector $(q_i)_{i=0}^{\infty}$ in the continuous-time model follows easily from this. We then translate this continuous-time result into a rigorous convergence result for the original discrete-time system.

3.2 Notation

Let $A \geq \max(A_1, A_2)$. Let $N_0(t)$ be the number of infertile vertices at (discrete) time t, and, for $k \geq 1$, let $N_k(t)$ be the number of fertile vertices with $k-1$ children at time t. Let $\widetilde{N}_A(t) = N_{\geq A}(t) = \sum_{k\geq A} N_k(t)$, let $\widetilde{N}_k(t) = N_k(t)$ if $k < A$, and let $W(t) = \sum_{k=1}^{A} \min\{k, A_2\}\widetilde{N}_k(t)$ be the combined attractiveness of all vertices. Let $n_k(t) = \frac{1}{t+1}N_k(t)$ and $\tilde{n}_k(t) = \frac{1}{t+1}\widetilde{N}_k(t)$. Finally, the vectors $(\widetilde{N}_k(t))_{k=1}^{A}$ and $(\tilde{n}_k(t))_{k=1}^{A}$ are denoted by $\widetilde{N}(t)$ and $\tilde{n}(t)$ respectively. Note that the index k runs from 1 to A, not 0 to A.

3.3 Evolution of the expected value

From the definition of the generalized preferential attachment model, it is easy to derive the probabilities for the various alternatives which may happen upon the arrival of the $(t+1)$-st node:

– With probability $A_2\widetilde{N}_A(t)/W(t)$, it attaches to a node of degree $\geq A$. This increments $\widetilde{N}_1$, and leaves $\widetilde{N}_A$ and all $\widetilde{N}_j$ with $1 < j < A$ unchanged.

– With probability $\min(A_2, k)\widetilde{N}_k(t)/W(t)$, it attaches to a node of degree k, where $1 \le k < A$. This increments $\widetilde{N}_{k+1}$, decrements $\widetilde{N}_k$, increments $\widetilde{N}_0$ or $\widetilde{N}_1$ depending on whether $k < A_1$ or $k \ge A_1$, and leaves all other $\widetilde{N}_j$ with $j < A$ unchanged.

It follows that the discrete-time process $(\widetilde{N}_k(t))_{k=0}^A$ at time t is equivalent to the state of the following continuous-time stochastic process $(\widehat{N}_k(\tau))_{k=0}^A$ at the random stopping time $\tau = \tau_t$ of the t-th event.

– With rate $A_2\widehat{N}_A(\tau)$, $\widehat{N}_1$ increases by 1.
– For every $0 < k < A$, with rate $\widehat{N}_k(\tau)\min(k, A_2)$, the following happens:

$$\widehat{N}_k \to \widehat{N}_k - 1 \quad ; \quad \widehat{N}_{k+1} \to \widehat{N}_{k+1} + 1 \quad ; \quad \widehat{N}_{g(k)} \to \widehat{N}_{g(k)} + 1$$

where $g(k) = 0$ for $k < A_1$ and $g(k) = 1$ otherwise.

Note that the above rules need to be modified if $A_1 = 1$; here the birth of a child of a degree-one vertex does not change the net number of fertile degree-one vertices, N_1. Let M be the following $A \times A$ matrix:

$$M_{i,j} = \begin{cases} -1 & \text{if } i = j = 1 < A_1 \\ -\min(j, A_2) & \text{if } \ 2 \le i = j \le A - 1 \\ \min(j, A_2) & \text{if } \ 2 \le i = j + 1 \le A \\ \min(j, A_2) & \text{if } \ i = 1 \text{ and } j \ge \max(A_1, 2) \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

Then, for the continuous time process, for every $\tau > \sigma$, the conditional expectations of the vector $\widehat{N}(\tau) = (\widehat{N}_k(\tau))_{k=1}^A$ are given by

$$\mathbb{E}\left(\widehat{N}(\tau)|\widehat{N}(\sigma)\right) = e^{(\tau-\sigma)M}\widehat{N}(\sigma). \tag{14}$$

