Degree distribution of the FKP network model

Noam Berger[1], Béla Bollobás[2,3], Christian Borgs[4], Jennifer Chayes[4], and Oliver Riordan[3]

[1] Department of Statistics, University of California, Berkeley, CA 94720. Research undertaken during an internship at Microsoft Research.
[2] Department of Mathematical Sciences, University of Memphis, Memphis TN 38152. Research supported by NSF grant DSM 9971788 and DARPA grant F33615-01-C-1900.
[3] Trinity College, Cambridge CB2 1TQ, UK. Research undertaken while visiting Microsoft Research.
[4] Microsoft Research, One Microsoft Way, Redmond, WA 98122.
noam@stat.berkeley.edu, {B.Bollobas,O.M.Riordan}@dpmms.cam.ac.uk, {borgs,jchayes}@microsoft.com

Abstract. Recently, Fabrikant, Koutsoupias and Papadimitriou [7] introduced a natural and beautifully simple model of network growth involving a trade-off between geometric and network objectives, with relative strength characterized by a single parameter which scales as a power of the number of nodes. In addition to giving experimental results, they proved a power-law lower bound on *part* of the degree sequence, for a wide range of scalings of the parameter. Here we prove that, despite the FKP results, the overall degree distribution is very far from satisfying a power law.
First, we establish that for almost all scalings of the parameter, either all but a vanishingly small fraction of the nodes have degree 1, or there is exponential decay of node degrees. In the former case, a power law can hold for only a vanishingly small fraction of the nodes. Furthermore, we show that in this case there is a large number of nodes with almost maximum degree. So a power law fails to hold even approximately at either end of the degree sequence range. Thus the power laws found in [7] are very different from those given by other internet models or found experimentally [8].

1 Introduction

In the last few years there has been an explosion of interest in "scale-free" random networks, based on measurements indicating that many large real-world networks have certain scale-free properties, for example power-law distributions of degrees and other parameters. The original observations of Faloutsos, Faloutsos and Faloutsos [8] and later many others have led to a host of proposals of random graph models to explain these power laws, and to better understand the mechanisms at work in the growth of real-world networks such as the internet or web graphs; see [2, 3, 9] for a few examples. For extensive surveys of the huge amount of work in this area, see Albert and Barabási [1] and Dorogovtsev and Mendes [6]; for a survey of the rather smaller quantity of mathematical work see [4].

Most of the models introduced use a small number of basic mechanisms, mainly preferential attachment or copying, to produce power laws, and do not involve any reference to underlying geometry. Thus, while they may be appropriate for the web graph, for example, they do not seem to be suitable for the internet graph itself.

In [7], Fabrikant, Koutsoupias and Papadimitriou (FKP) proposed a new paradigm for power law behaviour, which they called "heuristically optimized trade-offs": power laws may result from "complicated optimization problems with multiple and conflicting objectives." Their paradigm generalizes previous work of Carlson and Doyle [5] on "highly optimized tolerance," in which reliable design is one of the objectives.

In order to illustrate this paradigm, FKP introduced a simple, natural network model with such a mechanism. As in many models, a network is grown one node at a time, and each node chooses a previous node to which it connects. However, in contrast to other network 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 which is "central" in the network as a graph. Centrality may be measured by using, for example, the graph distance to the initial node.

Several variants of the basic model are considered by Fabrikant, Koutsoupias and Papadimitriou in [7]. The precise version we shall consider here is the principal version studied in [7]: fix a region $\mathcal{D}$ of area one in the plane, for example a disc or a unit square. The model is then determined by the number of nodes, $n+1$, and a parameter, α. We start with a point x_0 of $\mathcal{D}$ chosen uniformly at random, and set $W(x_0) = 0$. For $i = 1, 2, \ldots, n$ we choose a new point x_i of $\mathcal{D}$ uniformly at random, and connect x_i to an earlier point x_j chosen to minimize

$$W(x_j) + \alpha d(x_i, x_j)$$

over $0 \leq j < i$. Here $d(.,.)$ is the usual Euclidean distance. Having chosen x_j, we set $W(x_i) = W(x_j) + 1$. At the end we have a random tree $T = T(n, \alpha)$ on $n+1$ nodes $x_0, \ldots, x_n$, where each node has a weight $W(x_i)$ which is just its graph distance in the tree from x_0.

As in [7], we consider $n \to \infty$ with α some function of n, typically a power.

One might think from the title or a first reading of [7] that the form of the degree sequence of this model has been essentially established. In fact, as we shall describe in the next section, this is not the case. Indeed, two of our results, while of course consistent with the actual results of [7], go against the impression given there that the entire degree sequence follows a power law.

2 Results

As in [7] we consider α in two ranges. Roughly speaking, *large* α will mean $\alpha > n^{1/2}$, and *small* α will mean $\alpha < n^{1/2}$. In fact, to keep things simple we will allow ourselves a logarithmic gap.

Most of the time we will work in terms of the tail of the distribution. Let $\alpha = \alpha(n)$ be given. For each $k = 1, 2, \ldots$, let $q_k(\alpha, n)$ be the expected number of nodes of $T(n, \alpha)$ with degree at least k, and let $\rho_k(\alpha) = \lim_{n\to\infty} q_k(\alpha, n)/n$ be the limiting proportion of nodes with degree at least k.

