Heuristically Optimized Trade-offs: A New Paradigm for Power Laws in the Internet

Alex Fabrikant[1], Elias Koutsoupias[2*], and Christos H. Papadimitriou[1**]

[1] Computer Science Division
University of California at Berkeley
Soda Hall 689
Berkeley CA 94720, U.S.A.
alexf@csua.berkeley.edu, christos@cs.berkeley.edu
[2] University of Athens and UCLA
Computer Science Department
3731 Boelter Hall
Los Angeles, CA 90095, U.S.A.
elias@cs.ucla.edu

Abstract. We propose a plausible explanation of the power law distributions of degrees observed in the graphs arising in the Internet topology [Faloutsos, Faloutsos, and Faloutsos, SIGCOMM 1999] based on a toy model of Internet growth in which two objectives are optimized simultaneously: "last mile" connection costs, and transmission delays measured in hops. We also point out a similar phenomenon, anticipated in [Carlson and Doyle, Physics Review E 1999], in the distribution of file sizes. Our results seem to suggest that power laws tend to arise as a result of complex, multi-objective optimization.

1 Introduction

It was observed in [5] that the degrees of the Internet graph (both the graph of routers and that of autonomous systems) obey a sharp *power law*. This means that the distribution of the degrees is such that the probability that a degree is larger than D is about $cD^{-\beta}$ for some constant c and $\beta > 0$ (they observe βs between 2.15 and 2.48 for various graphs and years). They go on to observe similar distributions in Internet-related quantities such as the number of hops per message, and, even more mysteriously, the largest eigenvalues of the Internet graph. This observation has led to a revision of the graph generation models used in the networking community [14], among other important implications. To date, there has been no theoretical model of Internet growth that predicts this phenomenon. Notice that such distributions are incompatible with random graphs in the $G_{n,p}$ model and the law of large numbers, which yield exponential distributions.

* Research supported in part by NSF and the IST Program.
** Research supported in part by an NSF ITR grant, and an IBM faculty development award.

Power laws have been observed over the past century in income distribution [13], city populations [15, 6], word frequencies [10], and literally hundreds of other domains including, most notably, the degrees of the world-wide web graph [9]; they have been termed "the signature of human activity" (even though they do occasionally arise in nature)[1]. There have been several attempts to explain power laws by so-called generative models (see [12] for a technical survey). The vast majority of such models fall into one large category (with important differences and considerable technical difficulties, of course) that can be termed *scale-free growth* or *preferential attachment* (or, more playfully, *"the rich get richer"*). That is, if the growth of individuals in a population follows a stochastic process that is independent of the individual's size (so that larger individuals attract more growth), then a power law will result (see, from among dozens of examples, [6] for an elegant argument in the domain of city populations, and [8] for a twist involving copying in the world-wide web, crucial for explaining some additional peculiarities of the web graph, such as the abundance of small bipartite graphs).

Highly optimized tolerance (HOT, [2]) is perhaps the other major class of models predicting power laws. In HOT models, power laws are thought to be the result of optimal yet reliable design in the presence of a certain hazard. In a typical example, the power law observed in the distribution of the size of forest fires is attributed to the firebreaks, cleverly distributed and optimized over time so as to minimize the risk of uncontrolled spread of fire. The authors of [2] refer briefly to the Internet topology and usage, and opine that the power law phenomena there are also due to the survivability built in the Internet and its protocols (whereas it is well known that this aspect of the Internet has not had a significant influence on its development beyond the very beginning [3]).

In this paper we propose a simple and primitive model of Internet growth, and prove that, under very general assumptions and parameter values, it results in power-law-distributed degrees. By "power law" we mean here that the probability that a degree is larger than d is *at least* $d^{-\beta}$ for some $\beta > 0$. In other words, we do not pursue here sharp convergence results *à la* [8, 4, 1], content to bound the distribution away from exponential ones. Extensive experiments suggest that much stronger results actually hold.

In our model a tree is built as nodes arrive uniformly at random in the unit square (the shape is, as usual, inconsequential). When the i-th node arrives, it attaches itself on one of the previous nodes. But which one? One intuitive objective to minimize is the Euclidean distance between the two nodes. But a newly arrived node plausibly also wants to connect to a node in the tree that is "centrally located", that is, its hop distance (graph theoretic distance with edge lengths ignored) to the other nodes is as small as possible. The former objective captures the "last mile" costs, the latter the operation costs due to communication delays. Node i attaches itself to the node j that minimizes the weighted sum of the two objectives:

$$min_{j<i} \alpha \cdot d_{ij} + h_j,$$

[1] They are certainly the product of one particular kind of human activity: looking for power laws...

where d_{ij} is the Euclidean distance, and h_j is some measure of the "centrality" of node j, such as (a) the average number of hops from other nodes; (b) the maximum number of hops from another node; (c) the number of hops from a fixed center of the tree; our experiments show that all three measures result in similar power laws, even though we only prove it for (c). α is a parameter, best thought as a function of the final number n of points, gauging the relative importance of the two objectives.

