On the Emergence of Highly Variable Distributions in the Autonomous System Topology

Marwan Fayed, Paul Krapivsky, John W. Byers, Mark Crovella, David Finkel, Sid Redner*

ABSTRACT

Recent studies observe that vertex degree in the autonomous system (AS) graph exhibits a highly variable distribution [14, 21]. The most prominent explanatory model for this phenomenon is the Barabási-Albert (B-A) model [5, 2]. A central feature of the B-A model is preferential connectivity — meaning that the likelihood a new node in a growing graph will connect to an existing node is proportional to the existing node's degree. In this paper we ask whether a more general explanation than the B-A model, and absent the assumption of preferential connectivity, is consistent with empirical data. We are motivated by two observations: first, AS degree and AS size are highly correlated [10]; and second, highly variable AS size can arise simply through exponential growth. We construct a model incorporating exponential growth in the size of the Internet and in the number of ASes, and show that it yields a size distribution exhibiting a power-law tail. In such a model, if an AS's link formation is roughly proportional to its size, then AS out-degree will also show high variability. Moreover, our approach is more flexible than previous work, since the choice of *which* AS to connect to does not impact high variability, thus can be freely specified. We instantiate such a model with empirically derived estimates of historical growth rates and show that the resulting degree distribution is in good agreement with that of real AS graphs.

*M. Fayed, J. Byers and M. Crovella are with the Dept. of Computer Science at Boston University and are supported in part by NSF grant ANI-9986397 and NSF CAREER award ANI-0093296. E-mail: {mfayed,byers,crovella}@cs.bu.edu. P. L. Krapivsky and S. Redner are with the Dept. of Physics, as well as the Center for BioDynamics and the Center for Polymer Studies, at Boston University and are supported by grants NSF DMR9978902 and ARO DAAD19-99-1-0173. E-mail: {paulk,redner}@bu.edu. D. Finkel is with the Dept. of Computer Science at Worcester Polytechnic Institute. E-mail: dfinkel@cs.wpi.edu.

1. INTRODUCTION

Many aspects of the Internet's structure are relatively unknown. These gaps in our knowledge pose problems when attempting to construct representative network topologies for simulation and modeling. In addition, filling these gaps may shed light on the forces behind the Internet's growth and the ways in which the network may fail.

One aspect of the Internet's structure that has drawn great interest is the autonomous system (AS) graph (the graph in which vertices represent ASes and edges represent AS-AS peering relationships). A particularly surprising aspect of these graphs is that vertex degree generally possesses a highly variable distribution [14, 21].

In discussing properties of the AS graph, it is useful to draw a distinction between *high variability* and *power-law tails.* High variability is a qualitative notion, referring to a probability distribution showing non-negligible values over a wide range of scales (typically at least three orders of magnitude). On the other hand, a distribution $p(\cdot)$ with power-law tails has the formal property that:

$$p(x) \sim x^{-\alpha}$$

with $\alpha > 0$, and where $a(x) \sim b(x)$ means that $\lim_{x\to\infty} a(x)/b(x) = c$.

Some authors have argued that AS vertex degree is well modeled as having power-law tails [14, 21]. Others have suggested that vertex degree does not clearly exhibit power-law tails, although it is highly variable [8]. Since such highly-variable distributions do not arise in simple random graphs, and since power-law tails do provide a simple (albeit crude) approximation for the behavior of the true distribution, a number of papers have proposed mechanisms (more complicated than purely random connection) that may give rise to power-law degree distributions in graphs [5, 19, 18].

The most prominent model attempting to explain the emergence of power-law degree distributions is the Barabási-Albert model (or B-A model) [5, 2]. In fact, it has been considered in a number of papers as a model for AS graphs [3, 7, 26, 23, 31]. The B-A model assumes the network is formed through incremental addition of nodes. In the simplest form of the model, a new node forms a connection to an existing node with probability proportional to the existing node's degree. This preferential connectivity leads to a "rich get richer" phenomenon in which high degree nodes tend to increase in degree faster than low degree nodes.

In this paper we examine whether explanations more general than the B-A model may suffice to explain *highly variable* degree distributions in the AS graph. We are motivated by two observations. First, the authors in [10] point out that AS degree is strongly correlated with AS size (measured in number of nodes) — and that AS size also shows a highly variable distribution. Second, we observe that during the last 10 years or so, the Internet has undergone exponential growth in both number of nodes and number of ASes. Under such conditions, we show here that highly variable AS sizes (and, presumably as a consequence, highly variable AS degrees) may readily arise due to exponential growth alone.

We explore these observations in this paper by constructing a simple growth model for AS graphs. Our model makes three assumptions: (1) exponential growth in the number of hosts in the network; (2) exponential growth in the number of ASes in the network; and (3) an approximately proportional relationship between AS size and degree. The resulting model shows that highly variable AS degrees may easily arise without preferential connectivity, and in fact without any global knowledge of network state by individual ASes. Indeed, in our model, the methods by which ASes select peering partners can be freely unspecified.