2.1 Small α

The impression given on first reading [7] is that for small α the whole degree distribution follows a power law. However, the experimental results of [7] strongly suggest that there is a new kind of power law, holding over a large range of degrees, from 2 up to a little below the maximum degree, but involving only a very small proportion of the vertices.

On a second look the situation is more confusing; quoting the relevant part of the theorem (changing D to k for consistency with our notation):

> 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 k is greater than $c \cdot (k/n)^{-\beta}$ for some constants c and β (that may depend on α): $E[|\{i : \text{degree of } i \geq k\}|] > c(k/n)^{-\beta}$. Specifically, for $\alpha = o(\sqrt[3]{n^{1-\epsilon}})$ the constants are: $\beta \geq 1/6$ and $c = O(\alpha^{-1/2})$.

The usual form of a power law would be that a proportion $k^{-\beta}$ of vertices have degree at least k, which is not what is claimed above. There are other problems: the constant c depends on α which depends on n, so c is not a constant. Allowing c to be variable, the claim may then become meaningless if c is very small.

Turning to the proof in [7], a nice geometric argument is given to show that, for $\alpha = o(n^{(1-\epsilon)/3})$ and $k \leq n^{1-\epsilon}/(C\alpha^3)$, which is far below the maximum degree, the expected number $q_k(\alpha, n)$ of vertices with degree at least k is at least $cn^{-5/6}\alpha^{-1/2}k^{-1/6}$, where c and C are absolute constants. This supports the experimental results, showing that this interesting new model does indeed give power laws over a wide range; however, it tells us nothing about the vast majority of the vertices, namely all but $O(n^{1/6})$.

Now, in many examples of real-world networks, and in the preferential attachment and copying models of [2, 9] and others, the power-law degree distribution involves almost all vertices, and, less clearly, holds very nearly up to the maximum degree. (In the latter case, the power law is often called a Zipf law, though in fact Zipf's law is a power law with a particular power.) Thus it is interesting to see whether this is the case for the FKP model.

Theorem 1. *Let $\alpha = o(n^{1/2}/(\log n)^2)$. Then,* **whp** *the tree $T(n, \alpha)$ has at least $n - O(\alpha^{1/2}n^{3/4}\log n) = n - o(n)$ leaves.*

In other words, almost all vertices of $T(n, \alpha)$ have degree 1; in particular, when $\alpha = n^a$ for $a < 1/2$, the number of vertices with degree more than 1 is at most n^b for some $b < 1$. This contrasts strongly with the usual sense of power-law scaling, namely that the density of vertices of degree k converges to a function $f(k)$ which in turn decays like a power of k. This notion is implicit in [8] and [1], for example.

Our final result concerns the high degree vertices, showing that a 'Zipf-like' law does not hold. As usual we write $O^*(\cdot)$ for $O((\log n)^C \cdot)$, suppressing constant powers of $\log n$, and similarly for $\Theta^*(\cdot)$. We write **whp** to mean *with high probability*, i.e., with probability $1 - o(1)$ as $n \to \infty$.

Theorem 2. *Suppose that $(\log n)^4 \leq \alpha \leq n^{1/2}/(\log n)^4$. Then there are constants $c, C > 0$ such that* **whp** *the maximum degree of $T(n, \alpha)$ is at most Cn/α^2, while $T(n, \alpha)$ has $\Theta^*(\alpha^2)$ nodes of degree at least cn/α^2.*

Taking $\alpha = n^a$ for $0 < a < 1/2$, for example, this says that there are many (a power of n) vertices with degree close to (within a constant factor of) the maximum degree. This contrasts sharply with a so-called Zipf distribution, where there would be a constant number of such vertices. In fact, our method will even show that there are many vertices with degree $(1 - o(1))$ times the maximum.

2.2 Large α

We now turn to the simpler case of large α. This case is interesting for three reasons: one is simply completeness. The second is that the case $\alpha = \infty$, while involving no trade-offs, is a very nice geometric model in its own right. Finally, the large α results will turn out to be useful in studying the small α case.

Theorem 3. *Suppose that $\alpha = \alpha(n)$ satisfies $\alpha/(\sqrt{n}\log n) \to \infty$. Then there are positive constants A, A', C, C' such that*

$$A'e^{-C'k} \leq \rho_k(\alpha) \leq Ae^{-Ck}$$

holds for every $k \geq 1$.

In other words, for large α the tail of the degree distribution decays exponentially, as for classical random graphs with constant average degree.

Our theorem strengthens the upper bound in [7], which says that $q_k(\alpha, n) \leq O(n^2)e^{-Ck}$, or, loosely speaking, that $\rho_k(\alpha) \leq O(n)e^{-Ck}$. Note that the upper bound of [7] gives information only for k larger than a constant times $\log n$, i.e., a vanishing fraction of the nodes. Furthermore, we complement our stronger upper bound with a matching lower bound.

