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 Barabási-Albert preferential attachment model 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] [19, 4] or copying [17], extending ideas known for many years [12, 20, 22, 21] to a network context. Variants

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

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 the 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.

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

[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).

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.

When Λ is equipped with an appropriate norm $\|\cdot\|$, we can use a simplified algorithm, minimizing the cost only over those points j which are closer to the root than is the new point:

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

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 $W(t) = \sum_j' \min\{d_j(t), A_2\}$, where the sum is only over fertile vertices. Second, 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.

Theorem 1 *The border toll optimization process is equivalent to a 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 Barabási-Albert model of preferential attachment. 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* $k \geq 0$

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

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

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

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

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

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 paper. 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 who 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 attractiveness threshold, which thereby exerts a subtle influence on the decisions of *every* future edge, not only those who 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 key observation is that, in this continuous-time model, the expected number of vertices of each degree grows exponentially at a rate determined by the largest eigenvalue, w, of this matrix, while the variance of the number of vertices of each degree has an exponential growth rate which is at most the second eigenvalue. For the matrix in question, the top eigenvalue has multiplicity 1, thus ensuring that the variance grows more slowly than the mean. We then translate this continuous-time result into a rigorous convergence result for the original discrete-time system.

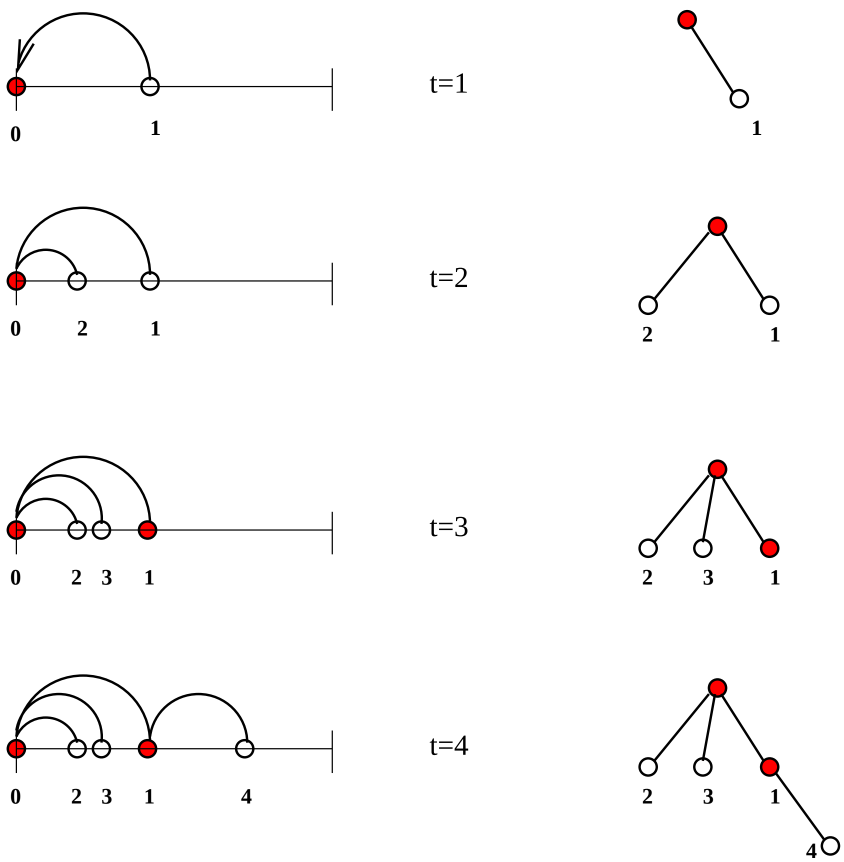

Fig. 1. A sample instance of BTOP for $A = 3$, showing the process on the unit interval (on the left), and the resulting tree (on the right). Fertile vertices are marked red, infertile ones are marked white. Note that vertex 1 became fertile at $t = 3$.

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 $j \in \{0, 1 \dots, t\}$ lies to the right of $i \in \{0, 1 \dots, t\}$ 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 $k \in \{1, \dots, t\}$ 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 $k \in \{1, \dots, t\}$ such that $x_i < x_k < x_j$ and k has the property in question.

Definition 3 *A vertex i is called* fertile at time t *if a new point that arrives at time $t+1$ and lands directly to the right of x_i attaches itself to the node i. Otherwise i is called* infertile *at time t.*

This definition is illustrated in Fig. 1.

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 right most 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 lies not 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 right most fertile vertex at time t, the statements in iii) hold with j replaced by the right endpoint of the interval $[0,1]$, 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 $\ell = \min\{A-1, k\}$ left most of them are infertile at time t, and any others are fertile.

Proof. Statement i) is trivial, and statements v) and vi) follow immediately from iii) and iv), 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. It is not worth reproducing here. The interested reader can find it in Appendix A.

