Accepted for publication in the *IEEE INFOCOM 2004*

Comparing static and dynamic measurements and models of the Internet's topology

Seung-Taek Park[1] David M. Pennock[3] C. Lee Giles [1,2]

[1]Department of Computer Science and Engineering
[2]School of Information Sciences and Technology
Pennsylvania State University
University Park, PA 16802 USA
separk@cse.psu.edu, giles@ist.psu.edu

[3] Overture Services Inc.
74 N. Pasadena Ave. 3rd floor
Pasadena, CA, 91103 USA
david.pennock@overture.com

***Abstract*— Capturing a precise snapshot of the Internet's topology is nearly impossible. Recent efforts have produced topologies with noticeably divergent characteristics [1], [2], [3], even calling into question the widespread belief that the Internet's degree distribution follows a power law. In turn, this casts doubt on Internet modeling efforts, since validating a model on one data set does little to ensure validity on another data set, or on the (unknown) actual Internet topology. We examine six metrics—three existing metrics and three of our own—applied to two large publicly-available topology data sets. Certain metrics highlight differences between the two topologies, while one of our static metrics and several dynamic metrics display an invariance between the data sets. Invariant metrics may capture properties inherent to the Internet and independent of measurement methodology, and so may serve as better gauges for validating models. We continue by testing nine models—seven existing models and two of our own—according to these metrics applied to the two data sets. We distinguish between growth models that explicitly add nodes and links over time in a dynamic process, and static models that add all nodes and links in a batch process. All existing growth models show poor performance according to at least one metric, and only one existing static model, called *Inet*, matches all metrics well. Our two new models—growth models that are modest extensions of one of the simplest existing growth models—perform better than any other growth model across all metrics. Compared with *Inet*, our models are very simple. As growth models, they provide a possible explanation for the processes underlying the Internet's growth, explaining, for example, why the Internet's degree distribution is more skewed than baseline models would predict.**

keywords: Simulations, Network measurements, Graph theory, Statistics.

I. INTRODUCTION

Researchers have explored characteristics and models of the Internet, mainly validating their conclusions using *Oregon RouteViews* (hereafter, simply *Oregon*), a well-known collection of (sampled) snapshots of the Internet's autonomous-systems (AS) level topology. Because of the Internet's distributed nature, recording an accurate picture of its topology at any given time is nearly impossible, casting some doubt on the validity of measurements and models based on necessarily incomplete data. Recently, using new methodologies for measuring the Internet's topology, researchers have created an extended source of data [2], [3] (hereafter, simply *Extended*), combining several existing sources, including *Oregon*, *Looking Glass*, *RIPE*, and other publicly available full BGP routing tables, and capturing 20-50% more physical links than *Oregon*. Since most pronouncements regarding Internet characteristics and models—including the most cited property of a power-law degree distribution—are based on *Oregon* data, the new findings raise several questions.

- *What are the differences in characteristics of the Oregon and Extended topology data sets?* Researchers have looked at differences in the two topologies' degree distributions, though other characteristics of the *Extended* topologies are still largely unexplored.
- *What metrics, if any, are invariant between the two topologies?* Even *Extended* is a partial view of the true Internet topology; it is not clear whether *Oregon* or *Extended* better represents the true Internet, or if neither represent it well enough. However, identifying meaningful *invariant* metrics that are the same for both data sets may help identify properties inherent to the Internet and less dependent on measurement methodology, and help validate competing In-

ternet models.

- *What models match with characteristics observed in the two data sets? To what extent do those models capture some essential aspect of the Internet's growth mechanism?* Models must be evaluated on two (often conflicting) dimensions: (1) their correspondence with data, and (2) their ability to abstract away inessential details while retaining some essential aspects of the system being modeled.

To begin to answer the first two questions, we compare *Oregon* and *Extended* using three existing metrics and three new metrics of our own: link-degree ratio, average node-degree ratio, and skewness. We find that, while the two data sets diverge according to most metrics, they agree nearly perfectly according to average node-degree ratio, suggesting that this metric is a good candidate for an invariant measure. We also find that, though most of the metrics' absolute values differ, their relative changes over time are very similar between the two data sets. So dynamic changes in metrics over time may serve as additional candidate invariant measures.