We remark again that our results contain logarithmic factors that are presumably unnecessary; these help keep the proofs relatively simple.

3 The pure geometric model

In this section we consider the case $\alpha = \infty$. In this case, each node x_i simply connects to the closest node among $x_0, \ldots, x_{i-1}$. Although this model is not our main focus, it is of interest in its own right, and it is somewhat surprising that it does not seem to have been extensively studied, unlike related objects such as the minimal spanning tree, for example (see [11, 12]). We study this case for two reasons. First, for large α, $T(n, \alpha)$ approximates $T(n, \infty)$. Second, certain results about $T(n, \infty)$ will be useful to study $T(n, \alpha)$ even for very small α. We start with a simple but surprising exact result.

Lemma 1. *In the random tree $T(n, \infty)$, for $1 \leq t \leq n$ the probability that x_t is at graph distance r from x_0, i.e., has weight r, is exactly*

$$\sum_{1 \leq i_1 < i_2 < \ldots < i_{r-1} < t} \frac{1}{i_1 i_2 \ldots i_{r-1} t}$$

Proof. We write $i \to j$ if $j < i$ and x_i is adjacent (joined directly) to x_j. The key observation is as follows: suppose we fix the points $x_s, x_{s+1}, \ldots, x_n$, and also the *set* of points $S_{s-1} = \{x_0, x_1, \ldots x_{s-1}\}$, leaving undetermined the order of the points in S_{s-1}. Then x_s is joined to the closest point in S_{s-1}, which is a certain point x. When we choose the ordering of the points in S_{s-1}, the point x is equally likely to be x_0, x_1, or any other x_j, $j < s$. Taking $s = t$, it follows that the probability that $t \to j$ is exactly $1/t$. Using the same observation for $s = j$ we see that, given $t \to j$, the probability that $j \to k$ is $1/j$. Continuing, the probability that $t \to i_{r-1} \to i_{r-2} \to \cdots \to i_1 \to 0$ is $1/(t i_{r-1} i_{r-2} \cdots i_1)$. As these events are disjoint for different sequences, the lemma follows.

Another way of stating the lemma is that for any fixed t, the distribution of the graph distance from t to 0 is the same as in a uniform random recursive tree. These are trees grown one node at a time, in which each new node is joined to an earlier node chosen uniformly at random. Such objects have been studied for some time; see, for example, the survey [10]. The radius (here, maximum node weight) of such a tree was shown by Pittel [13] to be $(c+o(1))\log n$ for a certain constant $c = 1.79..$ given by a root of an equation. This result does not apply to $T(n, \alpha)$ because the dependence between nodes is different. We shall just give an upper bound.

Lemma 2. *Let $\alpha = \alpha(n)$ be arbitrary. Then as $n \to \infty$,* **whp** *every point in $T(n, \alpha)$ has weight at most $3 \log n$.*

Proof. For $\alpha = \infty$ this follows from Lemma 1 by straightforward calculation: the expected number of points with weight r is

$$\sum_{1 \le i_1 < i_2 < \ldots < i_{r-1} < t \le n} \frac{1}{i_1 i_2 \ldots i_{r-1} t} \le \frac{(1 + \log n)^r}{r!} \le (e(1 + \log n)/r)^r .$$

Set $r = \lfloor 3 \log n \rfloor$. Then the expectation above tends to zero, so **whp** there are no points with weight r, and the radius, or maximum weight, is at most $r - 1$.

We can compare finite α with $\alpha = \infty$. Consider the sequence of points as fixed, let $W(x_i)$ be the weights for some finite $\alpha = \alpha(n)$, and let $W_\infty(x_i)$ be the weights obtained with $\alpha = \infty$. For any α, the weight of a point x_i is always at most one more than the weight of the nearest earlier point x_j: if we connect to a more distant point x_k it must have smaller weight than x_j. Since we have equality for $\alpha = \infty$, it follows that for any α we have $W(x_i) \le W_\infty(x_i)$. As shown at the start of the proof, **whp** we have $W_\infty(x_i) \le 3 \log n$ for every i, so we are done.

The lemma has a simple heuristic explanation: for $\alpha = \infty$ the closest earlier x_j to x_i will typically have index j around $i/2$, so it will take order $\log n$ steps to reach the origin. For finite α, any bias is towards earlier points. One might expect monotonicity of the weights as α decreases from one finite value to another, but this does not hold in general.

3.1 Degrees for $\alpha = \infty$

Here we are interested in the quantities $\rho_k(\infty)$ defined in section 2; our aim is to prove the $\alpha = \infty$ case of Theorem 3.

This result easy to see intuitively. As noted above, for $i < t \leq n$ the probability that $t \to i$ is exactly $1/t$. Thus the expected degree of node i in $T(n, \infty)$ is exactly

$$\frac{1}{i+1} + \frac{1}{i+2} + \cdots + \frac{1}{n} = \log(n/i) + O(i^{-1}).$$

If every degree were close to its expectation, this would give the result. In fact, it turns out that the probability of the degree of node i exceeding its expectation by some amount x decreases exponentially with x. To see this heuristically we use the notion of Voronoi cells: given a region $\mathcal{D}$ and a set of points X in $\mathcal{D}$, the region $\mathcal{D}$ is tiled by Voronoi cells V_x, one for each $x \in X$, defined as the set of points of $\mathcal{D}$ closer to x than to any other $y \in X$.