It is easy to see that the matrix e^M has all positive entries, and therefore (by the Perron-Frobenius Theorem) M has a unique eigenvector $\hat{p}$ of ℓ_1-norm 1 having all positive entries. Let w be the eigenvalue corresponding to $\hat{p}$. Then w is real, it has multiplicity 1, and it exceeds the real part of every other eigenvalue. Therefore, for every non-zero vector y with non-negative entries,

$$\lim_{\tau\to\infty} e^{-\tau w}e^{\tau M}y = \langle \hat{a}, y\rangle \hat{p}$$

where $\hat{a}$ is the eigenvector of $M^\intercal$ corresponding to w. Note that $\langle \hat{a}, y\rangle > 0$ because y is non-zero and non-negative, and $\hat{a}$ is positive, again by Perron-Frobenius. Therefore, the vector $\mathbb{E}\left(e^{-\tau w}\widehat{N}(\tau)\right)$ converges to a positive scalar multiple of $\hat{p}$, say $\lambda\hat{p}$, as $\tau \to \infty$. Note that this implies, in particular, that $w > 0$. We can also show that $w \le 2$ by showing that $\|\widehat{N}(\tau)\| = \sum_{k=1}^A \widehat{N}_k(\tau)$ is stochastically dominated by the following process, known as the standard birth process X_τ, for which $\mathbb{E}(X_\tau) \sim e^{2\tau}$: X increases by one with rate $2X$ (a more precise definition with proof of all facts used here will come in the full version).

Intuitively, it should be clear that in the discrete time version, $\mathbb{E}\left(\frac{1}{t+1}\widetilde{N}(t)\right)$ converges to $\lambda\hat{p}$ as well. As it turns out, this does not follow immediately, and we establish it in a somewhat round-about way. After we show almost sure convergence to $\lambda\hat{p}$ in continuous time, almost sure convergence in the discrete time model follows once one shows that a.s., t is finite for all finite τ. Then, the a.s. convergence in the discrete time model yields convergence of the expected value in discrete time.

4 Power law with a cutoff

In the previous section, we saw that for every $A > \max\{A_1, A_2\}$, the limiting proportions up to $A-1$ are $\lambda\hat{p}$ where $\hat{p}$ is the eigenvector corresponding to the highest eigenvalue w of the A-by-A matrix M defined in Eqn. 13. Therefore, the components $p_1, p_2, \dots, p_A$ of the vector $\hat{p}$ satisfy the equation:

$$wp_i = -\min(i, A_2)p_i + \min(i-1, A_2)p_{i-1} \qquad i \geq 2 \tag{15}$$

where the normalization is determined by $\sum_{i=1}^{A} p_i = 1$. From (15) we get that for $i \leq A_2$,

$$p_i = \left(\prod_{k=2}^{i} \frac{k-1}{k+w}\right) p_1 \tag{16}$$

and for $i > A_2$

$$p_i = \left(\frac{A_2}{A_2+w}\right)^{i-A_2} p_{A_2} \tag{17}$$

Clearly, (17) is exponentially decaying. There are many ways to see that (16) behaves like a power-law with degree $1+w$. The simplest would probably be:

$$\begin{aligned}
\frac{p_i}{p_1} &= \left(\prod_{k=2}^{i} \frac{k-1}{k+w}\right) = \exp\left(\sum_{k=2}^{i} \log\left(\frac{k-1}{k+w}\right)\right) \qquad (18)\\
&= \exp\left(\sum_{k=2}^{i}\left(\frac{-1-w}{k+w}\right) + O(1)\right) = \exp\left((-1-w)\left(\sum_{k=2}^{i}(k+w)^{-1}\right) + O(1)\right)\\
&= \exp\left((-1-w)\left(\sum_{k=2}^{i} k^{-1}\right) + O(1)\right) = \exp\left((-1-w)\left(\sum_{k=2}^{i}\log\left(\frac{k+1}{k}\right)\right) + O(1)\right)\\
&= \exp\left((-1-w)\log(i/2) + O(1)\right) = O(1)i^{-1-w}.
\end{aligned}$$

Note that the constants implicit in the $O(\cdot)$ symbols do not depend on A_1, A_2 or i, due the fact that $0 < w \leq 2$. (18) can be stated in the following way:

Proposition 3 *There exist $0 < c < C < \infty$ such that for every A_1, A_2 and $i \leq A_2$, if $w = w(A_1, A_2)$ is as in (15), then*

$$ci^{-1-w} \leq \frac{p_i}{p_1} \leq Ci^{-1-w}. \tag{19}$$

The vector $(q_1, q_2, \dots, q_{A-1})$ is a scalar multiple of the vector $(p_1, p_2, \dots, p_{A-1})$, so equations (5), (6), and (7) in Theorem 2 (and the comment immediately following it) are consequences of equations (16), (17), and (19) derived above. It remains to prove the normalization conditions

$$\sum_{i=0}^{\infty} q_i = 1; \qquad q_0 = \sum_{i=1}^{\infty} q_i \min(i-1, A_1-1)$$

stated in Theorem 2. These follow from the equations

$$\sum_{i=0}^{\infty} N_i(t) = t+1; \qquad N_0(t) = \sum_{i=1}^{\infty} N_i(t)\min(i-1, A_1-1).$$

The first of these simply says that there are $t+1$ vertices at time t; the second equation is proved by counting the number of infertile children of each fertile node.

The proof of the monotonicity properties of w asserted in part 4 of Theorem 2 is deferred to the full version of this paper.

References

1. W. Aiello, F. Chung, and L. Lu. Random evolution of massive graphs. In *Handbook of Massive Data Sets*, pages 97–122. Kluwer, 2002.
2. R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, 2002.
3. D. J. Aldous. A stochastic complex network model. *Electron. Res. Announc. Amer. Math. Soc.*, 9:152–161, 2003.
4. A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
5. N. Berger, B. Bollobás, C. Borgs, J. T. Chayes, and O. Riordan. Degree distribution of the FKP network model. In *International Colloquium on Automata, Languages and Programming*, 2003.
6. B. Bollobás, C. Borgs, J. Chayes, and O. Riordan. Directed scale-free graphs. In *Proceedings of the 14th ACM-SIAM Symposium on Discrete Algorithms*, pages 132–139, 2003.
7. B. Bollobás and O. Riordan. Mathematical results on scale-free random graphs. In *Handbook of Graphs and Networks*, Berlin, 2002. Wiley-VCH.
8. B. Bollobás, O. Riordan, J. Spencer, and G. E. Tusnady. The degree sequence of a scale-free random graph process. *Random Structures and Algorithms*, 18:279–290, 2001.
9. J. M. Carlson and J. Doyle. Highly optimized tolerance: a mechanism for power laws in designed systems. *Phys. Rev. E*, 60:1412, 1999.
10. C. Cooper and A. M. Frieze. A general model of web graphs. In *Proceedings of 9th European Symposium on Algorithms*, pages 500–511, 2001.
11. S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Adv. Phys.*, 51:1079, 2002.
12. F. Eggenberger and G. Pólya. Über die statistik verketteter. *Vorgänge.* Zeitschrift Agnew. Math. Mech., 3:279–289, 1923.
13. A. Fabrikant, E. Koutsoupias, and C.H. Papadimitriou. Heuristically optimized trade-offs: a new paradigm for power laws in the internet. In *International Colloquium on Automata, Languages and Programming*, pages 110–122, 2002.
14. M. Faloutsos, P. Faloutsos, and C. Faloutsos. On the power-law relationships of the Internet topology. *Comput. Commun. Rev.*, 29:251, 1999.
15. R. Govindan and H. Tangmunarunkit. Heuristics for Internet map discovery. In *Proceedings of INFOCOM*, pages 1371–1380, 2000.
16. C. Kenyon and N. Schabanel. Personal communication.
17. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *Proc. 41st IEEE Symp. on Foundations of Computer Science*, pages 57–65, 2000.
18. M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.
19. D. J. de S. Price. A general theory of bibliometric and other cumulative advantage processes. *J. Amer. Soc. Inform. Sci.*, 27:292–306, 1976.
20. H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.
21. G. U. Yule. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis. *Philos. Trans. Roy. Soc. London*, Ser. B 213:21–87, 1924.
22. G. K. Zipf. *Human Behavior and the Principle of Least Effort.* Addison-Wesley, Cambridge,MA, 1949.