In our model, M (the total number of hosts) and N (the total number of ASes) are described by the simple linear growth equations $dN/dt = qN$ and $dM/dt = pM + qN$, where q and p are the growth parameters. We show that in the asymptotic time limit, this model leads to a stationary size distribution with power-law tails. We then show that if these growth rules are used to construct a graph, such that as each AS grows it forms links to other ASes in approximate proportion to its own size, then the resulting degree distribution also shows high variability.

We validate the degree distributions produced by this simple model using empirical measurements of AS degree distributions. For this purpose we use measurements from BGP tables stored at Routeviews [27], as well as overlay maps produced by mapping routers from the Mercator [16] and Skitter [29] datasets to their corresponding ASes. We find that the resulting degree distributions in our simulated graphs are in good agreement with empirical data.

We conclude that, for topology generation, it is not necessary to incorporate preferential connectivity in order to generate highly variable AS degree distributions. This leaves the door open for more practically justified bases for forming inter-AS links, *e.g.,* based on economic and geographical considerations.

In summary, in this paper we explore a model for the AS graph that is more general than the B-A model, and is based on empirical observations of Internet growth dynamics. It allows for inter-AS connections to be formed in a way that need not be based on AS degree. We show that it yields highly-variable degree distributions, and that its outputs agree well with empirical measurements of AS graph degree distribution.

2. RELATED WORK

Until recently, Internet topologies have been generated using random and hierarchical models. Among the more significant of these is work due to Zegura *et al.* [32]. That paper proposes generating smaller domain-like networks and connecting them together to create a hierarchical structure whose properties are specified by input parameters. Unfortunately, these random and hierarchical approaches fail to capture many significant attributes of Internet topology as well as the power-law models [31, 23] discussed below.

Since attention was drawn to power-laws in Internet topologies by [14], modeling efforts have shifted to reproducing these power-law properties. The most notable effort in this direction has been the Barabási-Albert preferential attachment model [5]. This model was first formulated and solved by Simon [28] and further developed by Price [11, 12]. In this model, the network is formed through incremental addition of nodes. The model's key assumption is that a new node forms a connections to an existing node based the existing node's degree. The probability that a new node will connect to an existing node i is proportional to $\Pi(i) = k_i/\Sigma_j k_j$, where k_i is the degree of node i. The resulting rate at which nodes acquire new edges is given by $\delta k_i/\delta t = k_i/2t$, where t is the time elapsed from the start of the process. The resulting degree distribution exhibits a power-law tail, with a fixed exponent of $\alpha = 3$.

Later work has built upon and extended the B-A model. The same authors in [3] extended the model to allow re-wiring, in which edges may also be deleted or moved at each timestep; this allows the exponent to vary. The work in [26] investigates the case where only a subset of all nodes in the network are available for connection. With only slight modifications to the B-A model they show that a power-law degree distribution emerges. Additionally, a "generalized linear preference" model is proposed in [7] that better matches the clustering behavior and path lengths of empirical Internet measurements. These extensions have improved the flexibility of the B-A model, albeit with a corresponding increase in complexity.

The generation of power-laws through random graph models has also received considerable recent attention. An overview of existing models appears in [1], along with a method which generalizes all of them; this family of models is analyzed in [20]. In these models, nodes are periodically added to the graph with some probability and are initially assigned an in-weight and out-weight of 1. At each timestep, t, with some fixed probability, a new directed edge is created between nodes i and j. The probability of selecting an edge from i to j is in proportion to i's out-weight and j's in-weight, respectively. Then, the out-weight of i and the in-weight of j are increased by 1; hence, at every timestep the total in-weight in the system is exactly t. This general method can generate graphs with arbitrary degree distributions, but are not proposed as realistic models for the dynamics of Internet growth.

In contrast to the approaches above which focus on reproducing statistical properties, another family of models explores the implications of optimization-based algorithms for network structure. One such model has been suggested in [13]; it assumes that nodes arrive uniformly at random within some Euclidean space, and the newly created edges attempt to balance the distance d from its new neighbour with the desire to minimize the average number of hops h to other nodes. A new node i forms an edge to j by minimizing the weighted sum $\gamma \cdot d_{ij} + h_j$. The resulting degree distribution exhibits a power-law tail. A second optimization-based model is described in [4]; this paper explores a similar heuristic but at the ISP level.

The investigation in [10] evaluates the merits of the B-A model and its applicability to the Internet. The authors conclude that, while the B-A family of models do succeed in producing power-laws, the model itself is not representative of the dynamics that drive Internet evolution: its growth processes (preferential connectivity) do not match those observed in the Internet. Also, they present evidence to suggest that AS-level degree distribution is not a pure power-law, though it is still highly variable. Based on these observations, together with evidence in [30] which links degree to size, [10] suggests that other (perhaps simpler) mechanisms decide the evolution of the Internet.