Here we consider $V_{i,t}$, the Voronoi cell of x_i with respect to $x_0, x_1, \ldots, x_t$. Note that $t \to i$ if and only if x_t is in $V_{i,t-1}$. Keeping i fixed, as t increases $V_{i,t}$ shrinks whenever x_t lands close enough to x_i. In particular, $V_{i,t}$ gets smaller whenever x_t lands in $V_{i,t-1}$ itself; the key point is that in this case the area of $V_{i,t}$ is on average less than that of $V_{i,t-1}$ by a factor f strictly less than 1. On average, $V_{i,i}$ has area $1/(i+1)$, and $V_{i,n}$ area $1/(n+1)$. Hence it is very unlikely that i has degree much bigger than $\log(n/i)$; otherwise the area of $V_{i,t}$ would decrease by too much as t increases from i to n.

Proof (of Theorem 3 for $\alpha = \infty$). We make the argument outlined above rigorous. The key observation is as follows: let V be a convex region and C a point of V. Let X be a point of V chosen uniformly at random, and let V' be the set of points of V closer to C than to X. Then the expected area of V' is at most $15/16$ times the area of V. To see this, taking C as the origin divide V into four parts Q_1, Q_2, Q_3, Q_4, the intersections of V with the four quadrants of $\mathbf{R}^2$. Suppose X falls in a certain Q_i. If Y is any other point of Q_i then $(X+Y)/2$ is closer to X then to C. This is easy to see geometrically: the vector $(X+Y)/2 - X = (Y-X)/2$ is shorter than $(Y+X)/2$, as the angle between X and Y is less than 90 degrees. Hence $V \setminus V'$ contains a copy of Q_i shrunk by a factor two in each direction, so in this case $\mathrm{area}(V \setminus V') \geq \mathrm{area}(Q_i)/4$. Averaging, noting that the probability that X lies in Q_i is proportional to $\mathrm{area}(Q_i)$,

$$\mathbf{E}(\mathrm{area}(V \setminus V')) \geq \sum_{i=1}^{4} \frac{\mathrm{area}(Q_i)^2}{4\,\mathrm{area}(V)} \geq \frac{\mathrm{area}(V)}{16},$$

where the last step follows by convexity. Thus $\mathbf{E}(\mathrm{area}(V')) \leq \frac{15}{16}\,\mathrm{area}(V)$. Hence, fixing $x_0, \ldots, x_{t-1}$, conditional on $t \to i$, i.e., on $x_t \in V_{i,t-1}$, the expected area of $V_{i,t}$ is at most $\frac{15}{16}$ times the area of $V_{i,t-1}$.

Fix $0 \leq i \leq n$. Continuing the construction of $T(n, \infty)$ indefinitely, let $t_1 < t_2 < t_3 < \cdots$ be the points that send edges to i. Let $W_0 = V_{i,i}$ and $W_j = V_{i,t_j}$ be the Voronoi cells of i looked at at time i, and at each time when a new node joins to i. Note that $\mathbf{E}(\mathrm{area}(W_0)) = 1/(i+1)$ as this is the cell corresponding to one of $i+1$ points chosen independently. It may be that the Voronoi cell containing i shrinks at intermediate times as well, but certainly given W_j, we have $\mathbf{E}(\mathrm{area}(W_{j+1})) \leq \frac{15}{16}\,\mathrm{area}(W_j)$. Hence

$$\mathbf{E}(\mathrm{area}(W_k)) \leq \frac{1}{i+1}(15/16)^k. \tag{1}$$

We now consider time n: fix x_i and consider the n remaining points of $x_0, \ldots, x_n$ as random. Ignoring effects from the boundary of the region, if no other point lies within distance d of x_i, then the Voronoi cell $V_{i,n}$ contains a circle of radius $d/2$. In other words, for area$(V_{i,n})$ to be smaller than $\pi(d/2)^2$, one of the n points must lie in a disc of radius d, with area πd^2, an event with probability at most $n\pi d^2$. It turns out that boundary effects go the right way, so

$$\Pr(\text{area}(V_{i,n}) \leq x) \leq 4nx. \tag{2}$$

Finally, if i has degree at least $k+1$ in $T(n,\infty)$ then at least k of the first n points join to i, so $t_k \leq n$, and area$(V_{i,n}) \leq$ area(W_k). For any x, the probability of this is at most

$$\Pr(\text{area}(W_k) \geq x) + \Pr(\text{area}(V_{i,n}) \leq x),$$

which is at most

$$\frac{1}{x}\frac{1}{i+1}(15/16)^k + 4nx,$$

from (1), Markov's inequality and (2). The optimum choice

$$x = (15/16)^{k/2}/\sqrt{4n(i+1)}$$

yields

$$\Pr(\deg(i) \geq k+1) \leq 4\sqrt{\frac{n}{i+1}}(15/16)^{k/2}. \tag{3}$$

Summing over i by comparison with an integral, the expected number of nodes with degree at least $k+1$ is at most $(8+o(1))n(15/16)^{k/2}$, so $\rho_{k+1} \leq 8(15/16)^{k/2}$, proving the upper bound.