In response to the third question, we compare the performance of nine generative models of the Internet, two of which are new. We examine both *growth models* that posit a particular mechanism of growth over time, and *static models* that input a number of nodes and edges and generate graphs all at once, without explicitly formulating a growth procedure. Among existing growth models, a subset show relatively good performance on some static metrics, though none follow the observed dynamic behavior of the Internet. A static model called *Inet* does well at matching both static and dynamic Internet characteristics, but may be over-tuned to the data(*Oregon*); the model says little about the underlying processes governing Internet growth, only mimicking it using a quite complicated procedure. In short, we believe that, while *Inet* certainly excels according to the first criteria of a good model (item (1) of question three above), it arguably falls short according to the second criteria (item (2) of question three). Our new models, on the other hand, are quite simple, and do make statements about the potential mechanisms underlying Internet growth. Our models fit the static characteristics of the Internet more closely than any other growth model, and as closely as *Inet*. However our models still fail to capture the dynamic evolution of the Internet; it remains an open problem to discover a plausible growth mechanism that meshes well with the dynamic characteristics clearly visible in both *Oregon* and *Extended* data.

II. Previous Work

The Internet's topology has been studied at macroscopic level [4], the link architecture [5], [6], the end-to-end path level [7], [8]. Authors have also looked at temporal characteristics stemming from properties of its connectivity and growth [9], [10], [11]. Scaling factors, such as power-law relationships and Zipf distributions, arise in all aspects of network topology [4], [12] and web-site hub performance [13].

Recent research [14], [15], [16], [17], [18], [19] has argued that the performance of network protocols can be seriously effected by the network topology and that building an effective topology generator is at least as important as protocol simulations. Previously, the Waxman generator [20], which is a variant of the Erdos-Renyi random graph [21], was widely used for protocol simulation. In this generator, the probability of link creation depends on the Euclidean distance between two nodes. However, since real network topologies have a hierarchical rather than random structure, next generation network generators such as Transit-Stub [22] and Tiers [23], which explicitly inject hierarchical structure into the network, were subsequently used. In 1999, Faloutsos et al. [4] discovered several power-law distributions in Internet data, leading to the creation of new Internet topology generators.

Tangmunarunkit et al. [24] divide network topology generators into two categories: *structural* and *degree-based* network generators. The major difference between these two categories is that the former explicitly injects hierarchical structure into the network, while the later generates graphs with power-law degree distributions without any consideration of network hierarchy. Tangmunarunkit et al. argue that even though degree-based topology generators do not enforce hierarchical structure in graphs, they present a loose hierarchical structure, which is well matched to real Internet topology. Other recently proposed generators [25], [12], [26], [27], [28], [29] can be thought of as degree-based generators.

Characteristics of the Internet topology and its robustness against failures have been widely studied [25], [30], [4], [12], [31], with focus on extracting common regularities from several snapshots of the real Internet topology (e.g., power-law degree distributions). Properties measured on a single snapshot of the Internet's topology at a given time are examples of *static* metrics. On the other hand, researchers have shown that, for example, the clustering coefficient of the Internet is growing while the average diameter is decreasing over the past few years [29], [32]. A second class of reasonable metrics for characterizing the Internet are such *dynamic* metrics.

Park et al. [31], in examining the fault tolerance properties of Internet network models, also uncover some dynamic patterns of the real Internet's growth that are not captured by most existing models. One could of course simulate network protocols (and failures) using the full details of the sampled Internet topology instead of using models, but this limits one's ability to develop, for example, network protocols that best fit future conditions. Though degree-based generators seem to represent the Internet's topology better than structural ones, some degree-based topology generators seem to try more to mimic generic properties than to provide explanatory power regarding the Internet's growth mechanism.

III. Comparison of two Internet AS topologies

Recently, [2], [3] provided more extended Internet topologies constructed using several sources– including *Oregon RouteViews*, *Looking Glass* data, *RIPE* database, and other publicly available full BGP routing tables. Their extended topologies contain more nodes (2%) and links ($20\% \sim 50\%$ more). Also, degree-frequency distributions of their extended topologies do not follow strict power-law distribution while original topologies do. Chen et. al reported that their extended topologies showed more ASs with degree between 4 and 300, resulting in a curve line in the distribution. This result is shown in Figure 1 (a).

Then, our first question is that how different two topologies are. Since two topologies are still partial of the whole Internet topology and we do not really know which one is more similar to the real Internet topology, we compare two topologies with several metrics. Our second question is that which metrics will be more consistent over two Internet topologies. If we can find them, these metric will be quite useful to determine validation of existing Internet models.

