Does AS Size Determine Degree in AS Topology? [*]

Hongsuda Tangmunarunkit [†] USC-ISI
John Doyle Caltech
Ramesh Govindan USC-ISI
Sugih Jamin Univ. of Michigan
Scott Shenker ACIRI
Walter Willinger AT&T

1 Introduction

In a recent and much celebrated paper, Faloutsos *et al.* [6] found that the inter Autonomous System (AS) topology exhibits a power-law degree distribution. This result was quite unexpected in the networking community, and stirred significant interest in exploring the possible causes of this phenomenon.[1] The work of Barabasi *et al.* [2], and its application to network topology generation in the work of Medina *et al.* [9], have explored a promising class of models that yield strict power-law degree distributions. These models, which we will refer to collectively as the *B-A model*, describe the detailed dynamics of the network growth process, modeling the way in which connections are made between ASs. There are two simple connectivity rules that define the evolution of AS connectivity over time: *incremental growth* where a new AS connects to existing ASs, and *preferential connectivity* where the likelihood of connecting to an AS is proportional to the vertex outdegree of the target AS. These simple rules, which are similar to the classical "rich get richer" model originally proposed by Simon [12], lead to power-law degree distributions.

While the B-A model provably yields power-law vertex degree distributions, recent empirical evidence indicates that the model may not be consistent with the dynamics underlying the evolution of the actual AS topology. First, there is strong evidence [3, 4] that the degree distribution of the actual AS topology does not conform to a strict power law. However, the distribution is certainly *heavy-tailed* or *highly-variable* in the sense that the observed vertex degrees typically range over three or four orders of magnitude; in some cases, the *tail* of the degree distribution may fit a power law. These observations were gleaned from more complete pictures of AS-level connectivity (obtained by augmenting BGP route tables with peering relationships from other sources) than those used by earlier work [2, 6, 9]. Second, the B-A model's AS connectivity evolution rules can be shown to be inconsistent with empirical AS growth measurements [16]. As such, while the B-A model appears to produce topologies whose degree distribution characteristics exhibit power-law behavior, it cannot be a valid *explanation* for the connectivity evolution in the AS topology.

Clearly, some of these empirical observations don't corroborate the claim that the B-A model explains the phenomenon of highly variable vertex degrees in the Internet's AS topology [2]. However, the B-A model was originally proposed as a simple illustration of how some elementary mechanisms or rules can give rise to power law vertex degree distributions. As such, it is likely that the B-A model can be modified to accommodate these more recent findings [1], but we will neither discuss here such modifications nor comment on their possibility for success. Instead, we merely note that any such resulting model would seek, as does the original B-A model, to explain the highly variable degree distribution of the AS topology through the detailed dynamics of how connections between ASs are established.

The purpose of this note is to raise the question—motivated by the B-A approach—of whether the underlying cause of the high variability phenomenon of the vertex degree distribution lies in the detailed dynamics of network growth, or if there are alternative explanations. To that end, we briefly outline an alternative explanation for the AS topology degree distribution. We do not claim to have proven that this explanation holds; our purpose here is merely to expand the dialog to a larger class of explanations for the variability of the AS topology degree distribution.

---

[*] This work was supported in part by the Defense Advanced Research Projects Agency under grant F30602-00-2-055. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Defense Advanced Research Projects Agency.

[†] Send correspondence to hongsuda@isi.edu.

[1] As an aside, note that we do not discuss the degree distribution of the router-level Internet topology: there seems to be some debate about the characteristics of that distribution [3].

2 An Alternative Explanation

To motivate our explanation, we first note that high variability is the norm in the distribution of sizes of many real-world entities. Cities by population size [12, 17], companies by size of income [10] or by size of assets [11] are all known to exhibit power-law tails. The distribution of countries or oil reserves by size appears to exhibit a Weibullian distribution [8]. In the computing literature, file [15] and Web document [5] sizes have been known to have heavy tails.

Next, we find that AS sizes are no exception to this rule; in Section 3, we show that the distribution of the *size* of an AS (as measured by the number of routers[2] in the AS) exhibits high variability[3].

Finally, we observe, as described in Section 3, that AS sizes are highly correlated with degree. That is, large ASs tend to have large degrees and small ASs tend to have small degrees.

These observations suggest one possible, and quite general, explanation for the AS degree distribution: rather than arising from connectivity dynamics, *the highly variable degree distribution may arise merely from its correlation with a highly variable size distribution*. Assuming each individual AS corresponds to a business entity[4], that degree follows size captures the intuition that large businesses, by setting up a large initial capital investment and building out a nationwide network, are able to attract more customers and peers than smaller businesses. If our explanation turns out to be correct, it removes the mystery of AS topology degree distribution, but replaces it with the much older mystery of company size distribution. However, given that highly-variable size distributions are quite common, the reason why AS size exhibits a power law distribution may have little to do with the fact that the AS topology represents connectivity in a data communication network.