The lower bound also follows easily; the bound (3) shows that an individual degree is very unlikely to be much larger than its expectation. It follows that deg(i) has a significant (at least 1%, say) chance of being at least half its expectation, and the lower bound follows.

4 Observation

In the remaining proofs we will use again and again the following simple observation. At time t the points currently placed approximate a Poisson process with density $1/t$, so the closest earlier point x_j to x_t is 'typically' at distance $\Theta(1/\sqrt{t})$. In particular, for a fixed t, if $\omega \to \infty$ then **whp** $\omega^{-1}t^{-1/2} \leq d(x_t, x_j) \leq \omega t^{-1/2}$.

Furthermore, for any positive constant c, **whp** at time t every disk of radius $c \log nt^{-1/2}$ contains a point already placed. (This is easy to check, and also follows from a more general and more precise result of Penrose [11].)

5 Large α

Proof (of Theorem 3). The case $\alpha = \infty$ was proved in section 3; to extend this result to α large requires only a little further work.

Suppose that $\alpha/(\sqrt{n}\log n) \to \infty$. Fix $\delta > 0$, and consider a point x_i with $i \geq \delta n$, and the nearest earlier point x_j. Since all weights are within $3\log n$ of one another, for x_i to join to some other point x_k we must have

$$d(x_i, x_k) \leq d(x_i, x_j) + 3\log n/\alpha = d(x_i, x_j) + o(n^{-1/2}). \tag{4}$$

As noted above, **whp** we have $d(x_i, x_j) \le \omega i^{-1/2}$. Considering x_i and x_j as fixed, the other x_k, $k < i$, are distributed uniformly outside the circle centered at x_i with radius $d(x_i, x_j)$, and for a particular x_k to satisfy (4) it must lie in an annulus around this circle with thickness $o(n^{-1/2})$. This annulus has area $o(d(x_i, x_j)n^{-1/2}) = o((in)^{-1/2})$ (taking $\omega \to \infty$ slowly enough). Since there are $i-1$ points to consider, the probability that x_i does not join to the closest point x_j is at most $o(\sqrt{i/n}) = o(1)$. Thus, **whp**, almost all points join to the nearest earlier point. In particular, the final tree $T(n, \alpha)$ differs in only $o(n)$ edges from $T(n, \infty)$, and hence the numbers ρ_k are the same as for $\alpha = \infty$.

The conclusion that $\rho_k(\alpha) = \rho_k(\infty)$ should hold provided only that $\alpha/\sqrt{n} \to \infty$; this is likely to be harder to show.

6 Critical α

If $\alpha = \Theta(\sqrt{n})$ then we expect the behaviour of the tree to be similar to that for $\alpha = \infty$. In particular, for $\alpha = cn^{1/2}$, $c > 0$, we expect limiting proportions $\rho_k = \rho_k(c)$ with $\rho_k(c) \to \rho_k(\infty)$ as $c \to \infty$ but $\rho_k(c)$ not in general equal to $\rho_k(\infty)$. Also, the radius, or maximum weight, should be $A(c) \log n$. We have not stated a result for this case, which is likely to be harder to analyze precisely.

Note that one might hope for a complete power law in the critical case, but this does not happen, as shown by, for example, the weak exponential upper bound in [7].

7 Small α

This case is the heart of our paper. Here *small* would ideally mean $o(n^{1/2})$; in fact, for simplicity we shall work with extra logarithmic factors. Throughout this section it will be convenient to re-scale by a factor of α: rather than choosing points in the unit square or disc, we choose points in a square $\mathcal{D}$ of side α; correspondingly, we join x_i to the earlier point x_j minimizing $W(x_j) + d(x_i, x_j)$. Note that the final density n/α^2 of points is high (compared to 1). The reason to consider this scaling is that differences in re-scaled distances of order 1 are what is relevant; in particular, as all weights are within $3 \log n$ of each other, no point ever connects to a point more than $3 \log n$ further away than its nearest point.

Considering the process defining $T(n, \alpha)$ as points arrive one by one, there is a transition in the behaviour around time $t = \alpha^2$. This is because in the re-scaled process, the density of points at time t is t/α^2. At times much smaller than α^2, this density is very small, so distances and their differences are typically large, and the process looks very much like the $\alpha = \infty$ case of connecting to the nearest point.

On the other hand, at times much later than α^2, the density of points is already very high. We expect that certain 'attractive' early points will have established 'regions of attraction' of order unit size; almost all later points then just join to the nearest attractive point by a short edge. In particular, almost all later points will themselves never be joined to.

7.1 Small degrees

We now prove Theorem 1 from section 2, a precise version of the final observation from the paragraph above, that almost all points are leaves in $T(n, \alpha)$, i.e., have degree 1. In the proof we shall use the following simple geometric lemma.