Characteristics of the Internet topology can be divided into two categories: *static* and *dynamic* characteristics [31]. For example, several common regularities, e.g., power-law degree distributions, can be extracted from a snapshot of the Internet topology and those regularities can be defined as *static characteristics* because of their consistency over time. On the other hand, several growth patterns of the Internet can be derived by tracing the behaviors of the Internet topologies over time. For example, clustering coefficient of the Internet has been growing and average diameter of the Internet has been decreasing over the past few years. We define these as *dynamic characteristics* of the Internet. Based on these definitions, we choose six basic metrics, three static (including two new metrics of our own) and three dynamic metrics (including one of our own), for our analysis. In the following section, we will briefly explain these metrics.

A. Metrics

1) Static metrics: Our first static metric is the cumulative *degree-frequency* distribution. It is well known that the degree distribution of the Internet follows a power law. Let V be the set of all nodes in the graph and V_k the set of nodes of degree equal or less than k. Then, $F(k) = |V_k|/|V|$. On plots of the degree distribution, the horizontal axis is the degree of nodes and the vertical axis plots $1 - F(k)$.

We define a second metric called the cumulative *link-degree ratio* distribution. Let low_i (lower degree node) and $high_i$ (higher degree node) be the the two nodes connected by link i. k^i_{low} denotes the degree of the lower degree node and k^i_{high} denotes the degree of the higher degree node. Then the degree ratio σ_i of the link i can be calculated as k^i_{low}/k^i_{high}. The cumulative distribution of σ can be drawn similarly to the previous metric.

Finally, we define a third metric called the cumulative *average-node-degree ratio* distribution. Let V'_i be the set of neighbor nodes of the node i, and let k^i_{avg} be the average degree of V'_i. Then the average-node-degree ratio δ_i of node i is defined as k_i/k^i_{avg}. The cumulative distribution of δ can be drawn as above.

2) Dynamic metrics: We use three metrics for tracing the behavior of the Internet topology over time.

We define *skewness* to measure how preferential the network is. Consider the degree-rank distribution of a network. Let n denote the number of nodes in the network and r_i be the rank of node i according to its degree. The highest degree node has rank one and any two nodes cannot have the same rank. Skewness Sk is defined as that the sum over all nodes of the product of rank times degree:

$$Sk = \frac{\sum_r (r_i * k_i)}{Sk_u} \tag{1}$$

where Sk_u is the skewness of an idealized uniform network,

$$Sk_u = \sum_i (r_i * k_i) = \overline{k} * \sum_{r=1}^{n} r = \overline{k} * \frac{n*(n+1)}{2}, \tag{2}$$

where $\overline{k}$ denotes the average (uniform) degree of the network.

Note that Sk_u is upper bound of $\sum_r (r_i * k_i)$, so $1 \geq Sk > 0$. Sk values close to 0 mean that the network is extremely preferential; Sk values close to 1 means that the network is extremely random or uniform.

Average diameter and *clustering coefficient* [33], [30], [34] are widely used metrics for the analysis of networks. Average diameter or average shortest path length, $\overline{d}$, is defined as follows. Let $d(v,w)$ be the length of the shortest path between nodes v and w, where $d(v,w)=\infty$ if there is no path between v and w. Let Π denote the number of distinct node pairs (v,w) such that $d(v,w)\neq\infty$.

$$\overline{d} = \frac{\sum_{(v,w)\in\Pi} d(v,w)}{|\Pi|}, \tag{3}$$

where $v \neq w$.

The clustering coefficient gives a measure of the probability of connection between node i's neighbors. Let V_i be the set of neighbor nodes of node i, and μ a number of links between neighbors. Then, the clustering coefficient C_i for node i is defined as follows:

$$C_i = \frac{\mu}{|V_i| * (|V_i|-1)/2}. \tag{4}$$

Then the clustering coefficient of the network is:

$$C = \frac{\sum_{i\in V} Ci}{|V|}, \tag{5}$$

where V denotes a set of all nodes in the network.