We are not claiming, of course, that this process is an accurate model of the way the Internet grows. But we believe it is interesting that a simple and primitive model of this form leads to power law phenomena. Our model attempts to capture in a simple way the *trade-offs* that are inherent in networking, but also in all complex human activity (arguably, such trade-offs are key manifestations of the aforementioned complexity).

The behavior of the model depends crucially on the value of α, and our main result (Theorem 1) fathoms this dependency: If α is less than a particular constant depending on the shape of the region, then Euclidean distances are not important, and the resulting network is easily seen to be a *star* —the ultimate in degree concentration, and, depending on how you look at it, the exact opposite, or absurd extreme, of a power law. If α grows at least as fast as $\sqrt{n}$, where n is the final number of points, then Euclidean distance becomes *too* important, and the resulting graph is a dynamic version of the Euclidean minimum spanning tree, in which high degrees do occur, but with exponentially vanishing probability (our proof of this case is a geometric argument). Again, no power law. If, however, α is *anywhere in between* — is larger than a certain constant, but grows slower than $\sqrt{n}$ if at all — then, almost certainly, the degrees obey a power law. This part is proved by a combinatorial-geometric argument, in which we show that for any value for the desired degree, there are likely to be enough nodes with large enough "regions of influence," disjoint from one another, such that any future node falling into this region is certain to have an edge to the given node. Our technique proves a lower bound of $\beta = \frac{1}{6}$ for $\alpha = o(n^{1/3})$ (and smaller bounds for $\alpha = o(\sqrt{n})$), while our experiments (see Section 2.2) suggest that the true value is around 0.6-0.9.

In Section 3 we prove a result in a different but not unrelated domain: we present a simple (naïve is more accurate) model of file creation, inspired by [2], and prove that it predicts a power law in the distribution of file sizes under very broad assumptions. The model is this: We have a set of n data items, all of the same size, that we must partition into files. The i-th item has *popularity* p_i — say, the expected number of times it will be retrieved for Internet transmission each day. We want to partition the items into files so that the following two objectives are minimized (a) total transmission costs (the sum over all partitions of the product of the partition size times the total partition popularity), and (b) the total number of files. It is easy to see that the optimum partition will include items in sorted order of popularity, and can be found by dynamic programming. We show that, when the popularities are i.i.d. from any one of a large class of distributions (encompassing the uniform, exponential, Gaussian, power law, etc. distributions), then the optimum file sizes are, almost surely, power-law distributed. The technical requirement on the distribution from which the p_i's are drawn is essentially that the cumulative distribution Φ do not start exponentially slow at zero (see Theorem 2). The authors of [2] propose a similar model, and make an observation in the same direction: They consider a few examples of distributions of the p_i's (*not* distributions from which they are drawn, as in our model, but distributions of the drawn samples) and for these they point out that the file sizes obey a power law.

Our results seem to suggest that power laws are perhaps the manifestation of *trade-offs*, complicated optimization problems with multiple and conflicting objectives — arguably one of the hallmarks of advanced technology, society, and life. Our framework generalizes the HOT class of models proposed in [2], in the sense that HOT models are the trade-offs in which reliable design is one of the objectives being optimized.

As it turns out, within our proposed conceptual framework also lies a classical and beautiful model by Mandelbrot [10]. Suppose that you want to design the optimum language, that is, the optimum set of frequencies $f_1 \leq f_2 \leq \ldots \leq f_n$ assigned to n words. The length of the i-th word is presumably $\log i$. You want to maximize the information transmitted, which is the entropy of the f_i's ($-\sum f_i \log f_i$), divided by the expected transmission cost, $\sum f_i \log i$. The frequencies that achieve the optimum: a power law! Mandelbrot's multi-objective optimization differs from our two examples in that of his two objectives one is minimized and the other maximized; hence he considers their ratio, instead of their weighted sum (the two are obviously related by Lagrange multipliers).

2 A Model of Internet Growth

2.1 The Main Result

Consider a sequence of points $p_0, p_1, \ldots, p_n$ in the unit square, distributed uniformly at random. We shall define a sequence of undirected trees $T_0, T_1, \ldots$ on these points, with T_0 the tree consisting of p_0. Define h_i to be the number of hops from p_i to p_0 in T_i, and d_{ij} the Euclidean distance between points i[2] and j. Let α be a fixed number (we allow it though to be a function of n). Then T_i is defined as T_{i-1} with the point i and the edge $[i,j]$ added, where $j < i$ minimizes $f_i(j) = \alpha d_{ij} + h_j$. Let $T = T_n$. We will denote by $N_k(i)$ the neighborhood $\{j | [i,j] \in T_k\}$ of i in T_k; similarly, $N(i)$ will denote the neighborhood of i in T.