From the correlation between degree and size, and from the high-variability in degree, there is an alternative conclusion that we could draw, namely that the degree of an AS determines its size. In this case, a growth model like the one described in [9, 2] might be a plausible explanation for how such degree distributions arise. However, the ubiquity of highly-variable size distributions suggests to us that size variability is likely the cause, not the effect, of high-variability in AS degree. It is not clear to us how to establish the validity of this suggestion.

In summary, then, we raise the possibility that there exists an alternative explanation for the highly-variable degree distribution in the AS topology—namely, that (1) AS size determines AS degree and (2) AS sizes are highly-variable. This latter phenomenon, which may appear as mysterious as the original highly-variable degree distribution, is just another instance of the ubiquitous high-variability size distributions of various real-world entities.

Given that several other real-world networks exhibit highly variable degree distributions [13] one might assume that degree follows size more generally than just for the AS topology. Our initial findings suggest otherwise. In experiments with the graph of actor collaborations, where the measure of an actor's size was the number of movies the actor had participated in, we found only mild correlation between size and degree (on the order of 0.5), even though the size and degree distribution were highly-variable.

3 Methodology and Results

To determine the size distribution of ASs (recall that our definition of the size of an AS is the number of routers in the AS), we computed an *AS overlay* from a router-level topology.

Our router-level topology was collected using *Mercator*. The topology discovery methods employed by *Mercator*, and their limitations, are documented elsewhere [7]. Briefly, *Mercator* randomly probes addressable parts of the IP address space and, using traceroutes, infers adjacencies. It is also able to resolve *aliases*—interfaces belonging to the same router. The map inferred by *Mercator* is not complete, but we believe it captures a significant part of the transit portion of the network. In this note we present results from three different snapshots of the router-level topology, each taken more than six months after the previous one[5]. These topologies vary widely in size, an artifact of the different durations of each run of *Mercator*.

From the IP addresses obtained by *Mercator*, we then labelled each router with the AS it belonged to, thereby

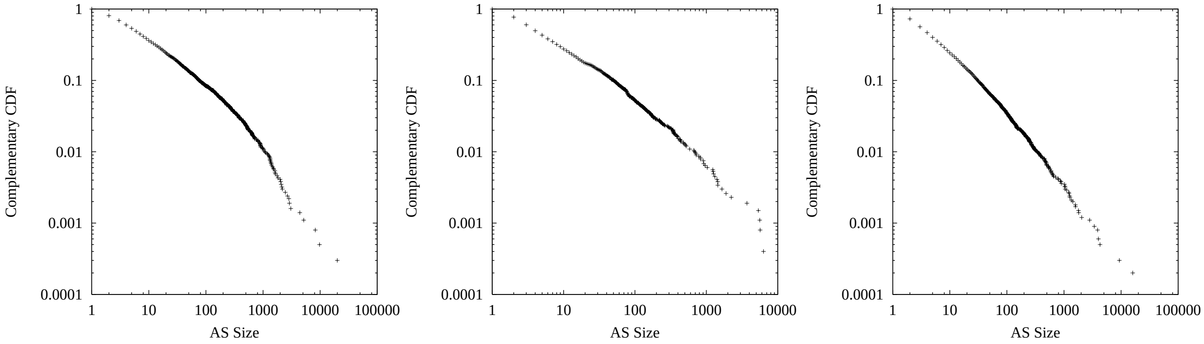
---

[2] We looked at other measures of AS size, including revenue, number of employees, and market capitalization. All these measures exhibit heavy-tails, and were correlated with AS degree. As an aside, it appears that for ASs, the number of routers in an AS is a good surrogate for any notion of "size".

[3] We emphasize that the exact form of this distribution is not of concern for the purposes of this paper. Of relevance is the qualitative observation that the distribution is highly-variable (or heavy-tailed).

[4] This assumption is only approximately true. In practice, some ISPs configure their networks to have several AS numbers.

[5] We have only computed correlations on three router-level topology snapshots. It is, of course, entirely possible that other snapshots might disprove our findings. However, we believe that the remarkably consistent results from our three snapshots argues that the likelihood of this is small.



(a) August 1999

(b) April 2000

(c) May 2001

Figure 1: Complementary cumulative distributions of AS size

generating an AS overlay. The techniques we used for associating a router with an AS, together with their limitations, are described in [14]. Briefly, the AS overlay technique uses the BGP routing tables to infer the ASs to which routers belong. In [14], we validate this technique and show how it appears to give AS maps that are qualitatively consistent with those obtained from BGP routing tables.

Having described our methodology, we now describe our main findings.