The work in this paper shows that preferential connectivity, or indeed any dependence on degree in making connection decisions, is not necessary for power-law degree distributions to emerge. Furthermore, our paper is the first model that models highly variable degree distributions as well as the size and growth of autonomous systems themselves.

3. A SIMPLE GROWTH MODEL

In this section we first motivate our model using observations regarding the rates of growth of ASes and hosts over time. We then analyze the model and explore its properties.

3.1 Exponential Growth

We start by assessing the growth of the number of ASes in the Internet. For this, we look to a history of routing number allocations made publicly available by the Internet registrars[1]. These agencies (ARIN, RIPE, and APNIC) are collectively responsible for assigning all Internet routing numbers. Each publishes a table of every AS number[2] and IP block it allocates, and the date the allocation was made.

[1]The strengths and drawbacks of various data sources for AS tracking are discussed in [15].

[2]RIPE does not publish AS number allocations, though many of these allocations have been recorded by ARIN.

Using these tables, we can measure the number of AS numbers allocated at any point in time. The result is shown in Figure 1, on (a) a linear scale and (b) a semi-log scale. Here we assume that allocations provide a good estimate for rate of growth in total number of ASes (since we are primarily interested in the overall *rate* of growth). Fitting a line to this logscale plot shows that, over the recent past, AS numbers have indeed been allocated at an exponentially growing rate. We estimate the rate of growth by the slope of the linear regression fit to the curve, or approximately 8.7×10^{-4} (units are $\ln$(ASes)/day).

The registries provide a good record of AS births, but it is inaccurate to use their records of allocated IP blocks to estimate growth of hosts in the Internet, because most IP blocks are not fully utilized. The best estimate of the number of Internet hosts seems to be that of the widely cited Internet Software Consortium's "Internet Domain Survey" (IDS) project. The host count they develop is based on a reverse DNS process; details can be found at [17].

Using the numbers published by IDS, we plot host growth in Figure 2 (again, (a) is linear scale, and (b) is semi-log scale). Although the linear regression of the whole curve fits reasonably well, we note that the slope of the curve starting about 1996 is noticeably different from the slope before that point. Using the linear fit shown in the figure, we estimate the the more conservative growth rate (the rate post-1996) to be about 1.1×10^{-3} (units are $\ln$(hosts)/day).

We emphasize that while host count may well underestimate the actual number of hosts on the Internet, we are primarily interested in estimating the rate of growth represented by the slope of the curve.

Figures 1 and 2 provide strong historical evidence of exponential growth both in size of the Internet and number of ASes. Next we construct a simple evolutionary model which relies on the observation that both measures grow exponentially.

3.2 Model Development and Analysis

We wish to construct a model which builds on the observations that the number of ASes and the number of hosts in the Internet have both grown exponentially in the recent past. Let $N(t)$ be the total number of ASes and $M(t)$ be the total number of hosts (or 'mass') in the system. The simplest growth model consistent with the observations in the previous section is mathematically described by linear equations

$$\frac{dN}{dt} = qN, \qquad \frac{dM}{dt} = pM + qN. \tag{1}$$

Here q is the rate of creation of new ASes and p is the rate of creation of new nodes. When a new AS is created, the host is given that new label, explaining the qN term in the left equation in (1). (We assume that there is no merging of ASes; moreover, we assume that links do not affect growth processes, and that hosts and links never disappear. For a model that in-

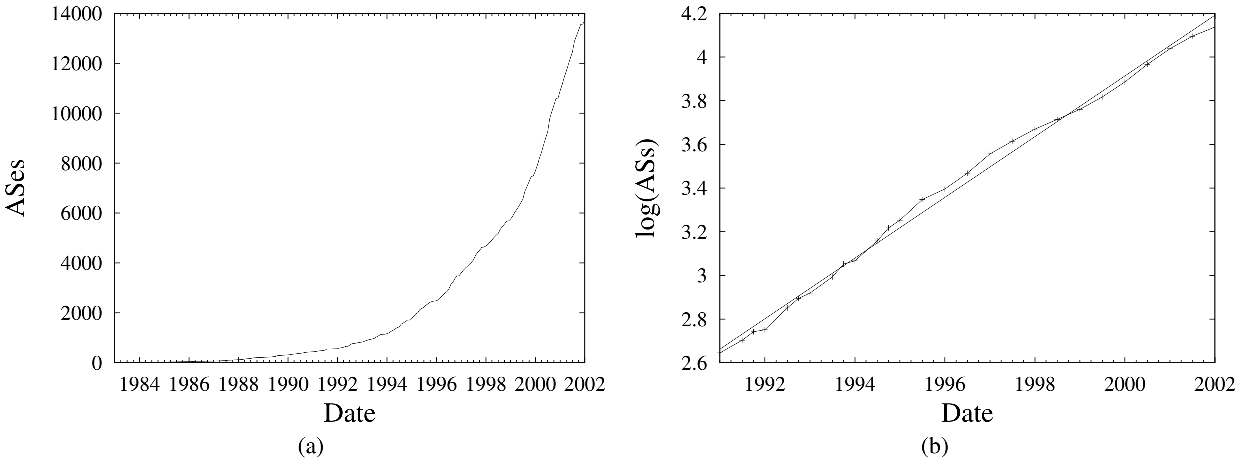

Figure 1: Growth in the number of Autonomous Systems.