Theorem 1. *If T is generated as above, then:*

(1) If $\alpha < 1/\sqrt{2}$, then T is a star with p_0 as its center.
(2) If $\alpha = \Omega(\sqrt{n})$, then the degree distribution of T is exponential, that is, the expected number of nodes that have degree at least D is at most $n^2 \exp(-cD)$ for some constant c: $E\left[\,|\{i : \text{ degree of } i \geq D\}|\,\right] < n^2 \exp(-cD)$.

[2] p_i and "point i" are used interchangeably throughout

(3) If $\alpha \geq 4$ and $\alpha = o(\sqrt{n})$, then the degree distribution of T is a power law; specifically, the expected number of nodes with degree at least D is greater than $c \cdot (D/n)^{-\beta}$ for some constants c and β (that may depend on α): $E[\,|\{i:$ degree of $i \geq D\}|\,] > c(D/n)^{-\beta}$. Specifically, for $\alpha = o(\sqrt[3]{n^{1-\epsilon}})$ the constants are: $\beta \geq 1/6$ and $c = O(\alpha^{-1/2})$.

Proof. We prove each case separately. The proof of the third case is the more involved.

(1) The first case follows immediately from the objective function; since $d_{ij} < \sqrt{2}$ for all i, j, and $h_j \geq 1$ for all $j \neq 0$, $f_i(0) < 1 \leq f_i(j)$ for $j \neq 0$, so every node p_i will link to p_0, creating a star.

(2) To obtain the exponential bound for (2), we consider the degree of any point p_i as consisting of 2 components — one due to geometrically "short" links, $S(i) = |\{j \in N(i) \,|\, d_{ij} \leq \frac{4}{\alpha}\}|$, and one due to "long" links, $L(i) = |\{j \in N(i) \,|\, d_{ij} > \frac{4}{\alpha}\}|$. By the union bound, $\Pr[\text{degree}_i \geq D] \leq \Pr[S(i) \geq D/2] + \Pr[L(i) \geq D/2]$.

For a fixed i and $\alpha \geq c_0\sqrt{n})$, any points contributing to $S(i)$ must fall into a circle of area $16\pi\alpha^{-2} \leq \pi {c_0}^{-2} n^{-1}$. Thus, $S(i)$ is (bounded by) a sum of Bernoulli trials, with $\mathrm{E}[S(i)] = 16\pi\alpha^{-2}n < c$ for a constant c depending only on c_0. By the Chernoff-Hoeffding bound, for $D > 3c$, $\Pr[S(i) > D/2] \leq \exp(-\frac{(D-2c)^2}{D+4c}) \leq \exp(-D/21)$.

For the other component, define $L_x(i) = |\{j \in N(i) \,|\, d_{ij} \in [x, \frac{3}{2}x]\}|$ (the number of points attached to i in a distance between x and $\frac{3}{2}x$ from point i). We will first show that $L_x(i) < 14$ for any $x \geq \frac{4}{\alpha}$. Indeed a geometric argument shows if points p_j and $p_{j'}$, $j < j'$, are both between x and $\frac{3}{2}x$ away from p_i, then $p_{j'}$ would prefer p_j over p_i whenever $|\angle p_j p_i p_{j'}| < c = \cos^{-1}(43/48)$ (see Figure 1); the bound on the angle would force $\alpha d_{ij'} > \alpha d_{jj'} + 1$ while $|h_j - h_i| \leq 1$. Since $c > \frac{2\pi}{14}$, $L_x(i) < 14$. We now bound $L(i)$ as follows: $L(i) = \sum_{k=1}^{-\log_{\frac{3}{2}} \delta_i} L_{(\frac{3}{2})^{-k}}(i) \leq -14 \log_{\frac{3}{2}} \delta_i$ where δ_i is defined as $\max\{\frac{4}{\alpha}, \min_j d_{ij}\}$. Since points are distributed uniformly at random, $\Pr[\delta_i \leq y] \leq 1 - (1 - \pi y^2)^{(n-1)} \leq \pi(n-1)y^2$. Therefore, $\Pr[L(i) \geq D/2] \leq \Pr[-14 \log_{\frac{3}{2}} \delta_i \geq D/2] \leq \pi(n-1)(\frac{3}{2})^{-D/14}$, completing the proof of (2).

It is worth noting that the only property used for this bound is $|h_j - h_i| \leq 1$ for $j \in N(i)$; this holds for a broad class of hop functions, including all 3 listed in the introduction.

(3) To derive the power law in (3), we concentrate on points close to p_0. While experimental evidence suggests the presence of high-degree points contributing to the power law throughout the area, the proof is rendered more tractable by considering only points $j \in N(0)$, with $d_{0j} \leq 2/\alpha$. Without loss of generality, we assume that p_0 is located at least $2/\alpha$ from the area boundary; the argument carries over to cases where it is near border with only a slight change in C.