B. Comparing the Oregon *and* Extended *Internet topologies*

1) Static measurements: Among other findings, the creators of the *Extended* data set noticed that their measurements do not corroborate the strict power-law degree-frequency distribution that the *Oregon* data display. This is recreated in Figure 1(a). We find that the separation between the two data sets is even larger when examined according to link-degree ratio, as seen in Figure 1(b). However, according to average-node-degree ratio, plotted in Figure 1(c), the two Internet topologies have nearly identical distributions. Average node-degree-ratio, then, might be considered one of the key measures along which validate Internet topology generative models, since there is a clear standard—constant across two distinct samples of the Internet—against which to compare. The above analyses were conducted using *Oregon* and *Extended* snapshots of the Internet, both from April 21, 2001.

2) Dynamic measurements: To trace the behaviors of two Internet topologies, we downloaded nine snapshots of each AS topologies from [35]. These data are collected in each week starting from March 31 to May 26 2001. According to metric Sk, *Extended* topologies are more preferential than *Oregon*. Skewness of *Oregon* topologies are between 0.37 and 0.38 while that of *Extended* are between 0.30 and 0.32. Also, the extended topologies show smaller average diameters, but larger clustering coefficients than the original topologies. Figure 2 shows these results.

One of interesting observations is that the behaviors of two Internet topologies over nine weeks are quite similar even though their absolute metric magnitudes are different. To confirm this observation, we trace several other properties of two topologies, i.e. number of nodes and links, average degree, node birth/death, and link birth/death and observe that all these results strongly support our argument[1]. So, we conclude that dynamic characteristics drawn from *Oregon* are very valuable metrics to validate network generative models.

IV. Existing Internet topology generators and our models

In this section we describe seven existing generative Internet topology models, and two new model of our own. We categorize the models according to whether they are *static* models, meaning that they build the full network *en masse* without an explicit model of growth over time, or dynamic *growth* models, meaning that they incorporate an explicit procedure for the network's growth over time. In growth models, node connectivities are in general time-dependent—older nodes tend to have higher probabilities of gaining edges—whereas there is no explicit notion of time in static models.

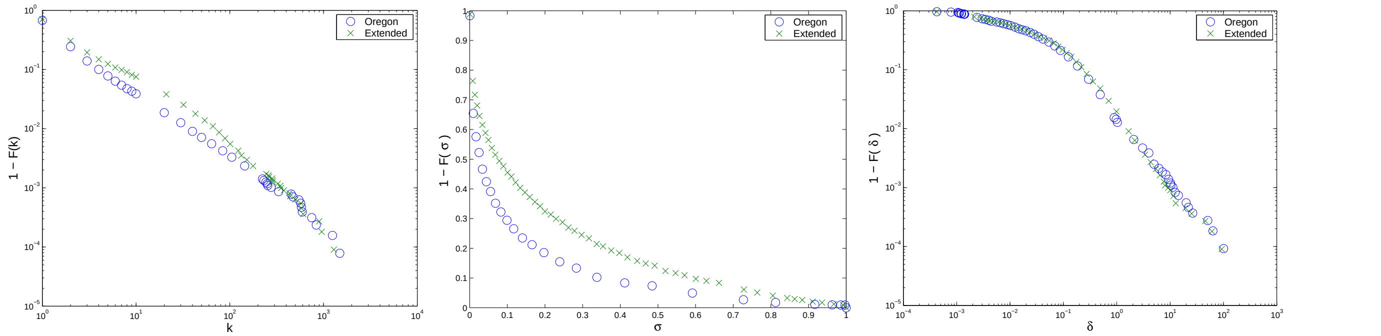
For growth models, there is a further distinction regarding the way in which links are added to (or removed from) the graph. Links can be added from a newly created node to the existing network; we call these *external* link additions. Or links can be added between already existing nodes in the network; we call these *internal* link additions.

Table I summarizes the characteristics of all nine models employed in our experiments. For all network models, we prohibit self links. Also, we prohibit network models from generating duplicate links rather than merging duplicate links at the end; we choose to prohibit duplicates because merging would reduce the number of links significantly. When a network model does not generate a fully connected graph, we only consider the largest connected component. (This process also potentially reduces the number of nodes and links significantly; however this method of canonicalization seems as appropriate as any). In this section we briefly explain each network model.