2.2 Proof of Theorem 1

In the BTOP, note that our cost function

$$\min\nolimits_{\mathrm{j}} \left[\alpha n_{tj}(t) + h_j(t-1)\right], \tag{8}$$

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

$$x_0 = x_{\pi_0(t)} < x_{\pi_1(t)} < \cdots < x_{\pi_t(t)}. \tag{9}$$

(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:

$$s_i(t) \equiv x_{\pi_{i+1}(t)} - x_{\pi_i(t)} \quad \text{if} \quad 0 \leq i \leq t-1 \quad \text{and} \quad s_t(t) = 1 - x_{\pi_t(t)}. \tag{10}$$

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 new 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

$$\pi_i(t+1) = \begin{cases} \pi_i(t) & \text{if} \quad i \leq i_0 \\ t+1 & \text{if} \quad i = i_0 \\ \pi_{i-1}(t) & \text{if} \quad i > i_0. \end{cases} \tag{11}$$

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:

$$\begin{aligned} &Pr\left[x_{t+1} \in I_k \,|\boldsymbol{\pi}(t), \boldsymbol{s}(t), \boldsymbol{\pi}(t-1), \boldsymbol{s}(t-1), \ldots, \boldsymbol{\pi}(0), \boldsymbol{s}(0)\right] \\ &\qquad\qquad\qquad = Pr\left[x_{t+1} \in I_k \,|\boldsymbol{\pi}(t), \boldsymbol{s}(t)\right] = s_k(t). \end{aligned} \tag{12}$$

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

$$\begin{aligned} Pr\left[x_{t+1} \in I_k \,|\boldsymbol{\pi}(t)\right] &= \int Pr\left[x_{t+1} \in I_k \,|\boldsymbol{\pi}(t), \boldsymbol{s}(t)\right] dP(\boldsymbol{s}(t)) \\ &= \int s_k(t) dP(\boldsymbol{s}(t)) = \frac{1}{t+1}, \end{aligned} \tag{13}$$

since after the arrival of t points, there exist $(t+1)$ intervals. The probability that the next point arrive 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 don't 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.

We thus have shown that the solution $G(t)$ of the optimization problem (8) can alternatively be described by the competition-induced preferential attachment model with parameter A.

3 Convergence of the Degree Distribution

3.1 Overview

To characterize the behavior of the degree distribution, we will derive a recursion which governs the evolution of the expected number of vertices of each degree, at the time when there are τ nodes in the network. The coefficients of this recursion are random variables depending on $W(\tau)$, the combined attractiveness of all vertices at time τ. To simplify the analysis of the recursion, we introduce a continuous-time model which is equivalent to the original discrete-time model up to a (random) reparametrization of the time coordinate. The kernel of this continuous-time Markov chain is a matrix M whose coefficients we identify explicitly. In this section we will prove that the *expected* degree distribution converges to a scalar multiple of the eigenvector $\hat{p}$ of M associated with the largest eigenvalue w. The much more difficult proof that the *empirical* degree distribution converges a.s. to the same limit is deferred to Appendix B.

3.2 Notation

Let $A \geq \max(A_1, A_2)$. At (discrete) time τ, let $N_0(\tau)$ be the number of infertile vertices at time τ, and, for $k \geq 1$, let $N_k(\tau)$ be the number of fertile vertices with $k-1$ children at time τ. Let $\tilde{N}_A(\tau) = N_{\geq A}(\tau) = \sum_{k \geq A} N_k(\tau)$, and $\tilde{N}_k(\tau) = N_k(\tau)$ if $k < A$. Finally let $\hat{N}_k(\tau) = \frac{1}{\tau+1}\tilde{N}_k(\tau)$, and let $n_k(\tau), \hat{n}_k(\tau)$ to be the expected values of $N_k(\tau), \hat{N}_k(\tau)$, respectively.

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 $(\tau+1)$-st node:

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

It follows that the discrete-time process $\{\tilde{N}_k(\tau)\}_{k=0}^{\infty}$ is equivalent to the state of the following continuous-time stochastic process (with time parameter t) at the time of the τ-th event.

- With rate $\tilde{A}_2 N_A(t)$, $\tilde{N}_1$ increases by 1.
- For every $0 < k < A$, with rate $\tilde{N}_k(t)\min(k, A_2)$, the following happens:

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

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

Let M be the following $A \times A$ matrix:

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

Then, for the continuous time process, for every $t > s$, the conditional expectations of the vector $\tilde{N}(t)$ are given by

$$E\left(\tilde{N}(t)|\tilde{N}(s)\right) = e^{(t-s)M}\tilde{N}(s). \tag{14}$$

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 n with non-negative entries,

$$\lim_{t\to\infty} e^{-tw}e^{tM}n = \langle \hat{a}, n\rangle \hat{p}$$

where $\hat{a}$ is the eigenvector of M^{T} corresponding to w. Note that $\langle \hat{a}, n\rangle > 0$ because n is non-zero and non-negative, and $\hat{a}$ is positive, again by Perron-Frobenius. Therefore, up to a scalar factor, the vector $\hat{n}(t) := E\left(e^{-tw}\tilde{N}(t)\right)$ converges to $\hat{p}$ as $t \to \infty$. Note that this implies, in particular, that $w > 0$. We can also show that $w \le 2$. This is a consequence of Claim 2 in Appendix B, which says that $\hat{N}_k(t)$ is stochastically dominated by a stochastic process X_t satisfying $E(X_t) \sim e^{2t}$.