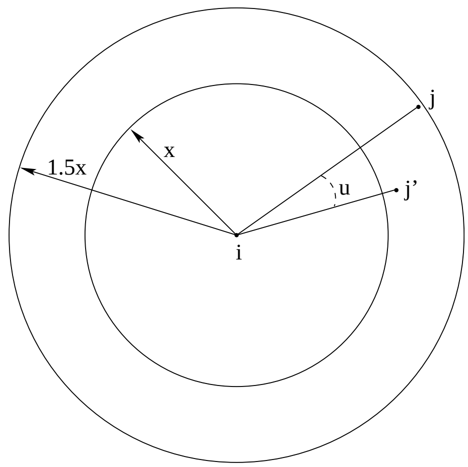

Fig. 1. If $u < cos^{-1}(43/48)$ then point j' prefers point j over point i because $\alpha d_{ij'} > \alpha d_{jj'} + 1$ for $x \geq 4/\alpha$.

First, we prove 2 lemmas for a point i which arrives so that $i \in N(0)$ and $\frac{3}{2\alpha} > d_{i0} > \frac{1}{\alpha}$. Let $r(i) = d_{i0} - \frac{1}{\alpha}$.

Lemma 1. *Every point arriving after i inside the circle of radius $\frac{1}{4}r(i)$ around i will link to i.*

Proof. Since i was linked to 0 on arrival, there was no $j \in N(0)$ prior to its arrival such that $\alpha d_{ij} + 1 < \alpha d_{i0}$, i.e. within distance $r(i)$ from it. Also, if a point j arrives after i so that $d_{ij} < \frac{1}{2}r(i)$, it will not link to 0, since by the triangle inequality $d_{j0} > 1/\alpha + \frac{1}{2}r(i)$. Now, if a point j' arrives after i so that $d_{ij'} < \frac{1}{4}r(i)$, it can't be linked to 0; for all $j \in N(0)\backslash\{i\}$, $d_{jj'} > \frac{1}{4}r(i)$, so j' would rather link to i; and for all other j, $h_j \geq 2$, so $f_{j'}(j) \geq 2$, while $f_{j'}(i) \leq \alpha\frac{1}{4}r(i) + 1 \leq \frac{9}{8}$. Thus, any such point j' will definitely link to i.

Lemma 2. *No point j will link to i unless $|\angle p_j p_0 p_i| \leq \sqrt{2.5\alpha r(i)}$ and $d_{j0} \geq \frac{1}{2}r(i) + 1/\alpha$.*

Proof. Note that if $f_j(0) < f_j(i)$, j will not link to i since i is not the optimal choice. That constraint is equivalent to $d_{j0} < d_{ij} + 1/\alpha$, which defines a region outside the cusp around p_i of a hyperbola with foci at p_0 and p_i and major axis length $1/\alpha$. The asymptotes for this hyperbola each make an angle of $\arctan\sqrt{\alpha^2 r(i)^2 + 2\alpha r(i)} \leq \arctan\sqrt{2.5\alpha r(i)} \leq \sqrt{2.5\alpha r(i)}$ with the segment $\overline{p_0 p_i}$, intersecting it at the midpoint m of $\overline{p_0 p_i}$. Since $|\angle p_i m x| \geq |\angle p_i p_0 x|$ for any point x, this guarantees that any point p_j inside the cusp around p_i satisfies $|\angle p_j p_0 p_i| \leq \sqrt{2.5\alpha r(i)}$. The triangle inequality also guarantees that any such p_j will also satisfy $d_{j0} \geq \frac{1}{2}r(i) + 1/\alpha$. Thus, any point not satisfying both of these will not link to i.

Lemma 1 provides a way for an appropriately positioned $i \in N(0)$ to establish a circular "region of influence" around itself so that any points landing there afterward will contribute to the degree of that point. Since point placement is independent and uniform, its exponentially unlikely that the circle will be populated with much fewer points than the expected number, thus giving a stochastic lower bound on $\deg i$. We use Lemma 2 to lower-bound the number of $i \in N(0)$ with sufficiently large $r(i)$. For a sketch of the geometrical features of the proof, refer to Figure 2.

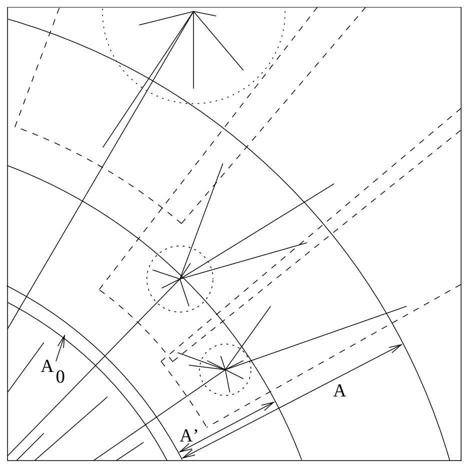

Fig. 2. A hypothetical small window about $1/\alpha$ from the root, of size on the order of $\rho^{2/3}$; figure not entirely to scale. Dotted contours indicate "regions of influence" provided by Lemma 1; dashed contours indicate the "sector" areas defined by Lemma 2. Note that the latter may overlap, while the former cannot.