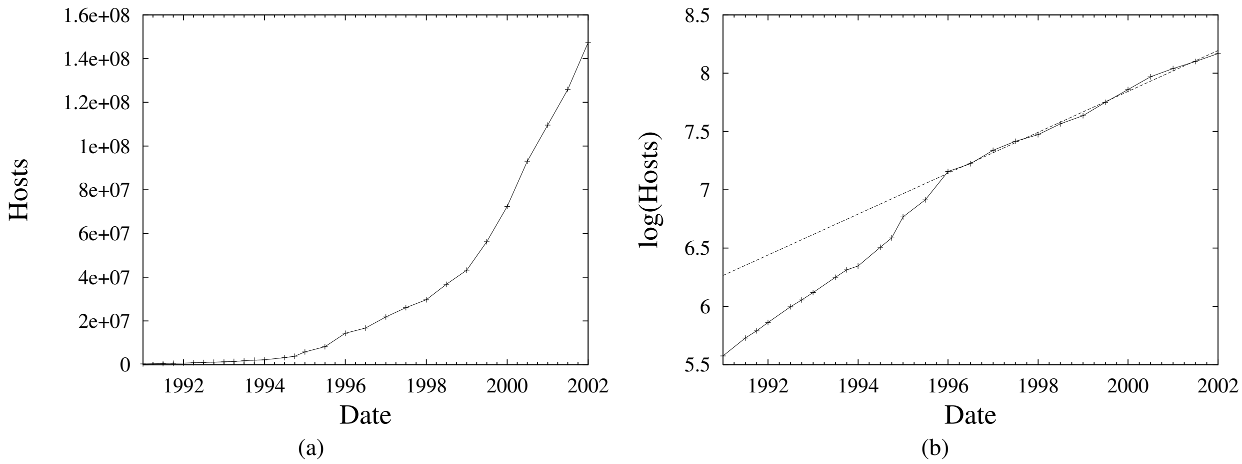

Figure 2: Growth in the number of Internet Hosts

cludes AS mergers, see [15].) Solving for N and M gives

$$
\begin{aligned}
N(t) &= N(0)\, e^{qt}, && (2)\\
M(t) &= A\, e^{pt} + BN(t), && (3)
\end{aligned}
$$

with A, B being simple functions of the initial data, and the parameters p and q. (At the special point $p = q$ the coefficients diverge ($A = B = \infty$), reflecting that the exact solution is actually a linear combination of e^{pt} and $t\, e^{pt}$.) Thus the average AS size $\langle s \rangle \equiv M(t)/N(t)$ could exhibit the following asymptotic behaviors:

$$
\langle s \rangle \sim \begin{cases} \text{finite} & \text{when } p < q, \\ \ln N & \text{when } p = q, \\ N^{(p-q)/q} & \text{when } p > q. \end{cases} \tag{4}
$$

In [15] we show that the average AS size grows over time (and with N), in agreement with measurements showing that $p > q$.

Let $N_s(t)$ be the number of ASes with s nodes. Unlike $N(t)$ and $M(t)$, $N_s(t)$ is a random variable; however in the large time limit it becomes highly localized around its average value. Under the assumption that $N_s(t)$ is equal to its expectation, its size distribution satisfies the rate equation

$$
\frac{dN_s}{dt} = p\left[(s-1)N_{s-1} - sN_s\right] + qN\delta_{s,1} \tag{5}
$$

where $\delta_{s,1}$ is the Kronecker delta function. The first term on the right-hand side accounts for growth that proceeds with rate p: when a node is added to an AS with $s-1$ nodes, the number of ASes with s nodes increases by one; similarly when a node is added to an AS with s nodes, the number of ASes with s nodes decreases by one. The second term on the right-hand side of Eq. (5) accounts for births of new ASes (with size one) which take place with rate q.

We already know $N(t) = N(0)\, e^{qt}$. Solving Eqs. (5) recursively and expressing in terms of N rather than t yields

$$N_s = n_s N + \sum_{j=1}^{s} C_{sj} N^{-jp/q} \tag{6}$$

where n_s is the fraction of ASes with s nodes. The coefficients C_{sj} depend on initial conditions while n_s are universal. Asymptotically, only the linear term $n_s N$ matters. To determine this dominant contribution, we insert $N_s(t) = n_s N(t)$ into Eq. (5). We arrive at the recursion relation

$$\left(s + \frac{q}{p}\right) n_s = (s-1) n_{s-1} \tag{7}$$

for $s \geq 2$, while for $s = 1$ we have $n_1 = q/(q+p)$. The solution to recursion (7) reads

$$n_s = \frac{q}{q+p}\, \frac{\Gamma(s)\, \Gamma\left(2 + \frac{q}{p}\right)}{\Gamma\left(s + 1 + \frac{q}{p}\right)}. \tag{8}$$

Asymptotically, the ratio of gamma functions simplifies to the power law,

$$n_s \sim C\, s^{-\alpha}, \tag{9}$$

with $\alpha = 1 + q/p$ and $C = \frac{q}{q+p}\, \Gamma\left(2 + \frac{q}{p}\right)$. That is, the model yields an AS size distribution exhibiting a power-law tail with exponent $-(1 + q/p)$.