To conclude that the discrete time version, $\hat{n}(\tau)$, converges to $\hat{p}$ as well, one needs show that, a.s., τ is finite for all finite t. This is done in Claims 1 and 2 in Appendix B.

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 $\hat{p}$ where $\hat{p}$ is the eigenvector corresponding to the highest eigenvalue w of the A-by-A matrix

$$M_{i,j} = \begin{cases} -\min(j, A_2) & \text{if } \ 1 \le i = j \le A-1 \\ \min(j, A_2) & \text{if } \ 1 \le i = j+1 \le A \\ \min(j, A_2) & \text{if } \ i = 1 \text{ and } j > A_1 \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

Therefore, the proportions p satisfy the equation:

$$wp_i = -\min(i, A_2)p_i + \min(i-1, A_2)p_{i-1} \qquad i \ge 2 \tag{16}$$

where the normalization is determined by $\sum_{i=1}^{A} p_i = 1$. From (16) we get that for $i \le A_2$,

$$p_i = \left(\prod_{k=2}^{i} \frac{k-1}{k+w}\right) p_1 \tag{17}$$

and for $i > A_2$

$$p_i = \left(\frac{A_2}{A_2 + w}\right)^{i-A_2} p_{A_2} \tag{18}$$

Clearly, (18) is exponentially decaying. There are many ways to see that (17) behaves like a power-law with degree $1 + w$. The simplest would probably be:

$$\begin{aligned}
\frac{p_i}{p_1} &= \left(\prod_{k=2}^{i} \frac{k-1}{k+w}\right) = \exp\left(\sum_{k=2}^{i} \log\left(\frac{k-1}{k+w}\right)\right) \qquad (19)\\
&= \exp\left(\sum_{k=2}^{i}\left(\frac{-1-w}{k+w}\right) + O(1)\right) = \exp\left((-1-w)\left(\sum_{k=2}^{i}(k+w)^{-1}\right) + O(1)\right)\\
&= \exp\left((-1-w)\left(\sum_{k=2}^{i} k^{-1}\right) + O(1)\right) = \exp\left((-1-w)\left(\sum_{k=2}^{i}\log\left(\frac{k+1}{k}\right)\right) + O(1)\right)\\
&= \exp\left((-1-w)\log(i/2) + O(1)\right) = O(1)i^{-1-w}.
\end{aligned}$$

Note that the constants implicit in the $O(\cdot)$ symbols do not depend on A_1, A_2 or i, due the fact that $0 < w \leq 2$. (19) 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 (16), then*

$$ci^{-1-w} \leq \frac{p_i}{p_1} \leq Ci^{-1-w}. \tag{20}$$

The vector $q = (q_1, q_2, \ldots)$ is a scalar multiple of $\hat{p}$, so equations (5), (6), and (7) in Theorem 2 (and the comment immediately following it) are consequences of equations (17), (18), and (20) 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 monotonicity properties of w asserted in part 4 of Theorem 2 are proved in Appendix C. The concentration of the empirical degree distribution is proved in Appendix B.

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 Structure 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.

A Proof of Lemma 1

In this appendix, we complete the proof of Lemma 1.

To this end, let us first recall that the only non-trivial part is the fact that ii) at time t implies iii) and iv) for a new vertex arriving at time $t+1$. Assume thus that ii) holds at time t.

At time $t+1$, a new vertex arrives, and falls directly to the right of some vertex k. Let i be the nearest vertex to the left of k that was fertile at time t (if k is fertile at time t, we set $i = k$) and let j be the nearest vertex to the right of i that was fertile at time t (we assume for the moment that i is not the right most fertile vertex at time t), let ℓ be the number of vertices between i and j at time t.

Let us first prove that the vertex $t+1$ connects to i. If $i = k$, this is obvious, since i is fertile at time t. We may therefore assume that $k \neq i$. For the new vertex $t+1$, the cost of connecting to the vertex i is then equal to $\alpha(n_{ik}(t)+1)$. Let us first compare this cost to the cost of connecting to a fertile vertex i' to the left of i. Let $i_0 = i'$, let $i_s = i$, and let $i_1, \dots, i_{s-1}$ be the fertile vertices between i' and i, ordered from left to right. If $h_{i_{m-1}} < h_{i_m}$, we use the inductive assumption ii) to conclude that the number of infertile vertices between i_{m-1} and i_m is equal to $A-1$, and $h_{i_{m-1}} = h_{i_m} - 1$. A decrease of q in the hop cost is therefore accompanied by an increase in the jump cost of at least $\alpha A q \geq q$. As a consequence, it never pays to connect to a fertile vertex i' to the left of i. The cost of connecting to an infertile vertex to the left of i is even higher, since the hop count of an infertile vertex is at best equal to the hop count of the next fertile vertex to the right. We therefore only have to consider the connection cost to some of the infertile children of i. But again, the hop count is worse by 1 when compared to the hop count of i, and the jump cost is at best reduced by $(A-1)\alpha < 1$, proving that the cost of connecting to i is minimal.