Here, we only treat the case for $\alpha = o(\sqrt[3]{n^{1-\epsilon}})$, which yields $\beta = 1/6$; the case for $\alpha = o(\sqrt{n})$ can be analyzed similarly, but with β approaching 0 as α approaches $\Theta(\sqrt{n}$.

For any $\epsilon > 0$ and sufficiently high n, suppose $\alpha = o(\sqrt[3]{n^{1-2\epsilon}})$, and let $D \leq \frac{n^{1-\epsilon}}{256\alpha^3}$. Set, with foresight, $\rho = 4\sqrt{D/n}$ and $m = \lceil \frac{1}{2\rho} \rceil$, and consider T_m. Specifically, consider the sets of points in T_m falling within annuli A_0, A, and A' around p_0, with radius ranges $[1/\alpha, 1/\alpha + \rho]$, $(1/\alpha + \rho, 1/\alpha + \rho^{2/3}]$, and $(1/\alpha + \rho, 1/\alpha + 0.5\rho^{2/3}]$, respectively. By our choice of ρ, any $j \in N_m(0)$ in A' will have a region of influence of area at least $\pi D/n$, thus expected to contain $\pi D(n-m)/n > \pi D/2$ points that link to j in T. By our choice of m, the number of points expected to arrive in A_0 is $1/\alpha + \rho/2 < \frac{1}{2}$.

Consider any point i arriving to A'. It cannot link to any j with $h_j \geq 2$, since $f_i(0) < 2 \leq f_i(j)$. By Lemma 2, it cannot link to any $j \in N(0)$ outside the outer radius of A. By triangle inequality, it cannot link to any $j \in N(0)$ such that $d_{j0} < 1/\alpha$ (since $d_{i0} - d_{ij} < 1$). Thus, it can only link to 0 or any $j \in N(0)$ that lies in $A_0 \cup A$.

But, to link to a $j \in N(0)$ which is in $A_0 \cup A$, i must, by Lemma 2 obey $|\angle p_i p_0 p_j| < \sqrt{2.5\alpha r(j)} < \sqrt{2.5\alpha}\rho^{1/3}$. Thus, each time a new point arrives in $A_0 \cup A$ and links to $N(0)$, it "claims" a sector of A' of angle no larger than $\sqrt{10\alpha}\rho^{1/3}$; i.e. no point arriving outside that sector can link to j (note that we disregard the constraint on d_{j0}; it is not needed here). The number of points in T_m expected to be in A' is $m(\rho^{2/3}/\alpha - 2\rho/\alpha + 0.25\rho^{4/3} - \rho^2) > 1/(4\rho^{1/3})$. This can be cast as an occupancy problem by partitioning annulus $A_0 \cup A$ into $N = 1/(8\sqrt{\alpha}\rho^{1/3})$ congruent sectors[3] of angle $16\pi\sqrt{\alpha}\rho^{1/3} > \sqrt{10\alpha}\rho^{1/3}$. Each partition is considered occupied if a point has landed in it or either of its adjacent partitions and linked to p_0, and, by the above argument, a point landing in the intersection of an unoccupied partition and A' will link to p_0, so the number of partitions occupied after m points arrive is at most $3|N(0) \cap (A \cup A_0)|$. By the Chernoff bound, with probability at least $p_1 = 1 - \exp(-N/8)$, at least $1/(8\rho^{1/3})$ points in T_m land in A'. Note that if a point lands in the intersection of a partition and A', that partition is definitely occupied, so:

$$
\begin{aligned}
p_2 &= \Pr[N/2 \text{ partitions occupied by points in } T_m] \\
&= 1 - \sum_{k=N/2}^{N} \binom{n}{k} \Pr[k \text{ partitions unoccupied}] \\
&\geq p_1 \left(1 - \sum_{k=N/2}^{N} \binom{n}{k} \left(1 - \frac{k}{N}\right)^{\sqrt{\alpha}N}\right) \\
&\geq p_1 \left(1 - \sum_{k=N/2}^{N} \binom{n}{k} \frac{1}{2^{2N}}\right) \\
&\geq p_1(1 - 2^{-N})
\end{aligned}
$$

Hence, with probability $p_2 \geq 1 - 2\exp(-N/8)$, there are at least $N/6$ points in $N(0) \cap (A \cup A_0)$. By the Chernoff bound again, we find that the probability that more than $N/12$ of these are in A_0 is $\exp(-\frac{2}{m}(N/12 - 1/2)^2) < \exp(-N/8)$, so, with probability at least $1 - 3\exp(-N/8)$, there are $N/12$ points in $N(0) \cap A$, each with expected degree at least $\pi D/2$ in T.