4. AS DEGREE FORMATION

The previous section showed that a power-law size distribution emerges in the presence of exponential growth of ASes and hosts. In this section we extend this idea to incorporate AS degree.

The key assumption we make is that as an AS grows, it will establish links with other ASes. We show that if link formation occurs in rough proportion to an AS's growth, then the AS degree distribution will show high variability. More precisely, if at each time step a new node is added to an AS it forms an inter-AS link to some other randomly chosen AS with a fixed probability, then AS degree distribution will show high variability. Furthermore, this need only be in "rough proportion;" for example, the result still holds if connection probability varies with the log of the AS size.

Any such link formation process is simple since it only depends on growth, it is flexible since there are no influencing agents other than size, and no global knowledge of other AS degrees is required to make link formation decisions.

We embody the growth model of Section 3.2 and the aforementioned link formation process in the following algorithm. Recall the notation from Section 3.2 where t is time and $N(t)$ is the number of ASes in the system. Let $M_i(t)$ be the number of hosts in AS i, and t_i be the time AS i is introduced into the system.

This algorithm requires three input parameters: p and q are the growth rates in the number of hosts and ASes, respectively, and x is some fixed probability of an inter-AS edge forming with each unit of host growth. At each timestep t two kinds of events occur: some new ASes are born, and existing ASes grow. Starting at $t = 1$:

i. Calculate the total number of ASes according to $N(t) = N(0)e^{qt}$. In our simulations we use $N(0) = 1$.

ii. Introduce $\lfloor N(t) \rfloor - \lfloor N(t-1) \rfloor$ new ASes with out-degree of 1, where the neighboring AS is chosen uniformly at random.

iii. For each AS i in existence before time t,

   (a) Calculate the number of total hosts within AS i according to $M_i(t) = Ae^{p(t-t_i)} + BN(t)$. In our simulations we use $A = 1, B = 0$.

   (b) Insert $\lfloor M_i(t) \rfloor - \lfloor M_i(t-1) \rfloor$ new hosts into AS i. Each new host creates an inter-AS edge with probability x, and if an edge is created, then invoke a *select* operation to determine to whom the new AS-to-AS link is created.

The *select* operation is left unspecified to emphasize the flexibility of the link formation process and its dependence only on the AS size. We consider only the simplest selection operation, where a target AS is chosen uniformly at random.

Even though this is a random connection process, ASes that are larger in size will generally have higher out-degree. Thus, like the underlying size distribution, the degree distribution that results will tend to be highly variable. We show in the following sections that a highly variable degree distribution does result, and that this distribution fits well when compared against distributions observed in the Internet.

5. VALIDATION

We validate our analysis and simulation results against empirical degree distributions in the following sections.

5.1 Empirical Data Sources

There are a number of sources from which we can draw AS-level degree distribution. We infer empirical degree distribution through two distinct methods, applied to three different sources.

The first method is to infer AS degrees from BGP tables. For this purpose we use BGP tables from the RouteViews project

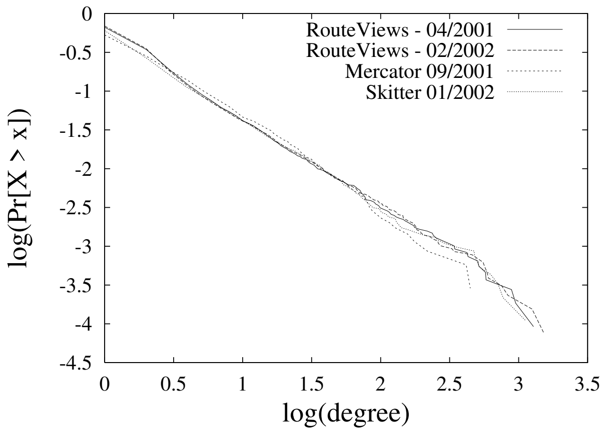

Figure 3: Degree Distributions Inferred from 4 Sources.

[27] collected in April 2001 and February 2002. An entry in a BGP table consists of an IP block represented by its prefix, followed by a sequence of ASes (an AS path) that must be traversed to reach an IP address within that range. We can infer an adjacency in the AS-level graph for a pair of ASes whenever they appear in succession within any path. While this inference method typically avoids false positives (adjacencies which are not actually present, but appear to be present), it suffers from false negatives, since not all AS adjacencies are advertised across BGP [10].