To discuss the fertility of the vertices in the graph $G(t+1)$, we need to consider the arrival of a second vertex, labeled $t+2$. If $t+2$ falls to the left of $t+1$, it will face an optimization problem that has not been changed by the arrival of the vertex $t+1$, implying that the fertility of the vertices to the left of $t+1$ is unchanged. If $t+2$ falls to the right of j, the cost of connecting to j or one of the vertices to the right of j is the same as before, and the cost of connecting to a vertex to the left of j is at best equal (the cost of connecting to any vertex to the left of $t+1$ is in fact higher, due to the additional cost of jumping over the vertex $t+1$). Therefore, the vertex $t+2$ will still prefer to connect to either j or one of the vertices to the right of j, implying that the fertility of the vertices to the right of j has not changed at all. We therefore are left with analyzing the case where $t+2$ falls between $t+1$ and j. Again, the vertex $t+2$ will prefer i over any vertex to the left of i (the cost analysis is the same as the one used for $t+1$ above), so we just have to compare the costs of connecting to the different vertices between i and j. If $\ell+1 < A$, this will again imply that $t+2$ connect to i; but if $\ell+1 = A$, the vertex $t+2$ will only connect to i if it does not fall to the right of the right most of the now $\ell+1$ vertices between i and j. If it falls to the right of this vertex, it will be as expensive to connect to the right most of the now $\ell+1$ vertices between i and j as it is to connect to i. Recalling out convention of connecting to the nearest vertex to the left if there is a tie in costs, this proves that now $t+2$ connects to the right most vertex between i and j, implying that this vertex is fertile.

The above considerations prove the fertility statements in iii), and thus completes the proof of iii). The case where i is the right most fertile vertex at time t is similar (in fact, it is slightly easier since it involves less cases), and leads to the proof of iv). This completes the proof of Lemma 1

B Concentration of $\hat{N}_k(t)$

B.1 Concentration of the continuous time process

In order to show concentration of the continuous time process, we will prove the following two lemmas:

Lemma 2. *For every $u < w$ and every $1 \leq k \leq A$, a.s. for every t large enough,*

$$\tilde{N}_k(t) > e^{ut}.$$

and

Lemma 3. *There exists $v < w$ s.t. for every $1 \leq k < j \leq A$ a.s. for every t large enough,*

$$p_j \tilde{N}_k(t) - p_k \tilde{N}_j(t) < e^{vt}.$$

In order to prove Lemmas 2 and 3, we need to use some estimates considering the standard birth process described below.

Definition 4 *Let $\{o_n\}_{n=1}^{\infty}$ be independent exponential random variables, so that $\mathbb{E}(o_n) = \frac{1}{2}n^{-1}$. For $t \in [0,\infty)$, let $X_t = \inf\{n : \sum_{k=1}^{n} o_k > t\}$. Then X is called the* **standard birth process***.*

The following claim will be proved in Appendix D.

Claim 1 *X_t is almost surely finite for every t. Furthermore, there exists a constant C_s such that for every $t_2 > t_1$ and x and k,*

$$\mathbf{P}\left(X_{t_2} > kxe^{2(t_2-t_1)} \middle| X_{t_1} = x \right) < \frac{C_s}{x(k-1)^2}. \tag{21}$$

The standard birth process is connected to our discussion through the following easy claim:

Claim 2 *Let $\|\tilde{N}(t)\| = \sum_{k=1}^{A} \tilde{N}_k(t)$. Let $T \geq 0$, let $x \geq y$, and let X be a standard birth process. Then $\{\{X_t\}_{t\geq T} | X_T = x\}$ stochastically dominates $\left\{ \{\|\tilde{N}(t)\|\}_{t\geq T} \middle| \|\tilde{N}(T)\| = y \right\}$.*

Proof. Let $\{r_n\}_{n=1}^{\infty}$ be i.i.d. exponential random variables with mean 1. Then $\sum_{k=1}^{n} o_k$ has the same distribution as $\sum_{k=1}^{n} r_k/2k$. The time at which the n-th node is born has the same distribution as $\sum_{k=1}^{n} r_k/W(k)$, where $W(k)$ denotes the combined attractiveness of all nodes at time k. The claim follows now from the observation that $W(k) \leq 2k$.

Proof (Proof of Lemma 3). We use a martingale to bound the variance. Fix T, and let

$$L_t = \mathbb{E}\left(p_j \tilde{N}_k(T) - p_k \tilde{N}_j(T) \middle| \tilde{N}(t) \right)$$

Clearly, L_t is a (continuous time) martingale. Let $b = b^{(j,k)}$ be the vector

$$b_i = \begin{cases} -p_k & \text{if } i = j \\ p_j & \text{if } i = k \\ 0 & \text{otherwise.} \end{cases}$$

By (14), we know that $L_t = b^{\mathsf{T}} e^{M(T-t)} n(t)$. $J\hat{p} = 0$, and therefore the norm of $Je^{M(T-t)}$ is bounded by $e^{(T-t)v'}$ for some $v' < w$. Without loss of generality, we may assume that $v' > w/2$.

Claim 3

$$var\left(p_j\tilde{N}_k(T) - p_k\tilde{N}_j(T)\right) < C\exp(2v'T) \tag{22}$$

For some constant C.