Lastly, by another application of the Chernoff bound, the probability that the degree of any such point is less than D is $\exp(-\frac{1}{2}N(1 - 2/\pi)^2) < \exp(-N/20)$. Thus, with exponentially high probability $1 - (3 + N/24)\exp(-N/20)$, for $C = \frac{1}{2^{16/3}3} n^{-5/6}\alpha^{-1/2}$, and any D in the above specified range, the probability that a randomly chosen point in T has degree at least D is at least $N/24n = CD^{-1/6}$. That is, the degree distribution is lower-bounded by a power law at least up to a constant power of n.

[3] Note that $N = \Theta(D^{-1/6}n^{1/6}\alpha^{-1/2}) = \Omega(n^{\epsilon/6})$. Also, since the analysis here assumes $N \geq 1$, it only applies for $D \leq n/(2048\alpha^3)$.

Allowing p_0 to be placed closer than $2/\alpha$ from the border causes only a fraction of the annuli to be within the region for some values, but since at least a quarter of each annulus will always be within the region, N changes only by a small constant factor. To extend the above argument to $\alpha = o(\sqrt{n})$, the outer radii of A and A' would have to be reduced, making the ring thinner and allowing us to partition it into asymptotically more sectors. However, much fewer of them will be occupied, leading to a decrease in β.

2.2 Experiments

An implementation of both this model and several natural variations on it has shown that the cumulative density function (c.d.f.) of the degree distribution produced indeed appears to be a power law, as verified by a good linear fit of the logarithm of the c.d.f. with respect to the logarithm of the degree for all but the highest observed degrees. Using $n = 100,000$ and $\alpha \leq 100$, the β values observed from the slope of the linear fit ranged approximately between 0.6 and 0.9. When we used higher values of α, the range of D where the c.d.f. exhibited linear behavior shrunk too much to allow good estimation of β.

The c.d.f. generated for 2 tests are shown in the Appendix. Specifically, Figures 3 and 4 show the c.d.f. for $n = 100,000$ and $\alpha = 4$ and $\alpha = 20$ (the straight line shown is not actually a linear fit of the entire data set, but visually approximates a large part of the data). We also show the associated trees. (Actually due to enormous (postscript) file sizes, these are only the partially trees $T_{10,000}$.) The full trees (which do not differ substantially) and a more complete set of experimental results can be found in
`http://research.csua.berkeley.edu/~alexf/hot/`.
Tests with varying values of n and fixed α produced consistent values for β throughout, as expected.

Furthermore, replacing the "rooted" h_j centrality measure with other alternatives mentioned in the introduction also produced power-law behavior in the c.d.f., with very similar values for β. Also, the "maximum hops" centrality measure, which is identical to the "rooted" measure provided that the first node remains the global minimum of h_j, while being more decentralized (and thus better mirroring a real network), was indeed observed to retain p_0 as the "most central" node in more than 75% of the cases.

Clearly the experiments suggest sharper power laws than our theoretical results and they also seem to occur for much wider range of the parameters. Our proofs could in fact be improved somewhat to give tighter results but we opted for simplicity.

It is straightforward to extend the proof of the main theorem (Theorem 1) to higher dimensions and other metrics —with different constants of course. The experiments indeed verify that the power law behavior also holds when the 2-dimensional square was replaced by a circle, as well as by hypercubes or hyperspheres in $\mathbb{R}^d$ for various $d \geq 2$ with the L_i (for various $i \geq 1$) or L_∞ metrics. Interestingly, one-dimensional domains do not seem to give rise to power laws.

3 A Model of File Creation

Suppose that we are given n positive real numbers $p_1, \ldots, p_n$, intuitively capturing the "popularity" of n data items, the expected number of times each will be retrieved. We wish to find a partition Π of the items into sets so as to minimize $\min_{\Pi} \left[\sum_{S \in \Pi} |S| \cdot \sum_{i \in S} p_i \right] + \alpha \cdot |\Pi|$. That is, the trade-off now is between transmission costs and file creation overhead, with α capturing the relative importance of the latter. It is very easy to see the following:

Proposition 1. *The optimum solution partitions the p_i's sorted in decreasing order, and can be found in $O(n^2)$ time by dynamic programming.*

It is shown by a rough calculation in [2] that, if the p_i's decrease exponentially, or polynomially, or according to a "Gaussian" $\exp(-ci^2)$ law, then the optimum file sizes obey a power law.

In this section we prove a different result, starting from the assumption that the p_i's are drawn i.i.d. from a distribution f.

We start with two inequalities capturing certain properties of the optimum solution. Suppose that the partitions $S_1, \ldots, S_k$ have sizes $s_i = |S_i|$, and that the average item in S_i is a_i. It is not hard to see that the following is true:

Lemma 3. *$s_i + s_{i+1} \geq \sqrt{\alpha/a_i}$, and $s_i \leq \sqrt{2\alpha/a_i}$.*

The proof uses that, by optimality, it is not advantageous to merge two sets, or to split one in the middle.