Lemma 3. *Let $\mathcal{D}$ be a convex set in the plane, and let $X = \{x_1, \ldots, x_k\}$ be a set of points in $\mathcal{D}$. For $r > 0$ let $X(r)$ be the set of points in $\mathcal{D}$ at distance at most r from some x_i. For $0 < r_1 < r_2$ we have*

$$\mathrm{area}(X(r_2)) \leq \frac{r_2^2}{r_1^2}\,\mathrm{area}(X(r_1)).$$

Proof. A point $x \in \mathcal{D}$ lies in $X(r)$ if and only if $d(x, x_i) \leq r$ for x_i the closest point of X to x. Let us partition $\mathcal{D}$ into the Voronoi cells $V_i = \{x \in \mathcal{D} : d(x, x_i) = \min_j d(x, x_j)\}$. (We may ignore the boundaries.) Then, for any r, we have $\mathrm{area}(X(r)) = \sum_i \mathrm{area}(X(r) \cap V_i)$. But V_i is convex; thus if $X(r_2) \cap V_i$ is a certain region A, then $X(r_1) \cap V_i$ certainly contains the region obtained by shrinking A by a factor r_2/r_1 around the point x_i. Hence, $\mathrm{area}(X(r_1) \cap V_i) \geq r_1^2/r_2^2\,\mathrm{area}(X(r_2) \cap V_i)$, and the lemma follows.

Of course, a corresponding result holds in any dimension, with exactly the same proof. Also, the result holds for an arbitrary set X.

Proof (of Theorem 1). If x_i is joined to the earlier point x_j, we call $x_i x_j$ *the edge from* x_i. We consider edges with lengths in three ranges: writing γ for $\alpha^{1/2} n^{-1/4} = o(1/\log n)$, we call an edge of length l *short* if $l < 1$, *long* if $l > 1 + \gamma$, and *medium* if $1 \leq l \leq 1 + \gamma$.

The key observation is that if the edge $x_i x_j$ from x_i is short, then x_i has degree 1 in the final graph $T(n, \alpha)$. To see this, note that no later point x_k can possibly join to x_i, since $W(x_i) = W(x_j) + 1$, while $d(x_k, x_j) < d(x_k, x_i) + 1$, so x_k would join to x_j in preference to x_i. To complete the proof we shall show that the number of medium and long edges is small.

Suppose that the edge $x_i x_j$ from x_i is medium. Writing w for $W(x_j)$, at time $i - 1$ there is no point with weight w within distance 1 of x_i, but there is such a point within distance $1 + \gamma$. Turning this around, let $X = \{x_j : W(x_j) = w, 1 \leq j \leq i - 1\}$. Then x_i lies in $X(1 + \gamma)$, but not in the interior of $X(1)$. By Lemma 3, $\mathrm{area}(X(1 + \gamma)) \leq (1 + \gamma)^2\,\mathrm{area}(X(1))$. Hence, given $x_0, \ldots x_{i-1}$, the probability that x_i lies in $X(1 + \gamma) \setminus X(1)$ is at most $\frac{(1+\gamma)^2 - 1}{(1+\gamma)^2} \leq 2\gamma$. It follows Lemma 2 that there are at most $\log n$ values of w to consider, so the probability that for a given i the edge $x_i x_j$ is medium is at most $2\gamma \log n = o(1)$. It follows that **whp** there are at most $2\gamma n \log n = 2\alpha^{1/2} n^{3/4} \log n = o(n)$ medium edges in the final tree.

We now consider long edges, i.e., edges of length at least $1+\gamma$. The key observation is that when the edge from x_i is long, this edge provides a useful shortcut in future: new points near x_i have a better connection route than if x_i were deleted. To formalize this, given the final set of points $x_0, \ldots, x_n$ and their weights, for $1 \leq i \leq n$ let us define a function $c_i : \mathcal{D} \to \mathbf{R}$ by $c_i(x) = \min_{j<i}\{W(x_j) + d(x, x_j)\}$. Note that c_i only depends on the locations of $x_0, \ldots, x_{i-1}$, and that $c_i(x)$ is the 'cost' of connecting a potential new point at x to the existing tree on $x_0, \ldots, x_{i-1}$. In particular, x_i joins to

the x_j attaining the minimum defining $c_i(x_i)$, and receives weight $W(x_j)+1$. Suppose that x_ix_j is long, i.e., has length at least $1+\gamma$, and let $w = W(x_j)$. Then we have $c_i(x_i) = w + d(x_i, x_j) \geq w + 1 + \gamma$, but $c_{i+1}(x_i) = w + 1$. Hence

$$c_{i+1}(x_i) \leq c_i(x_i) - \gamma.$$

Our strategy is to consider the quantities $I_i = \int_{\mathcal{D}} c_i(x)$, $1 \leq i \leq n$. We shall show that I_i is positive, and decreases with i. Also, we shall show that **whp** I_{i_0} is not too large for some $i_0 = o(n)$, and that if the edge from i is long, then $I_i - I_{i+1}$ is not too small; together these observations will give a bound on the number of long edges.

It is immediate from the definition that $c_i(x)$ and hence I_i are positive. Also, it is immediate that $c_{i+1}(x) \leq c_i(x)$—the minimum is taken over a larger set. Hence $I_{i+1} \leq I_i$ for each i.