Proof. Let $0 < \epsilon < \exp(-10T)$ be such that $K = T/\epsilon$ is an integer number. Then, $\{U_k = L_{k\epsilon}\}_{k=0}^{K}$ is a martingale, and

$$\mathrm{var}\left(p_j\tilde{N}_k(T) - p_k\tilde{N}_j(T)\right) = \sum_{k=0}^{K-1}\mathrm{var}(U_{k+1} - U_k)$$

We want to estimate the variance of $(U_{k+1} - U_k)$. Let $v_k = \|\tilde{N}_{(k+1)\epsilon} - \tilde{N}_{k\epsilon}\|$. Clearly,

$$\mathrm{var}(U_{k+1} - U_k) \le \mathrm{var}(v_k)\exp\left[2v'(T - (k+1)\epsilon)\right]$$

Using Claims 1 and 2,

$$\begin{aligned}
\mathrm{var}(v_k) &= \mathbb{E}\left(\mathrm{var}\left(v_k \middle| \tilde{N}_{k\epsilon}\right)\right) + \mathrm{var}\left(\mathbb{E}\left(v_k \middle| \tilde{N}_{k\epsilon}\right)\right) \\
&\le \exp(wk\epsilon)\left(e^{4\epsilon} - 1\right) + \exp(4k\epsilon)\left(e^{2\epsilon} - 1\right)^2 \\
&\le 5\epsilon\exp(wk\epsilon) + 4\epsilon^2\exp(4k\epsilon) < C_0\epsilon\exp(wk\epsilon)
\end{aligned}$$

for $C_0 = 6$, by the choice of ϵ. Therefore,

$$\begin{aligned}
\mathrm{var}\left(p_j\tilde{N}_k(T) - p_k\tilde{N}_j(T)\right) &< C_0\epsilon\sum_{k=0}^{K-1}\exp\left(wk\epsilon + 2v'(T - (k+1)\epsilon)\right) \\
&\le C_0 e^{2v'T}\int_0^T e^{(w-2v')t}dt < C\exp(2v'T)
\end{aligned}$$

for

$$C = C_0\int_0^\infty e^{(w-2v')t}dt < \infty.$$

Choose some v strictly between v' and w in a way that $w - v < 0.25\min(0.1, v - v')$ and let $\delta = \min(0.1, v - v')$. Using Chebyshev's inequality,

$$\mathbf{P}\left(p_j\tilde{N}_k(T) - p_k\tilde{N}_j(T) > \frac{1}{3}e^{vT}\right) \le Ce^{-2\delta T} \tag{23}$$

Let $\{T_i\}_{i=1,2,\ldots}$ be such that $e^{2\delta T_i} = i^2$. By Borel-Cantelli, almost surely there exists i_0 such that for all $i > i_0$,

$$p_j\tilde{N}_k(T_i) - p_k\tilde{N}_j(T_i) < \frac{1}{2}e^{vT_i} \tag{24}$$

We want to show that almost surely for all T large enough,

$$p_j\tilde{N}_k(T) - p_k\tilde{N}_j(T) < e^{vT} \tag{25}$$

We know that $\mathbb{E}(\tilde{N}(T_i)) = O(\exp(wT_i))$, and using a martingale argument similar to the one in Claim 3, we get that $\mathrm{var}(\tilde{N}(T_i)) = O(\exp(2wT_i))$ and therefore

$$\mathbf{P}(\tilde{N}(T_i) > e^{(w+0.6\delta)T_i}) < C_l e^{-1.2\delta T_i} = C_l i^{-1.2}$$

for some constant C_l, and therefore, if $m(i)$ is the number of moves between T_i and T_{i+1}, then

$$\begin{aligned}
&\mathbf{P}\left(m(i) > \frac{1}{2}e^{vT_i}\right) \\
&\leq \mathbf{P}\left(\tilde{N}(T_i) > e^{w+(0.6\delta)T_i}\right) + \mathbf{P}\left(m(i) > \frac{1}{2}e^{vT_i}\middle| \tilde{N}(T_i) \leq e^{w+(0.6\delta)T_i}\right) \\
&\leq C_l i^{-1.2} + C_s e^{-(w+0.6\delta)T_i} \qquad (26)
\end{aligned}$$

where the last inequality uses Claim 1 and the fact that

$$\frac{1}{2}e^{vT_i} > 2e^{(w+0.6\delta)T_i}(\exp(2(T_{i+1}-T_i))-1)$$

Using Borel-Cantelli, we conclude that almost surely,

$$\sum_{k=1}^{A}\left|\tilde{N}_k(T)-\tilde{N}_k(T_i)\right| < \frac{1}{2}e^{vT_i} \qquad (27)$$

for all k and all i large enough and all T between T_i and T_{i+1}. (25) follows from (27).