ASs Sizes are Highly-Variable Figure 1 depicts log-log plots of the complementary cumulative distribution of ASs by size. In this figure, note that these distributions are *quantitatively* different. This can be attributed to the variation in size of our three router-level topologies. In each case, however, the complementary cumulative distribution of ASs by size is highly variable, spanning 3-4 orders of magnitude.

Degrees and Sizes are Well-Correlated Figure 2 describes the coefficient of correlation between AS size and degree, where the AS degree is measured by the number of neighbors of each AS as inferred from our AS overlay. For three topologies separated by several months, the correlations are high. The figure also depicts, for the topology snapshot of May 2001, a scatterplot revealing this correlation visually (the scatterplots for other snapshots are similar).

Date	Coefficient of Correlation
August 1999	0.941
April 2000	0.936
May 2001	0.959

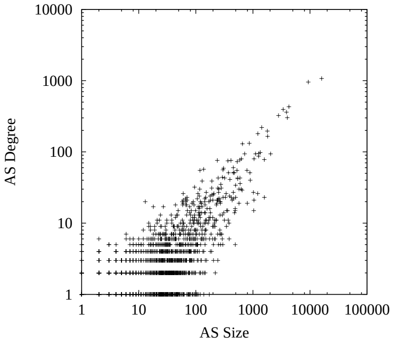


Figure 2: Correlation between size and degree

4 Conclusions

In this note, we question whether the highly variable vertex degree distribution of the Internet AS topology can indeed be attributed to connectivity dynamics, as envisioned by the B-A model. Based on the ubiquity of highly-variable size distributions, and on our observed correlation between AS size and AS degree, we ask whether there exists an alternative explanation—namely, that AS size determines degree and the high-variability in degree follows naturally from the observed high-variability in AS size.

References

[1] ALBERT, R., AND BARABASI, A.-L. Statistical Mechanics of Complex Networks. *cond-mat/0106096* (June 2001).

[2] BARABASI, A.-L., AND ALBERT, R. Emergence of Scaling in Random Networks. *Science 286* (1999), 509–512.

[3] BROIDO, A., AND CLAFFY, K. C. Internet Topology: Local Properties. In *Proceedings of SPIE ITCom 2001* (Denver, CO, August 2001).

[4] CHANG, H., GOVINDAN, R., JAMIN, S., SHENKER, S., AND WILLINGER, W. On Inferring AS-level Connectivity from BGP Route Tables. submitted to the ACM Internet Measurements Workshop 2001.

[5] CROVELLA, M. E., AND BESTAVROS, A. Self-Similarity in World-Wide Web Traffic: Evidence and Possible Causes. *IEEE/ACM Transactions on Networking 5*, 6 (December 1997), 835–846.

[6] FALOUTSOS, M., FALOUTSOS, P., AND FALOUTSOS, C. On power-law relationships of the Internet topology. *Proceedings of the ACM SIGCOMM* (Sep 1-3 1999), 251–262.

[7] GOVINDAN, R., AND TANGMUNARUNKIT, H. Heuristics for Internet Map Discovery. In *Proceedings of the IEEE Infocom* (Tel-Aviv, Israel, March 2000).

[8] LAHERRERE, J., AND SORNETTE, D. Stretched Exponential Distributions in Nature and Economy: Fat tails with characteristic scales. *European Physics Journal B*, 2 (1998), 525–539.

[9] MEDINA, A., MATTA, I., AND BYERS, J. On the Origin of Power-Laws in Internet Topologies. *ACM Computer Communications Review 30*, 2 (April 2000).

[10] OKUYAMA, K., TAKAYASU, M., AND TAKAYASU, H. Zipf's Law in income distribution of companies. *Physica A 269* (1999), 125–131.

[11] RAMSDEN, J. J., AND KISS-HAYPAL, G. Company size distribution in different countries. *Physica A 277* (2000), 220–227.

[12] SIMON, H. On a Class of Skew Distribution Functions. *Biometrika* (1953).

[13] TANGMUNARUNKIT, H., GOVINDAN, R., JAMIN, S., SHENKER, S., AND WILLINGER, W. On network topologies, power laws and hierarchy. Tech. Rep. 01-746, University of Southern California, Computer Science Department, 2001.

[14] TANGMUNARUNKIT, H., GOVINDAN, R., SHENKER, S., AND ESTRIN, D. The Impact of Policy on Internet Paths. In *Proceedings of the IEEE Infocom* (Anchorage, AK, 2001).

[15] VOGELS, W. File System Usage in Windows NT 4.0. In *Proceedings of the ACM Symposium on Operating Systems Principles* (Kiawah Island, SC, December 1999).

[16] WILLINGER, W., GOVINDAN, R., JAMIN, S., PAXSON, V., AND SHENKER, S. Scaling Phenomena in the Internet: Critically Examining Criticality. Proceedings of the National Academy of Sciences. 2001 (to appear); also available at http://topology.eecs.umich.edu.

[17] ZIPF, G. K. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, 1949.