A second method for determining AS degrees is to annotate a router-level map with each router's associated autonomous system. Nodes in the router-level graph are labeled using IP addresses. In the overlay produced by annotating the router-level graph, each node is further labeled with its assocated AS. The approach is detailed in [9]; we summarize the approach here. An IP is associated with an autonomous system by performing a lookup in BGP tables. First, find the longest matching prefix of an IP address within the BGP table; the last entry in the path vector is the number of the AS which owns that IP address. A complete inspection of every edge in the annotated router-level graph reveals an inter-AS edge wherever any pair of nodes are labeled with distinct AS numbers.

This method has numerous advantages over AS maps inferred from BGP tables directly. It provides an AS map at a finer granularity; aggregated ASes are revealed, as are multiple links between ASes. However, this method suffers from the following drawback. Any single BGP table is potentially incomplete and can be limited by path hiding from parent ASes (in order to reduce message and table sizes). Sets of BGP tables are used to reduce the magnitude of this problem, with the belief that more BGP tables reveal more information. However, no AS can observe the existence of another AS which is hidden by its parents.

We draw on router-level maps gathered from the Mercator project [16] in August 2001, and another provided by the Skitter project [29] gathered in January 2002. Statistics, dates, and sources of all four datasets are summarized in Table 1.

Source	ASes	Edges	Date	Method
Route Views	10854	47847	04/01	BGP
Route Views	12875	57385	02/02	BGP
Mercator	3478	13590	08/01	AS Overlay
Skitter	9206	38334	01/02	AS Overlay

Table 1: Summary of Data Sources

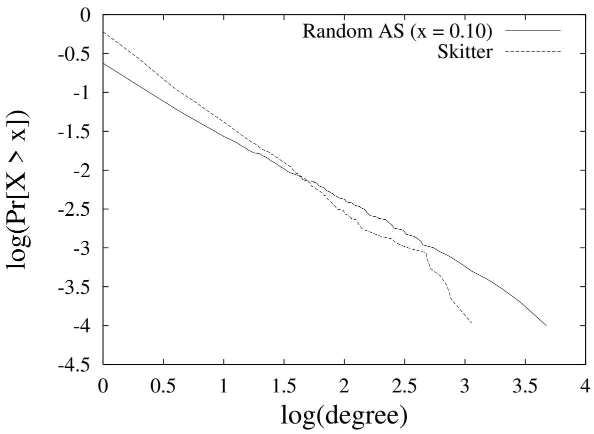

Figure 4: Predicted Degree Distribution under the Constant Connectivity Model.

The degree distributions plotted in Figure 3 show that all methods and sources yield similar results. For subsequent comparisons, we use the distribution drawn from the autonomous system overlay constructed from the Skitter dataset collected in January 2002 as a baseline for comparison against simulation results.

5.2 Constant Connectivity Models

Section 3.2 shows that the size distribution that results from our model has a power-law tail. However, since the growth model does not directly describe degree, we turn to our simulation to determine the influence of size and growth on degree.

The simulation is executed using the algorithm in Section 4 using rates $p = 1.1 \times 10^{-3}$ and $q = 8.7 \times 10^{-4}$ estimated in Section 3. The degree distribution predicted by our model is plotted against observed degree distributions in Figure 4. We found empirically that using fixed connection probability $x = 0.10$ results in vertices of our simulated graphs having a roughly commensurate average degree to that of the Skitter dataset. Where the discrepancy does occur, the general tendency is for our model to underestimate the degree of small to medium sized ASes, while overestimating the degree of larger ASes.

Size Range	Data Points Mercator	Data Points Skitter	Avg. Conductance Mercator	Avg. Conductance Skitter
2-10	1404	4254	0.49	0.87
11-100	1429	3502	0.24	0.60
101-1000	359	1050	0.13	0.31
1001-10000	38	131	0.11	0.21
10001-100000	1	10	0.20	0.25

Table 2: Conductance of ASes

Figure 4 shows that the predicted degree distribution is reasonably similar to that of the Skitter dataset. Discrepancies can potentially be removed by refining the decision processes used to form AS to AS connections in the model. In the following section, we explore a refined model that accounts for the size of the AS when determining the relationship between growth and link formation.

5.3 Size-Based Connectivity Models

The relationship between predicted and empirical distributions shown in Figure 4 suggest that there is room for other practical influences on inter-AS link formation. Here we discuss an approach that takes into account the actual size of the AS when choosing to create new links.

We presuppose the following notion: as an AS grows, the ratio of its degree to its size will shrink, and so a constant probability when deciding to create new links may not best relate degree to size. Intuitively, the ratio between the degree of an AS and its size is analogous to surface-to-volume ratio. In graph-theoretic terms, this ratio is often referred to as the conductance of a subgraph. Thus, we define the conductance of an autonomous system i with size M_i and degree d_i to be $\frac{d_i}{M_i}$.