Proof (Proof of Lemma 2). Using the same martingale argument as above, we can conclude that $\mathrm{var}(\tilde{N}_A(t)|\tilde{N}_A(0)=1) < C_1e^{2wt}$, while $\mathbb{E}(\tilde{N}_A(t)|\tilde{N}_A(0)=1) > C_2e^{wt}$. Therefore there exists $\rho > 0$ such that

$$\mathbf{P}\left(\tilde{N}_A(t) > \rho e^{wt}|\tilde{N}_A(0)=1\right) > \rho. \qquad (28)$$

Fix some large T, and let $t_i = iT$. Then using (28) and independence,

$$\mathbf{P}\left(\tilde{N}_A(t_i) > \frac{\rho^2}{2}e^{wT}\middle| \tilde{N}_A(t_{i-1})\right) > 1-e^{-\frac{1}{16}\tilde{N}_A(t_{i-1})} \qquad (29)$$

where (29) was obtained using Chernoff's bound. From (29), we get that almost surely, for all i large enough,

$$\tilde{N}_A(t_i) > \exp\left(i\left[wT+\log\left(\frac{\rho^2}{2}\right)\right]\right).$$

$\tilde{N}_A(t)$ is monotone increasing, and therefore

$$\tilde{N}_A(t) > C\exp\left(t\left[w-\frac{1}{T}\log\left(\frac{\rho^2}{2}\right)\right]\right). \qquad (30)$$

for all t large enough. Using Lemma 3 we conclude that

$$\tilde{N}_k(t) > C\exp\left(t\left[w-\frac{1}{T}\log\left(\frac{\rho^2}{2}\right)\right]\right) > e^{ut}$$

for all k and large enough t.

Proposition 4 *For every k and j, almost surely*

$$\lim_{t\to\infty}\frac{\tilde{N}_k(t)}{\tilde{N}_j(t)} = \frac{p_k}{p_j} \qquad (31)$$

Proof. This follows immediately from Lemma 2 and Lemma 3.

B.2 Back to discrete time

Proposition 5 *For the discrete time process, and $A > \max\{A_1, A_2\}$ there exists a vector $\hat{q}$ such that, for $k \leq A$, we have*

$$\lim_{\tau\to\infty} \frac{\tilde{N}_k(\tau)}{\tau+1} = q_k. \tag{32}$$

Proof. The i-th newcomer is of degree zero with probability

$$\frac{\sum_{k=1}^{A_1-1} \tilde{N}_k(\tau)}{\sum_{k=1}^{A} \tilde{N}_k(\tau)}$$

However, by (31), this expression tends to a limit, and therefore, using the law of large numbers,

$$\lim_{\tau\to\infty} \frac{N_0(\tau)}{\tau+1} = q_0 = \frac{\sum_{k=1}^{A_1-1} p_k}{\sum_{k=1}^{A} p_k.} \tag{33}$$

Using (31) once more, the proposition now follows for $k \geq 1$ with $q_k = (1-q_0)p_k$.

Note that the above proposition implies that q_k and hence p_k is independent of A if $A > k$, since the left hand side of (32) does not depend on A if $A > k$. So, in particular, p_1 does not depend on A.

C Monotonicity properties of w

In this section we will prove that the exponent $1+w$ of the power law in Proposition 3 is monotonically decreasing in A_1 and monotonically increasing in A_2. For this purpose, it will be useful to define a family of matrices, parametrized by two vectors $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, which generalizes the matrix M appearing in (15), whose top eigenvalue is w.

Given vectors $\mathbf{y} = (y_1, y_2, \ldots, y_n), \mathbf{z} = (z_1, z_2, \ldots, z_n) \in \mathbb{R}^n$, let $M(\mathbf{y}, \mathbf{z})$ denote the n-by-n matrix whose (ij)-th entry is:

$$M_{i,j}(\mathbf{y}, \mathbf{z}) = \begin{cases} -y_j & \text{if} \quad i = j \\ y_j & \text{if} \quad i = j+1 \\ z_j & \text{if} \quad i = 1 \text{ and } j > 1 \\ 0 & \text{otherwise.} \end{cases}$$

Thus, for instance, the matrix M defined in (15) is $M(\mathbf{y}, \mathbf{z})$, where

$$\begin{aligned} \mathbf{y} &= (1, 2, \ldots, A_2 - 1, A_2, A_2, A_2, \ldots, A_2) \\ \mathbf{z} &= (0, 0, \ldots, 0, \min(A_1, A_2), \min(A_1+1, A_2), \ldots, A_2, A_2) \end{aligned}$$

For the remainder of this section, we will assume:

- $y_i > 0$ for $1 \leq i \leq n$, (34)
- $z_i \geq 0$ for $1 \leq i \leq n$, (35)
- $z_n > 0$. (36)

All of these criteria will be satisfied by the matrices $M(\mathbf{y}, \mathbf{z})$ which arise in proving the desired monotonicity claim. It follows from (34),(35), and (36) that if we add a suitably large scalar multiple of the identity matrix to $M(\mathbf{y}, \mathbf{z})$, we obtain an irreducible matrix $M(\mathbf{y}, \mathbf{z}) + BI$ with non-negative entries. The Perron-Frobenius Theorem guarantees that $M(\mathbf{y}, \mathbf{z}) + BI$ has a positive real eigenvalue R of multiplicity 1, such that all other complex eigenvalues have modulus $\leq R$; consequently $M(\mathbf{y}, \mathbf{z})$ has a real eigenvalue $w = R - B$, of multiplicity 1, such that the real part of every other eigenvalue is strictly less than w.