A. Static exponential (random) model

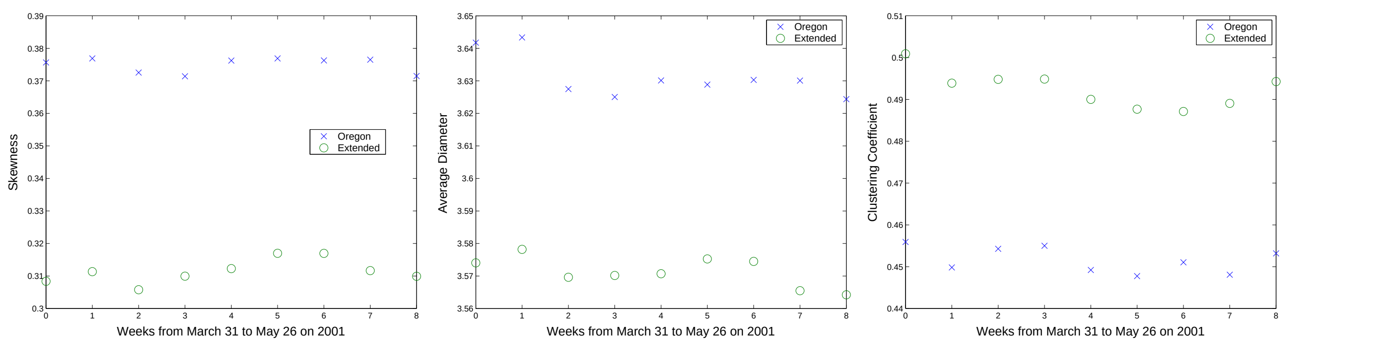
This model generates a random graph in the classic Erdos-Renyi sense. All nodes are added initially, then

[1] We do not show these results due to space limitation.



(a) Degree (b) Link-degree ratio (c) Average node degree ratio

Fig. 1. *Static metrics for* Oregon *and* Extended *topologies on April 21, 2001.* (a) Degree-frequency distribution: *Extended* shows a looser fit to a power law, while *Oregon* follows a nearly strict power law. (b) Link degree ratio: this metric clearly differentiates the two Internet topologies. (c) Average node degree ratio: this measure appears invariant under both the *Oregon* and *Extended* topologies.



(a) Skewness (b) Average diameter (c) Clustering coefficient

Fig. 2. Skewness, average diameter, and clustering coefficient; Our result clearly shows that the behaviors of two topologies are quite similar according to three different metrics.

TABLE I
COMPARING NINE GENERATIVE INTERNET TOPOLOGY MODELS.

	Static/Growth	Network Partition	Operations
Random	Static	Y	all nodes are added initially. internal link addition
GE	Growth	N	node birth with *m* links
BA	Growth	N	node birth with *m* links
AB	Growth	Y	node birth with *m* links, internal link addition, rewiring
GLP	Growth	N	node birth with *m* links, internal link addition
PG	Growth	Y	node birth without links, internal link addition
Inet-3.0	Static	N	all nodes are added initially, heuristics link addition
First model	Growth	N	node birth with m links, internal link addition
Second model	Growth	N	node birth with m links, internal link addition, dynamic generation of p

links are added one by one between pairs of (uniformly) randomly selected nodes. For every edge endpoint added, the probability that the edge endpoint attaches to a given node is

$$\Pi_{rand}(i) = \frac{1}{|V|}, \quad (6)$$

where V is the set of all nodes. Random graphs often partitions into several subgraphs; as mentioned we keep only the largest connected component. The model generates most nodes with roughly the same degree.

B. Growing exponential (GE) model

GE is a dynamic or growth-model version of the random graph model. At each time step, one node and m links are added. Links are connected *externally*, meaning that they all connect from the new node to one of the existing nodes. The identity of the existing node is chosen uniformly at random from among all nodes added to the graph in the past. The probability that a given edge endpoint attaches to a particular existing node is

$$\Pi_{ge}(i, t + \Delta t) = \frac{1}{|V(t)|}, \quad (7)$$

where $V(t)$ is the number of nodes in the graph at time t. Note that, although nodes are chosen uniformly at any given time step, as the network grows, older nodes tend to gain more links simply because they have more chances to.

C. Barabási-Albert (BA) model

The BA model [25] resembles GE except that destination nodes are chosen according to a linear *preferential attachment* function, rather than uniformly at random. Again, at each time step, one new node and m new links are added. Links are added externally from the new node to an existing node. The probability that existing node i is chosen is proportional to its degree:

$$\Pi_{ba}(k_i, t + \Delta t) = \frac{k_i(t)}{\sum_j k_j(t)}, \quad (8)$$

where $k_i(t)$ denotes the degree of node i at time t. The BA model is remarkable in its simplicity, and it seems to capture the minimal assumptions required to generate graphs with power-law degree distributions. However, in its basic form, it is not flexible enough to fit different power law exponents. The BA model, often cited as a more generic model (e.g., for the World Wide Web, the power grid, the co-star graph of Hollywood actors, etc.), touched off a wave of extensions and analysis among computer scientists and physicists.