Observations of conductance are estimated from Mercator and Skitter datasets discussed in Section 5.1, and shown in Table 2. This table shows that as an autonomous system grows, the average conductance shrinks. While the actual conductance of ASes of a given size varies considerably, this trend holds on average. Note that ASes of size 1 are excluded from the smallest range since an AS of size 1 *must* have conductance of at least 1, and so may bias observations. Also, average conductance in the largest ASes appear to break this trend. We believe that this may be an artifact of noise from a small number of data points.

We believe that this decrease in conductance is natural, driven by the decreasing necessity to add inter-AS links as an AS grows. For example, as previously mentioned, an AS of size 1 must have a minimum degree of 1 (otherwise it is not connected to other ASes, and hence cannot be a part of the AS-level map). We speculate that it is more often the case that hosts are added to a closed network to increase the capacity and range of the network itself, rather than to connect to other ASes, and so a connection probability that decreases as an AS grows is reasonable.

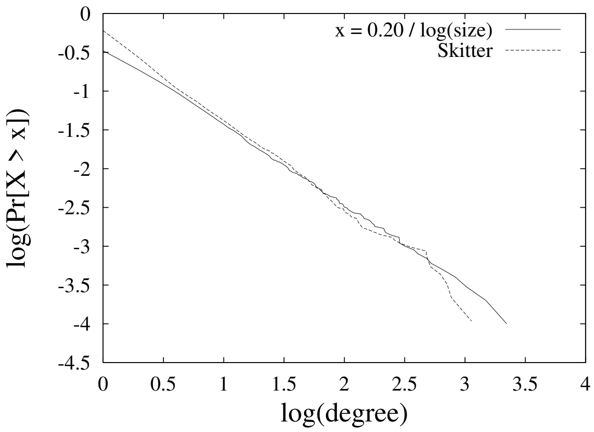

Figure 5: Validation of the Size-Based Connectivity Model.

The ratios and ranges in Table 2 show conductance diminishing in an approximately logarithmic fashion as AS size increases. To better fit the data observed in Table 2, we applied a logarithmic correction factor to implement a "diminishing probability" function, L. This function takes the size of the autonomous system M_i, and a fixed probability x as parameters, and returns a probability value:

$$L(x, M_i) = \begin{cases} x & \text{when } M_i < 10, \\ \frac{x}{\log_{10}(M_i)} & \text{otherwise.} \end{cases} \quad (10)$$

As before, we use the simple *select* operation which returns a neighboring AS chosen uniformly at random.

The distribution that results when applying the diminishing probability function is plotted against Skitter data in Figure 5, using $x = 0.20$, the value providing the best fit. The two curves are nearly identical, sharing a similar slope, and are virtually indistinguishable throughout the entire body of the distribution.

6. CONCLUSIONS

In this paper we have explored a model for how highly variable degree distributions may arise in the AS graph. It is instructive to compare this model with the B-A model.

Like the B-A model, we assume that high variability has arisen via a "rich get richer" phenomenon resulting from an exponential growth process. However the B-A model assumes *preferential connectivity,* meaning that new nodes probabilistically prefer to connect to well-connected existing nodes. Besides requiring that each AS be aware of the degree of each other AS (a strong assumption of global knowledge), the B-A model strongly constrains the resulting connection pattern. This is restrictive; as discussed in [25], many graph realizations are consistent with a given degree sequence, and different realizations may have very different properties. In fact, [24] shows that the AS graph exhibits a high degree of clustering, an effect that is not captured by the particular connection pattern created by the B-A model.

In contrast, the assumption in our model is that *AS sizes* are the underlying cause of high variability, and that a large AS will naturally tend to have a large degree. From this standpoint, our model allows for a much wider range of connection patterns than the B-A model, since the degree of an AS grows as a function of its size, but the choice of *which* AS to connect to can be specified independently, as a separate selection operation. In this paper we have explored the selection operation in which growing ASes choose peering partners uniformly at random; however we expect that any choice of peering partners that is made without regard to degree (and including those that exhibit a high degree of clustering) will likely show characteristic high variability.

Our results demonstrate that a simple and natural model incorporating exponential growth alone is sufficient to drive both a highly variable AS size distribution and a highly variable AS degree distribution. We motivated this model with datasets that demonstrate exponential growth both in the number of hosts and the number of ASes, and validated the model by comparing the degree distribution our model predicts against observed degree distributions drawn from BGP tables and AS overlay maps. We also provide an analysis of the power-law tail of the AS size distribution that results when our methods are employed.

We have integrated this model into the publicly available BRITE [6, 22] topology generation framework. In future work, we intend to investigate selection operations that incorporate real-world considerations such as locality, clustering and performance optimization, to provide an even more realistic AS growth model. As part of this effort, we are mining AS time-series data extracted from BGP logs to better understand the underlying nature of AS growth, interconnection and merging over time [15].

7. REFERENCES

[1] W. Aiello, F. Chung, and L. Lu. Random evolution in massive graphs. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, Las Vegas, NV, October 2001.

[2] R. Albert and A. Barabási. Topology of evolving networks: local events and universality. *Phys. Rev. Lett.*, 85:5234–5237, 2000.