Set $i_0 = \lfloor (\alpha \log n)^2 \rfloor = o(n)$. At time i_0 the overall density of points is at least $(\log n)^2$. Hence, **whp**, for every $x \in \mathcal{D}$ there is a $j < i_0$ with $d(x, x_j) < 1$. Since $W(x_j) \leq 3 \log n$ from section 5, we have $c_{i_0}(x) \leq 1 + 3 \log n$. Thus, **whp**,

$$I_{i_0} \leq (1 + 3 \log n)\text{area}(\mathcal{D}) = O(\alpha^2 \log n).$$

Finally, suppose that the edge from x_i is long. As shown above, we then have $c_{i+1}(x_i) \leq c_i(x_i) - \gamma$. Now each $c_k(x)$ is the minimum of a set of Lipschitz functions with constant 1, and is hence Lipschitz with constant 1. Thus for y at distance $\ell \leq \gamma/2$ from x_i we have $c_{i+1}(y) \leq c_i(y) - \gamma + 2\ell$. Integrating, we see that

$$I_{i+1} \leq I_i - \frac{1}{4} \int_{\ell=0}^{\gamma/2} (\gamma - 2\ell) 2\pi\ell \mathrm{d}\ell = I_i - \frac{\pi}{48}\gamma^3.$$

(The initial factor of $1/4$ allows for the fact that the little disc we are integrating over may not lie entirely within $\mathcal{D}$.)

Since I_i is decreasing and positive, from the two equations above we see that **whp** the number of x_i, $i \geq i_0$, from which we have long edges is at most $O(\alpha^2 \log n / \gamma^3)$. Thus, **whp** we have $i_0 + O(\alpha^2 \log n/\gamma^3) = O(\alpha^{1/2} n^{3/4} \log n)$ long edges.

Combining the cases above completes the proof: we have shown that there are $O(\alpha^{1/2} n^{3/4} \log n) = o(n)$ medium and long edges, and hence $n - o(n)$ short edges. But every short edge gives rise to a leaf in T, so almost all nodes are leaves.

The above result shows that for small α the degree sequence of $T(n, \alpha)$ is not a power law in the usual sense, which is that for fixed k there is a limiting proportion p_k of nodes with degree k, which falls off as some power of k. In particular, here $p_1 = 1$, while $p_k = 0$ for all $k \neq 1$.

7.2 Large degrees

We now turn to the opposite end of the degree sequence, showing that there is a bunching of degrees near the maximum, in the sense that for $\alpha = n^a$, $0 < a < 1/2$, a positive power of n nodes have degree within a constant factor of the maximum. This is easy to see heuristically: up to time α^2 the process looks like the $\alpha = \infty$ case, and all degrees are at most $O(\log n)$. Beyond this time, $\Theta(\alpha^2)$ attractive points will have become established, each of which will attract the $\Theta(n/\alpha^2)$ later points that fall in its zone of attraction, which will have re-scaled area $O(1)$, out of a total re-scaled area of α^2. Since no point can maintain a region of attraction much bigger than this for long, the maximum degree will also be of order $\Theta(n/\alpha^2)$.

As before, for simplicity we have allowed ourselves extra logarithmic factors when making this precise. In Theorem 2, which we now prove, the main case of interest is $\alpha = n^a$ for some constant a between 0 and 1/2.

Proof (of Theorem 2). We start with the maximum degree. Let $t_0 = \alpha^2/(\log n)^2$. Arguing as in section 5 we see that **whp** at time t_0 the tree is essentially $T(t_0, \infty)$, and that all degrees are $O(\log n)$.

Fix a point x_i. To bound the final degree of i we need only consider which x_j, $j > t_0$, join to x_i. Now at time t_0 the typical distance between points is $\log n$, and allowing for deviations no disk of radius $(\log n)^2$ is empty. (This is a rescaling of the final observation from section 4.) It follows that all later edges have length at most $2(\log n)^2$. Hence we need only consider a region R around x_i with radius $O((\log n)^2)$. We divide this into a 'good region', a disk of radius 1.1 around x_i, and a 'bad region', the rest of R. Note that $O(n/\alpha^2)$ points will fall into the good region, so we need only control the bad region. This is easy: the bad region can be covered by $O((\log n)^4)$ disks of radius 0.01. Within any such disk at most one point x_j, $j > i$, can join to i; a second point $x_{j'}$ landing in the same disk would rather join to x_j at distance < 0.01 than to x_i at distance at least 1.1, since the weight of x_j is only one larger than that of x_i. Hence the expected degree of x_i is at most

$$O(\log n) + O(n/\alpha^2) + O((\log n)^4) = O(n/\alpha^2).$$

Since the main term is at least $\Theta((\log n)^2)$ it is easy to check that large deviations are very unlikely, and hence that the maximum degree is $O(n/\alpha^2)$, as claimed.

Establishing the existence of 'attractive' points which remain attractive is not quite so easy, as the situation is not really as simple as the heuristic description suggests. However, with the flexibility allowed by logarithmic factors we can proceed as follows. Let us consider time $t_1 = \alpha^2$. Note that at this time typical distances between nearest points are 1, so **whp** all later edges have length at most $\log n$.