We will study how w varies under perturbations of the parameters $\mathbf{y}, \mathbf{z}$. Let $P(\lambda, \mathbf{y}, \mathbf{z})$ be the characteristic polynomial of $M(\mathbf{y}, \mathbf{z})$, i.e.

$$P(\lambda, \mathbf{y}, \mathbf{z}) = \det(\lambda I - M(\mathbf{y}, \mathbf{z})).$$

This is a polynomial of degree n in λ (with coefficients depending smoothly on $\mathbf{y}, \mathbf{z}$), whose largest real root $w(\mathbf{y}, \mathbf{z})$ exists and has multiplicity 1, provided $(\mathbf{y}, \mathbf{z})$ belongs to the region $V \subset \mathbb{R}^n \times \mathbb{R}^n$ determined by (34),(35), and (36). It follows from the Implicit Function Theorem that $w(\mathbf{y}, \mathbf{z})$ is a smooth function of $(\mathbf{y}, \mathbf{z})$ in V, satisfying:

$$\left(\frac{\partial P}{\partial y_i} + \frac{\partial w}{\partial y_i} \cdot \frac{\partial P}{\partial w} \right) \Bigg|_{(w, \mathbf{y}, \mathbf{z})} = 0 \tag{37}$$

If $\mathbf{x}$ is any vector in $\mathbb{R}^n \times \mathbb{R}^n$, and $\partial_{\mathbf{x}}$ is the corresponding directional derivative operator, we have from (37):

$$\partial_{\mathbf{x}} w(\mathbf{y}, \mathbf{z}) = - \frac{\partial_{\mathbf{x}} P(w, \mathbf{y}, \mathbf{z})}{(\partial P / \partial w)|_{(w, \mathbf{y}, \mathbf{z})}}. \tag{38}$$

We know that $(\partial P / \partial w)|_{(w, \mathbf{y}, \mathbf{z})} > 0$ because P is a polynomial with positive leading coefficient, w is its largest real root, and w has multiplicity 1. Thus we've established:

Claim 4 *For any vector $\mathbf{x} \in \mathbb{R}^n \times \mathbb{R}^n$, and any $(\mathbf{y}, \mathbf{z}) \in V$, put $w = w(\mathbf{y}, \mathbf{z})$. Then the directional derivatives $\partial_{\mathbf{x}} w(\mathbf{y}, \mathbf{z})$ and $\partial_{\mathbf{x}} P(w, \mathbf{y}, \mathbf{z})$ have opposite signs.*

This allows monotonicity properties of w to be deduced from calculations involving directional derivatives of P. Given the definition of $M(\mathbf{y}, \mathbf{z})$, it is straightforward to compute that

$$P(\lambda, \mathbf{y}, \mathbf{z}) \;=\; \det(\lambda I - M(\mathbf{y}, \mathbf{z})) \;=\; \prod_{i=1}^{n} (\lambda + y_i) \;-\; \sum_{j=2}^{n} P_j(\lambda, \mathbf{y}, \mathbf{z}), \tag{39}$$

where

$$P_j(\lambda, \mathbf{y}, \mathbf{z}) = \left(\prod_{i=1}^{j-1} y_i \right) z_j \left(\prod_{i=j+1}^{n} (\lambda + y_j) \right). \tag{40}$$

The following three lemmas encapsulate the requisite directional derivative estimates.

Lemma 4. $(\partial P / \partial z_k)|_{(w, \mathbf{y}, \mathbf{z})} < 0$ *for* $(\mathbf{y}, \mathbf{z}) \in V$.

Proof.

$$\partial P / \partial z_k = -\partial P_k / \partial z_k = - \left(\prod_{i=1}^{k-1} y_i \right) \left(\prod_{i=k+1}^{n} (w + y_i) \right) < 0.$$

Corollary 6 *w is monotonically decreasing in A_1.*

Proof. Increasing A_1 from k to $k+1$ has no effect on $\mathbf{y}$, and its only effect on $\mathbf{z}$ is to decrease z_k from $\min(k, A_2)$ to 0. As we move in the $-z_k$ direction, the directional derivative of P is positive, so the directional derivative of w is negative by Claim 4. Thus w decreases as we increase A_1 from k to $k+1$.

Lemma 5. *$(\partial P/\partial y_k)|_{(w,\mathbf{y},\mathbf{z})} < 0$ if $(\mathbf{y}, \mathbf{z}) \in V$ and $z_k = 0$.*

Proof.

$$
\begin{aligned}
\frac{\partial P}{\partial y_k} &= \frac{\partial}{\partial y_k}\left[\prod_{i=1}^{n}(w+y_i)\right] - \sum_{j=2}^{n}\frac{\partial P_j}{\partial y_k} \\
&= \frac{1}{w+y_k}\prod_{i=1}^{n}(w+y_i) - \frac{1}{y_k}\sum_{j=2}^{k-1}P_j - \frac{1}{w+y_k}\sum_{j=k+1}^{n}P_j \\
&< \frac{1}{w+y_k}\prod_{i=1}^{n}(w+y_i) - \frac{1}{w+y_k}\sum_{j=2}^{k-1}P_j - \frac{1}{w+y_k}\sum_{j=k+1}^{n}P_j \\
&= \frac{P(w,\mathbf{y},\mathbf{z})}{w+y_k} \\
&= 0
\end{aligned}
$$