D. Generalized linear performance (GLP) model

GLP [29] is one of the proposed extensions of BA. In this model, the probability of attachment is modified to better fit Internet-like graphs:

$$\Pi_{glp}(k_i, t + \Delta t) = \frac{k_i(t) - \beta}{\sum_j (k_j(t) - \beta)}, \quad (9)$$

where $-\infty < \beta < 1$. This model has two link addition operations:

1) with probability p, m links are added *internally*—links are added between two existing nodes. For each endpoint, a node is chosen with probability (9).
2) With probability $1 - p$, one new node and m new links are added externally from the new node to an existing node chosen according to (9).

In the simulation, we set parameters as $\beta = 0.7124$, $m = 1.13$, and $p = 0.4294$, which are the same as those in the [29]. The fractional m value of 1.13 means that 13% of new nodes are added with two links while 87% are added with one link, yielding an expected number of links/edge of 1.13.

E. Albert-Barabási (AB) model

The AB model [28] is the authors' own extension of their BA model. In this model, three operations are used as the network grows:

1) With probability p, m links are added internally. One edge endpoint is selected uniformly at random while the other endpoint is selected according to

$$\Pi_{ab}(k_i, t + \Delta t) = \frac{k_i(t) + 1}{\sum_j (k_j(t) + 1)}, \quad (10)$$

which is like (8) but with a "Laplacian smoothing"-like term.
2) With probability of q, m links are rewired. Node i is randomly selected and one of the links $l_{i,j}$ connecting i with j is randomly selected. Link $l_{i,j}$ is replaced with a new link $l_{i,k}$, where k is chosen according to (10).
3) With the probability $1-p-q$, one new node and m links are added externally from the new node to an existing node chosen according to (10).

The rewiring operation often causes the graph to become partitioned; we keep only the main connected component. In the experiments, we use parameters $m = 1$, $p = 0.45$, and $q = 0.1$.

F. "Pretty good" (PG) model

The PG model [36] is another extension of the BA model. This model adds a parameterized component of uniform attachment to the BA model's strictly preferential attachment policy. Specifically edge endpoints are chosen according to a mixture α of preferential attachment and $1-\alpha$ of uniform attachment:

$$\Pi(k_i, t+\Delta t) = \alpha \frac{k_i(t)}{\sum_j k_j(t)} + (1-\alpha)\frac{1}{|V(t)|} \tag{11}$$

This additional degree of freedom is enough to allow flexibility in fitting differing power-law exponents, and to fit typical divergences from the strict power law often observed in the low-degree region of a variety of naturally-occurring graphs, including communities on the World Wide Web. The PG model employs only *internal* link additions. That is, all edge endpoints are chosen according to (11), and new nodes are not explicitly differentiated.

Note that in the limit as $\alpha \to 0$, PG corresponds to GE, while as $\alpha \to 1$, PG corresponds to BA (modulo the internal/external distinction).

The main problem in adapting this model to our problem is that, because it employs only internal link additions, it generates too many disconnected nodes. For example, when $\alpha = 0.7$ and $m = 2$, around 50% of nodes are disconnected. Because we choose to keep only the largest connected component, the average degree within this component is artificially high. Alternative canonicalization policies might have yielded more comparable results for this model.

G. *Inet 3.0*

Inet-3.0 is the latest version of a complex yet very accurate model [37], [26]. The user provides the desired number of nodes N and the fraction k of nodes with degree one. The model proceeds in five steps. First, the model calculates the number of months (t) it would take the Internet to grow from its initial size in Nov. 1997 to size N according to:

$$N = exp(0.0298 * t + 7.9842). \tag{12}$$

Second, the model defines V_1, V_{top3}, and V', respectively, as the set of all degree-one nodes, the set of the three highest-degree nodes, and the set of all nodes except nodes in V_1 and V_{top3}. The model calculates the cumulative degree distribution (defined above in Section III-A.1) for all nodes in V' in order to match a power law:

$$1 - F(d) = e^c * d^{at+b}. \tag{13}$$

The degrees of particular nodes in V' are then assigned in order to agree with (13). The degrees of nodes in V_{top3} are assigned according to:

$$d = e^{pt+q} * r^R. \tag{14}$$

The parameters a, b, c, p, q, and R are known constants estimated from *Oregon* data, and t is the number of months since Nov. 1997.