Let us say that a point x_i, $i \le t_1$, of weight w is *good* if no other point x_j, $j \le t_1$, with smaller weight lies within distance $3(\log n)^2$ of x_i. Good points are useful for the following reason: suppose some later point x_k, $k > t_1$, lies within distance 1 of x_i. Then x_k will join to x_i; we have $x_k = x_{a_0} \to x_{a_1} \to x_{a_2} \to \cdots \to x_{a_{l-1}} \to x_{a_l}$ for some sequence $k = a_0 > a_1 \cdots > a_{l-1} > a_l$, with $a_{l-1} > t_1$, $a_l \le t_1$. If we do not have $x_k \to x_i$, then $a_l \ne j$. But then x_k is connected by a sequence of at most $3 \log n$ edges of length at most $\log n$ to a point x_j with $j \le t_1$ of smaller weight than x_i, contradicting that x_i is good. Thus a good point attracts all points after t_1 within distance 1, and will have final degree at least cn/α^2 **whp**. In fact, using only the Chernoff bounds, the deviation probability for one point is $o(n^{-1})$, so **whp** *every* good point has final degree at least cn/α^2.

It remains to show that at time $t_1 = \alpha^2$ there are many good points. We do this using a little trick.

Let $r_w = 3(\log n)^2(1 + 3\log n - w)$, so $r_0 = O^*(1)$, $r_{3\log n} \ge \log n$, and $r_w = r_{w-1} - 3(\log n)^2$. For $0 \le w \le 3\log n$ let T_w be the set of all points in $\mathcal{D}$ within distance r_w of some x_i, $i \le t_1$, with weight at most w. Note that T_0 has area $O((\log n)^6)$, which is much less than α^2. On the other hand, $T_{3\log n}$ is, **whp**, all of $\mathcal{D}$, as at time t_1 every point of $\mathcal{D}$ is within distance $\log n$ of some x_i, which has weight at most $3\log n$ by Lemma 2. Thus there is some w for which $T_w \setminus T_{w-1}$ covers at least a fraction $1/(4\log n)$ of the area of $\mathcal{D}$. Fix such a w, and suppose $y \in T_w \setminus T_{w-1}$. Then there is some x_i with $W(x_i) \le w$ and $d(y, x_i) \le r_w$. On the other hand, there is no x_j with $W(x_j) \le w-1$ within distance $r_{w-1} = r_w + 3(\log n)^2$ of y. It follows that x_i is good, so y is within distance r_w of a good x_i. As each such good x_i can only account for an area $\pi r_w^2 = O^*(1)$ of $T_w \setminus T_{w-1}$, which has area $\Theta^*(\alpha^2)$, it follows that there are at least $\Theta^*(\alpha^2)$ good points, and the proof is complete.

In fact, being a little more careful with the constants, we can show that both the maximum degree and the degrees of almost all good points (those not too near the boundary of $\mathcal{D}$) are $(1+o(1))\pi n/\alpha^2$. Thus there is a strong bunching of degrees near the maximum.

References

1. R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, *Rev. Mod. Phys.* **74** (2002), 47–97.
2. A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *Science* **286** (1999), 509–512.
3. B. Bollobás and O.M. Riordan, The diameter of a scale-free random graph, to appear in *Combinatorica*. (Preprint available from `http://www.dpmms.cam.ac.uk/~omr10/`.)
4. B. Bollobás and O. Riordan, Mathematical results on scale-free random graphs, in Handbook of Graphs and Networks, Stefan Bornholdt and Heinz Georg Schuster (eds.), Wiley-VCH, Berlin (2002).
5. J.M. Carlson and J. Doyle, Highly optimized tolerance: a mechanism for power laws in designed systems. *Phys. Rev. E* **60** (1999), 1412–1427.
6. S.N. Dorogovtsev and J.F.F. Mendes, Evolution of networks, *Adv. Phys.* **51** (2002), 1079.
7. A. Fabrikant, E. Koutsoupias and C.H. Papadimitriou, Heuristically optimized trade-offs: a new paradigm for power laws in the internet ICALP 2002, LNCS 2380, pp. 110–122.
8. M. Faloutsos, P. Faloutsos and C. Faloutsos, On power-law relationships of the internet topology, SIGCOMM 1999, *Comput. Commun. Rev.* **29** (1999), 251.
9. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins and E. Upfal, Stochastic models for the web graph, FOCS 2000.
10. H.M. Mahmoud and R.T. Smythe, A survey of recursive trees, *Th. of Probability and Math. Statistics* **51** (1995), 1–27.
11. M.D. Penrose, A strong law for the largest nearest-neighbour link between random points, *J. London Math. Soc.* (2) **60** (1999), 951–960.
12. M.D. Penrose, A strong law for the longest edge of the minimal spanning tree. *Ann. Probab.* **27** (1999), 246–260.
13. B. Pittel, Note on the heights of random recursive trees and random m-ary search trees, *Random Struct. Alg.* **5** (1994), 337–347.