Consider now the cumulative distribution Φ of f, and its inverse Ψ. That is, $\Psi(x)$ is the least y for which $\Pr[z \leq y] \geq x$. It is useful to think of $\Psi(y/n)$ as the expected number of elements with popularity smaller than y. Let $g = \Psi(\log n/n)$.

Lemma 4. *Almost certainly in the optimum solution there are at least $y/2\sqrt{2\alpha/g}$ sets of size at least $\sqrt{\alpha/2\Psi(2y/n))}$.*

Sketch of proof: With high probability, the popularity of the smallest element is no bigger than g, and, for large enough $y \leq n$, there are at least y elements with popularities smaller than $\Psi(2y/n)$. By the previous lemma, the sets that contain these elements have sizes that satisfy $s_i + s_{i+1} \geq \sqrt{\alpha/\Psi(2y/n)}$ and $s_i \leq \sqrt{2\alpha/g}$. Thus, these elements are divided into at least $y/\sqrt{2\alpha/g}$ sets (by the second inequality), half of them of size at least $\frac{1}{2}\sqrt{\alpha/\Psi(2y/n)}$ (by the first inequality). □

From this lemma we conclude the following for the distribution of file sizes:

$$\Pr[\text{size of a file} \geq \frac{1}{2}\sqrt{\alpha/\Psi(2y/n)}] \geq y/2n\sqrt{2\alpha/g}.$$

Now set $x = \frac{1}{2}\sqrt{\alpha/\Psi(2y/n)}$ or $\Psi(2y/n) = \alpha/4x^2$ or $2y/n = \Phi(\alpha/4x^2)$ or $y = n\Phi(\alpha/4x^2)/2$. Therefore we have

Theorem 2. *In the distribution of file sizes induced by the optimum solution,*

$$\Pr[\text{size of a file} \geq x] \geq \Phi(\alpha/4x^2)\sqrt{g/32\alpha}.$$

It follows that, with high probability, the file sizes are power-law distributed whenever $\lim_{z\to 0}\Phi(z)/z^c > 0$ for some $c > 0$; this condition does not hold only for distributions f that are extremely slow to start at zero. For example, any continuous distribution f that has $f(0) > 0$ (such as the exponential, normal, uniform, etc.) gives a power law.

4 Discussion and Open Problems

We have observed that power laws can result from trade-offs between various aspects of performance or fitness that must be optimized simultaneously (deployment vs. operational costs in one example, file creation overhead and transmission costs in another). In contrast, when the trade-off is weak (in that one of the criteria is overwhelmingly important), exponentially concentrated distributions result. This seems to be a new genre of a rigorous argument predicting power laws. Needless to say, our network model does exhibit a behavior of the form "the rich get richer," just like many previous models of such phenomena:Nodes that arrived early are more likely to both have high degree and small hop cost, and thus to attract new nodes. The point is, in our model this is not a primitive (and hard to defend) assumption, but a rather sophisticated consequence of assumptions that are quite simple, local, and "behavioral."

By extending our tree generation algorithm to non-tree graphs (attach each arriving node to the few most advantageous nodes, where the number of new edges is appropriately distributed to produce graphs with the correct average degree > 1) we obtain an interesting network generation model, about which we can prove next to nothing. However, it was suggested to us by Ramesh Govindan [7] that the graphs produced by this more elaborate model are passably Internet-like, in that they seem to satisfy several other observed properties of the Internet graph, besides the degree distribution.

It would be very interesting to extend our results to other definitions of the "hop" cost, and to strengthen them by proving stronger power laws (bigger exponents than we can prove by our simple techniques, hopefully closer to the ones observed in our experiments). It would be wonderful to identify more situations in which multi-criterion optimization leads to power laws, or, even more ambitiously, of generic situations in which multi-criterion optimization can be proved sufficient to create solutions with power-law-distributed features.

Finally, [5] observed another intriguing power law, besides the one for the degrees analyzed here: The largest *eigenvalues* of the Internet graph appear to also fit a power law. In joint work with Milena Mihail [11], we have discovered a way of explaining the power law distribution of the eigenvalues. Briefly, it is a corollary of the degree distribution: If a graph consists of a few nodes of very degrees $d_1 \geq d_2 \geq \cdots \geq d_k$ plus a few other edges, then with high probability the k largest eigenvalues will be about $\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_k}$. Indeed, in the power law observed in [5] the exponent of the eigenvalues are roughly half those of the degrees. The small difference can be explained by the fact that [5] examines only the 20 highest eigenvalues; these correspond to extreme degrees that do not fit exactly the eigenvalue power law (as is quite evident in the figures in [5]).