Third, the model builds a spanning tree among all nodes in V_{top3} and V'. The spanning tree construction proceeds one node at a time, although any interpretation in terms of the network's natural evolution seems unwarranted, since the final degree values have already been pre-assigned in step two. In each step a node is selected randomly. One of the node's pre-assigned edges connects to the existing graph according to:

$$P(i,j) = \frac{w_i^j}{\sum_{k \in G} w_i^k} \tag{15}$$

where

$$w_i^j = \max\left(1, \sqrt{\left(log\frac{d_i}{d_j}\right)^2 + \left(log\frac{f\,(d_i)}{f\,(d_j)}\right)^2}\right) * d_j \tag{16}$$

This procedure continues until all nodes in V_{top3} and V' are added to the graph. Note that P(i,j) depends not only the degree of destination node j but also the degree of departure node i. If the degrees of two nodes are very different, the probability for two nodes to be connected is higher than the linear preference assumption. Otherwise, it roughly follows the linear preference assumption.

Fourth, the model connects all degree-one nodes (V_1) to the graph according to (15). Fifth, the model connects the remaining free edge endpoints (edges that have been assigned one endpoint in step two, but have not yet been assigned a particular second endpoint), starting from the highest degree nodes, according to (15).

We consider *Inet* to be a static model, since the probabilities of connections are time-independent; each node's degree is assigned in a batch process in step two. One interesting characteristic of this model is that number of links is *not* an input parameter; this value is computed to match the proper degree distribution using (in part) parameter t. This model is extremely accurate in generating random topologies similar in many respects to the *Oregon* data; in fact it fits this data much better than every other model we tested. However, the model seems particularly well-tuned to *Oregon*, and its flexibility in adapting to other data sets appears limited; for example, the model does not fit *Extended* data as well. Since the model is effectively static—it generates graphs with the explicit intention of matching particular aggregate characteristics like the degree distribution (13)—it is limited in its ability to provide any bottom-up explanation of why those particular aggregate characteristics arise. For our experiments, we did not re-implement *Inet*; we used the code made publicly available by the model's authors [35].

H. Our models

In this section, we describe our own generative network models. Our models are very simple and provide one possible explanation for the degree distribution displayed in the *Oregon* data set, and why other growth models disagree. Two assumptions help motivate our models: (1) For each link, we consider that the higher-degree node is a service provider and the lower-degree node is a customer; and (2) customers decide which providers they would like to connect to. Our models posit reasonable policies for customers to choose providers.

1) First model: Our first model can be thought of as yet another extended BA model, with a new attachment probability equation. Let node i be the customer node, which tries to generate a new link and k_i the degree of node i. Also, let $V(k_i+\gamma)$ be the set of nodes with degree higher than $k_i+\gamma$. Then, consumer i chooses provider j according to:

$$\Pi(k_j, t+\Delta t) = \begin{cases} \frac{k_j(t)}{\sum_{l\in V(k_i+\gamma)} k_l(t)} & if k_j > (k_i+\gamma) \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

In other words, a customer node always selects a provider node that has degree higher than $k_i+\gamma$; among this group the customer still prefers higher-degree nodes according to the linear preference function. This assumption seems reasonable: customers prefer to link up to providers whose connectivity is strictly greater than their own. The assumption is supported by our observations that most links on the Internet are hierarchical (endpoints have greatly varying degrees) rather than peer-to-peer (endpoints have similar degrees).

Our model has two operations: node birth and link birth. With probability p, a new internal link is added between existing nodes. The customer node is randomly selected and connected to a provider according to (17). With probability $1-p$, one new node and m external links are added. The new node is considered a customer and the m links are connected to providers using (17). For the experiments, we set $m = 1.25$, meaning that 25% of new nodes are added with two links and 75% are added with one link [31].