[3] R. Albert and A. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, January 2002.

[4] D. Alderson, J. Doyle, R. Govindan, and W. Willinger. Toward an Optimization-Driven Framework for Designing and Generating Realistic Internet Topologies. In *ACM HotNets-I*, Princeton, NJ, October 2002.

[5] A.-L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286:509–512, October 1999.

[6] Boston University Representative Internet Topology Generator (BRITE). At *http://www.cs.bu.edu/brite*.

[7] T. Bu and D. Towsley. On Distinguishing between Internet Power Law Topology Generators. In *Proceedings of IEEE INFOCOM*, New York, NY, June 2002.

[8] H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Towards Capturing Representative AS-Level Internet Topologies. Technical Report UM-CSE-TR-454-02, University of Michigan Computer Science, 2002.

[9] H. Chang, S. Jamin, and W. Willinger. Inferring AS-level Internet Topology from Router-Level Path Traces. In *Proceedings of SPIE ITCom 2001*, Denver, CO, August 2001.

[10] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. The Origin of Power Laws in Internet Topologies Revisited. In *Proceedings of IEEE INFOCOM*, New York, NY, June 2002.

[11] D. J. deS. Price. Networks of scientific papers. *Science*, 149:510, 1965.

[12] D. J. deS. Price. A general theory of bibliometric and other cumulative advantage processes. *J. Amer. Soc. Inform. Sci.*, 27:292, 1976.

[13] A. Fabrikant, E. Koutsoupias, and C.H. Papadimitriou. Heuristically Optimized Trade-offs: A New Paradigm for Power Laws in the Internet. In *International Colloqium on Automata, Languages, and Programming (ICALP)*, Málaga, Spain, July 2002.

[14] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On Power-Law Relationships of the Internet Topology. In *Proceedings of ACM SIGCOMM*, pages 251–62, Cambridge, MA, September 1999.

[15] M. Fayed, P. Krapivsky, J. Byers, M. Crovella, D. Finkel, and S. Redner. On the size distribution of autonomous systems. Technical Report BUCS-TR-2003-001, Boston University, January 2003.

[16] R. Govindan and H. Tangmunarunkit. Heuristics for Internet Map Discovery. In *Proceedings of IEEE INFOCOM*, March 2000.

[17] Internet Domain Survey. At *http://http://www.isc.org/ids/*.

[18] P. L. Krapivsky and S. Redner. Organization of Growing Random Networks. *Phys. Rev. E*, 63:066123, 2001.

[19] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of Growing Random Networks. *Phys. Rev. Lett.*, 85:4629–4632, 2000.

[20] P. L. Krapivsky, G. J. Rodgers, and S. Redner. Degree distributions of growing networks. *Phys. Rev. Lett.*, 86:5401, 2001.

[21] D. Magoni and J. Pansiot. Analysis of the Autonomous System Network Topology. *ACM SIGCOMM Computer Communication Review (CCR)*, 31(3):26–37, July 2001.

[22] A. Medina, A. Lakhina, I. Matta, and J. Byers. BRITE: An Approach to Universal Topology Generation. In *IEEE MASCOTS*, Cincinnati, OH, August 2001.

[23] A. Medina, I. Matta, and J. Byers. On the Origin of Power Laws in Internet Topologies. *ACM SIGCOMM Computer Communication Review (CCR)*, 30(2):18–28, April 2000.

[24] M. Mihail, C. Gkantsidis, and E. Zegura. Spectral analysis of Internet topologies. In *Proceedings of IEEE INFOCOM*, April 2003.

[25] M. Mihail and N. Visnoi. On generating graphs with prescribed degree sequences for complex network modeling applications. In *Proceedings of Approx. and Randomized Algorithms for Communication Networks (ARACNE)*, 2002.

[26] S. Mossa, M. Barthélémy, H. Stanley, and L.A. Nunes Amaral. Truncation of Power Law Behavior in "Scale-Free" Network Models due to Information Filtering. *Phys. Rev. Lett.*, 88(13):138701, March 2002.

[27] RouteViews Project at University of Oregon. At *http://archive.routeviews.org/*.

[28] H.A. Simon. On a class of skew distribution functions. *Biometrica*, 42:425, 1955.

[29] The Skitter Project. At *http://www.caida.org/tools/measurement/skitter/*.

[30] H. Tangmunarunkit, J. Doyle, R. Govindan, S. Jamin, and S. Shenker. Does AS Size Determine Degree in AS Topology? *ACM SIGCOMM Computer Communication Review (CCR)*, 31(5):7, October 2001.

[31] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Network Topology Generators: Degree-Based vs. Structural. In *Proceedings of ACM SIGCOMM*, Pittsburgh, PA, August 2002.

[32] E. Zegura, K. Calvert, and M. J. Donahoo. A quantitative comparison of graph-based models for Internet topology. *IEEE/ACM Transactions on Networking*, 5(6), December 1997.