References

1. Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Structures and Algorithms*, 18(3):279–290, 2001.
2. J. M. Carlson and J. Doyle. Highly optimized tolerance: a mechanism for power laws in designed systems. *Physics Review E*, 60(2):1412–1427, 1999.
3. Manuel Castells. *The Internet Galaxy: Reflections on the Internet, Business, and Society*. Oxford, 2001. http://www.oup.co.uk/isbn/0-19-924153-8.
4. Colin Cooper and Alan M. Frieze. A general model of undirected web graphs. In *ESA*, pages 500–511, 2001.
5. C. Faloutsos, M. Faloutsos, and P. Faloutsos. On power-law relationships of the internet topology. In *Proc. SIGCOMM*, 1999.
6. X. Gabaix. Zipf's law for cities: an explanation. *Quarterly Journal of Economics*, 114:739–767, 1999.
7. R. Govindan. Private communication.
8. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 57–65, 2000.
9. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Extracting large scale knowledge bases from the Web. In *Proceedings of the 25th VLDB Conference*, 1999.
10. B. Mandlebrot. An informational theory of the statistical structure of languages. In W. Jackson, editor, *Communication theory*, pages 486–502. Betterworth, 1953.
11. M. Mihail and C. Papadimitriou. Power law distribution of the Internet eigenvalues: An explanation. Unpublished.
12. M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. Manuscript.
13. V. Pareto. *Cours d'Economie Politique*. Dronz, Geneva Switzerland, 1896.
14. H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Network topologies, power laws, and hierarchy. Technical Report 01-746, Computer Science Department, University of Southern Calofornia, 2001. To appear in Computer Communication Review.
15. G. Zipf. *Human behavior and the principle of least effort*. Addison-Wesley, Cambridge MA, 1949.

Appendix

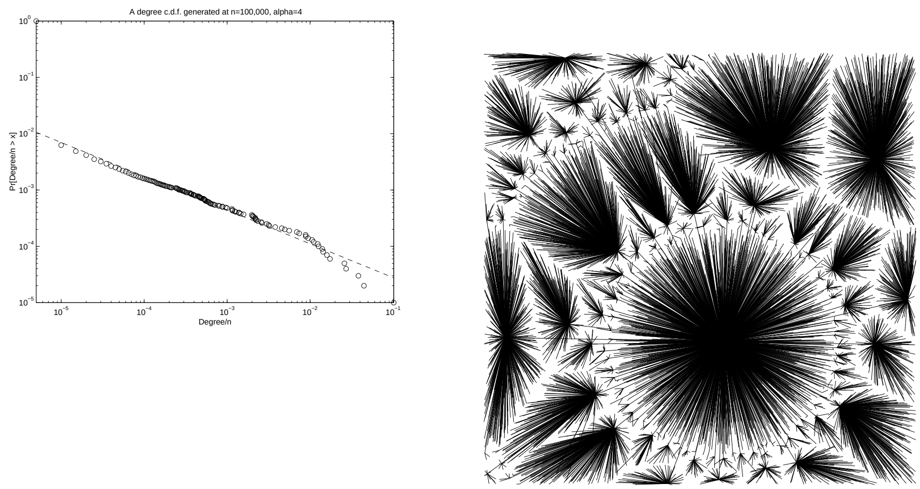

Fig. 3. c.d.f. and the associated tree generated for $\alpha = 4$ and $n = 100,000$ (only the first $10,000$ nodes are shown)

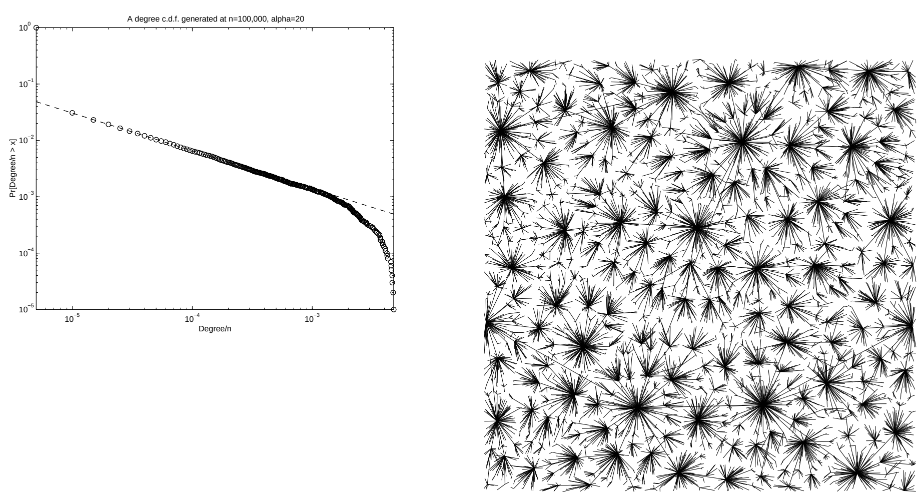

Fig. 4. c.d.f. and the associated tree generated for $\alpha = 20$ and $n = 100,000$ (only the first $10,000$ nodes are shown)