Heuristically Optimized Trade-offs Model

CS591 Graph Theory Final Project

Lu Min
Department of Computer Science
Boston University
Boston, MA 02215
mlu821@bu.edu

July 1, 2017

Abstract

This project proposes a plausible explanation of the power distributions of degrees observed in the graphs arising in the Internet topology based on a toy model of Internet growth in which two objectives are optimized simultaneously: "last mile" connection costs, and the transmission delays measured in hops. The results seem to suggest that power laws tend to arise as a result of complex, multi-objective optimization.

1 Introduction

It was observed that the degrees of the Internet graph (both the graph of routers and that of autonomous systems) obey a sharp power law. 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, among other important implications.

Most of the methods have approached power-law degree distribution from the preferential-attachment viewpoint: if the âĂIJrich get richerâĂİ, power laws might result. However, another point of view is that power laws can result from resource optimizations.

This project proposes 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.

2 Model

Heuristically Optimized Trade-offs Model is an optimization of the highly optimized tolerance (HOT) model.

2.1 Highly optimized tolerance (HOT) model

Highly optimized tolerance (HOT) 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.

HOT model addresses the following problem. Suppose we have a forest which is prone to forest fire. Each portion of the forest has a different chance of starting the fire. We wish to minimize the damage by assigning resources such as firebreaks at different positions in the forest. However, the total available resources are limited. The problem is to place the firebreaks so that the expected cost of fires is minimized.

2.2 Heuristically Optimized Trade-offs Model

In the model a tree is built as nodes arrive uniformly at random in the unit square. When the i-th node arrives, it attaches itself on one of the previous nodes. One intuitive objective to minimize and decide which node to choose 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 âĂIJcentrally locatedâĂİ, 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 âĂIJlast mileâĂİ costs, the latter the operation costs due to communication delays.

Thus, node i attaches itself to the node j that minimizes the weighted sum of two objectives:

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

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). $/alpha$ is a parameter, best thought as a function of the final number n of points, gauging the relative importance of the two objectives.

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. This 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 framework generalizes the HOT class of models proposed, in the sense that HOT models are the trade-offs in which reliable design is one of the objectives being optimized.

3 Results

3.1 Main Results

The behavior of the model depends crucially on the value of α, and our main result 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.

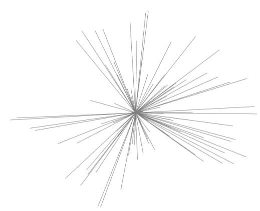

Figure 1: star

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. Again, no power law.

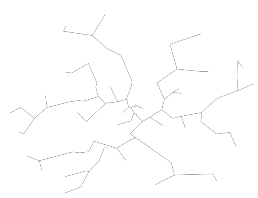

Figure 2: Euclidean minimum spanning tree

If, however, α is anywhere in between âĂŤ is larger than a certain constant, but grows slower than $\sqrt{n}$ if at all âĂŤ then, almost certainly, the degrees obey a power law.

3.2 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 following figures 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). they also show the associated trees.

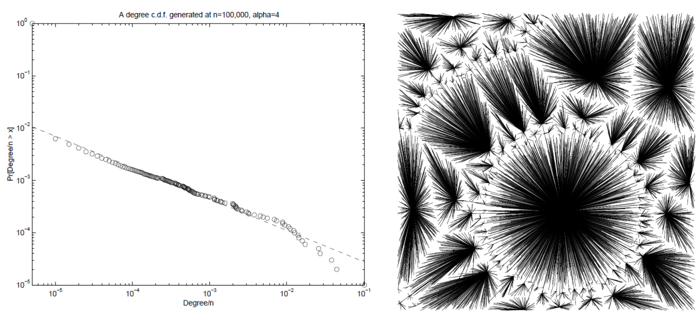

Figure 3: c.d.f and the associated tree generated for $\alpha = 4$ and n=100,000 (only the first 10,000 nodes are shown)

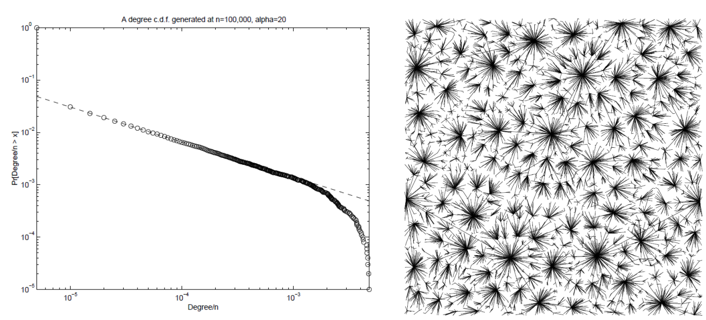

Figure 4: c.d.f and the associated tree generated for $\alpha = 20$ and n=100,000 (only the first 10,000 nodes are shown)

4 Conclusion and future work

4.1 Conclusion

As in the HOT model, power laws are seen to fall off as a by-product of resource optimization. However, only local optimizations are now needed, instead of global optimizations. This makes the heuristically optimized trade-offs model very appealing.

However, this model only generates graphs of density 1. This also implies that the graph is actually a tree, whereas many real-world graphs have cycles. Also, in addition to the degree distribution, the generated graphs need to be analyzed for all other graph patterns too. Further research needs to modify the basic model to address these issues.

4.2 Future work

I may also establish a model of file creation. For example, suppose that we are given n positive real numbers $p_1, ..., p_n$, intuitively capturing the "popularity" of n data items, the expected number of times each will be retrieved. I wish to find a partition of the items into and minimize.

It would be very interesting to extend results to other definitions of the "hop" cost, and to strengthen them by proving stronger power laws. 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.

References

[1] Alex Fabrikant, Elias Koutsoupias, Christos H. Papadimitriou(2002) Heuristically Optimized Trade-offs: A New Paradigm for Power Laws in the Internet. University of California at Berkeley, Berkeley CA 94720, U.S.A.