Lemma 6. *$(\partial P/\partial y_k + \partial P/\partial z_k)|_{(w,\mathbf{y},\mathbf{z})} < 0$ if $(\mathbf{y}, \mathbf{z}) \in V$ and $y_k = z_k$.*

Proof.

$$
\begin{aligned}
\frac{\partial P}{\partial y_k} + \frac{\partial P}{\partial z_k} &= \frac{\partial}{\partial y_k}\left[\prod_{i=1}^{n}(w+y_i)\right] - \sum_{j=2}^{n}\frac{\partial P_j}{\partial y_k} - \frac{\partial P_k}{\partial z_k} \\
&= \frac{1}{w+y_k}\prod_{i=1}^{n}(w+y_i) - \frac{1}{y_k}\sum_{j=2}^{k-1}P_j - \frac{1}{w+y_k}\sum_{j=k+1}^{n}P_j - \frac{1}{z_k}P_k \\
&< \frac{1}{w+y_k}\prod_{i=1}^{n}(w+y_i) - \frac{1}{w+y_k}\sum_{j=2}^{k-1}P_j - \frac{1}{w+y_k}\sum_{j=k+1}^{n}P_j - \frac{1}{w+y_k}P_k \\
&= \frac{P(w,\mathbf{y},\mathbf{z})}{w+y_k} \\
&= 0
\end{aligned}
$$

Corollary 7 *w is monotonically increasing in A_2.*

Proof. If we change A_2 from k to $k+1$, this changes

$$\mathbf{y} = (1, 2, \ldots, k-1, k, k, \ldots, k)$$

into

$$\mathbf{y}' = (1, 2, \ldots, k-1, k, k+1, \ldots, k+1)$$

and it changes

$$\mathbf{z} = (0, 0, \ldots, 0, \min(A_1, k), \min(A_1+1, k), \ldots, k, k)$$

into

$$\mathbf{z}' = (0, 0, \ldots, 0, \min(A_1, k+1), \min(A_1+1, k+1), \ldots, k+1, k+1).$$

Letting $\mathbf{e}_j^{(y)}$ denote a unit vector in the $+y_j$ direction, and $\mathbf{e}_j^{(z)}$ a unit vector in the $+z_j$ direction, the direction of change is expressed by the vector

$$\mathbf{x} = (\mathbf{y}', \mathbf{z}') - (\mathbf{y}, \mathbf{z}) = \sum_{k+1 \le j < A_2} \mathbf{e}_j^{(y)} + \sum_{\max(k+1, A_2) \le j} (\mathbf{e}_j^{(y)} + \mathbf{e}_j^{(z)})$$

and $\partial_{\mathbf{x}} P$ is negative, by the preceding two lemmas. By Claim 4, this means w increases monotonically as we move along this path.

D Proof of Claim 1

To see the finiteness of X_t, we need to show that $\sum_{n=1}^{\infty} o_n = \infty$ a.s. But this follows easily from the fact that $\sum_{n=1}^{\infty} n^{-1} = \infty$. To see (21), we use the following argument,

The standard birth process is equivalent to the following process: Start with one cell at time 0. At each time, every cell multiplies with rate 2. X_t is the number of cells at time t.

Lemma 7 (Joel Spencer). *For every $t > 0$ and every positive integer k, $\mathbb{E}(X_t^k) < \infty$.*

Proof. Let $T = (V, E)$ be an infinite rooted binary tree, and let $\{u_e\}_{e \in E}$ be i.i.d. exponential variables with expected value 0.5. Then, X_t is dominated by the size of

$$Y_t = \left\{ v \in V : \sum_{e \in \gamma(v)} u_e < t \right\}$$

where $\gamma(v)$ is the path from the root to v. For v of depth n, the probability that v is in Y_t is

$$\mathbf{P}(\text{Poisson}(2t) > n) = o\left(((n/2)!)^{-1} \right)$$

Therefore,

$$\mathbf{P}(|Y_t| > h) \le \sum_{h = \left[\frac{\log h}{log 2}\right]}^{\infty} 2^n \mathbf{P}(\text{Poisson}(2t) > n) \le C(t) h^{-\log h}$$

and therefore all moments of X_t are finite.

Claim 1 will follow from Chebyshef if we show that

$$\mathbb{E}(X_{t_2} | X_{t_1}) = X_{t_1} O\left(e^{2(t_2 - t_1)} \right) \tag{41}$$

and

$$\text{var}(X_{t_2} | X_{t_1}) = X_{t_1} O\left(e^{4(t_2 - t_1)} \right) \tag{42}$$

for $t_2 > t_1$. To show (41) and (42), it is enough to show that $\mathbb{E}(X_t) = O(e^{2t})$ and $\text{var}(X_t) = O(e^{4t})$. $\mathbb{E}(X_t) = O(e^{2t})$ follows because $f(t) = \mathbb{E}(X_t)$ satisfies the differential equation

$$\frac{df}{dt} = 2f(t).$$

$\text{var}(X_t) = O(e^{4t})$ follows using the exact same martingale argument as in Claim 3.