2) Second model: Since the average degree of the Internet changes continuously over time, Our second model adapts the probability p (the internal link addition probability) dynamically. We compute $P(N)$, the average ratio of internal link additions compared to all link addition, from the *Oregon* data using

$$\begin{aligned} In &= L - N*m \\ P(N) &= In/(N+In), \end{aligned} \quad (18)$$

where N is the number of nodes, L is the number of links, and In is the number of internal links added after November 1997. Then the probability p can be computed as follows:

$$\begin{aligned} p(N+\Delta N) &= p(N) + \frac{dP(N)}{dN} \\ &= p(N) - (3*10^{-9}*N) + \\ &\quad 3.6*10^{-5}, \end{aligned} \quad (19)$$

where $p(0) = 0.3$, determined empirically. So, the number of internal link additions versus external addition more closely reflects the trends seen on the Internet. This change to the model causes the average degree of nodes to increase over time, as the number of internal link additions grows. Figure 3 shows that the resulting trend in average degree growth for our model matches the trend found in the *Oregon* data quite closely.

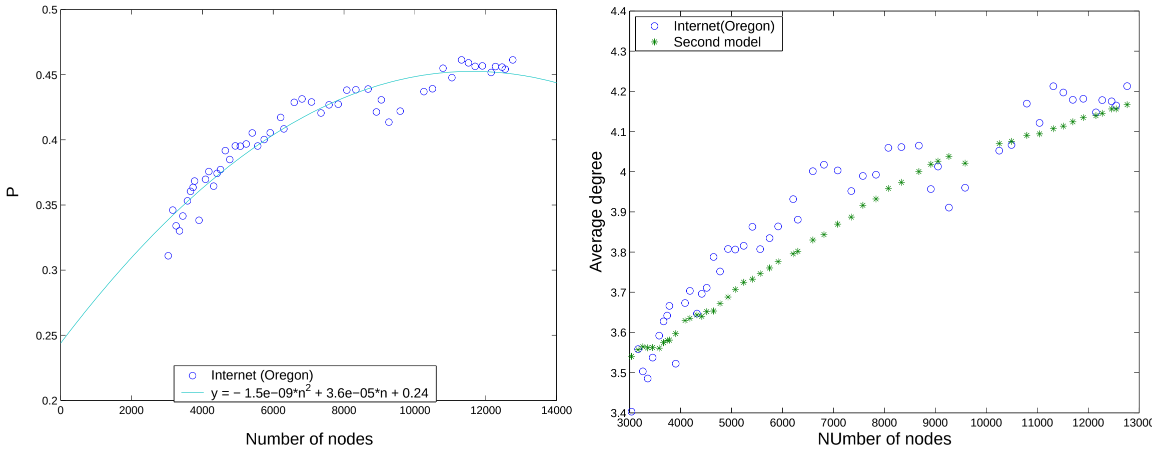
Note that γ determines how preferential a generated network is. In *BA* and its other extensions, $\gamma = -\infty$, meaning that all existing nodes have a certain probability to be chosen as a provider. However, in our models, customers choose providers only among candidate nodes which have higher degree than their own. We find that our models generate very similar Internet-topology-like graphs when $\gamma = 1$. All experiments show results for $\gamma = 1$.

V. Model comparison

In this section, we compare nine Internet models according to the three static and three dynamic metrics defined in Section III-A.

A. Static metric performance

We first compare the cumulative degree-frequency distribution for the nine models. Figure 4(a) shows a few of the models that do not perform particularly well according to this metric. Figure 4(b) shows that *AB*, *GLP*, *Inet*, and our two model do match the Internet (*Oregon*) data relatively well; with our models and Inet performing best. Note, however, that all models fit *Extended* considerable less well. Link-degree ratio clearly differentiates the models. Figures 4(c) and 4(d) split the models according to the same partition used in separating Figures 4(a) and 4(b). *Inet* matches *Oregon* the best, and our two model match *Oregon* very closely as well; *GLP* matches *Extended* best.



(a) Cumulative internal-link probability

(b) Average degree

Fig. 3. Average degree growth of our second model compared to the Internet (*Oregon*).

According to average node-degree-ratio, we find that our models, along with *AB*, *Inet* and *GLP*, show relatively good performance. Again, Figures 4(e) and 4(f) categorize models by their ability to fit the *Oregon* degree distribution.

Our models seem to exhibit excellent performance according to the static metrics. Our models show better agreement to the Internet than any other growth models across all three metrics. Only *Inet* show slightly better performance than our models. In general, we do not find any noticeable differences between the first and second model and conclude that the average degree increment over time does not affect the static metric performance of our model.

B. *Dynamic metric performances*

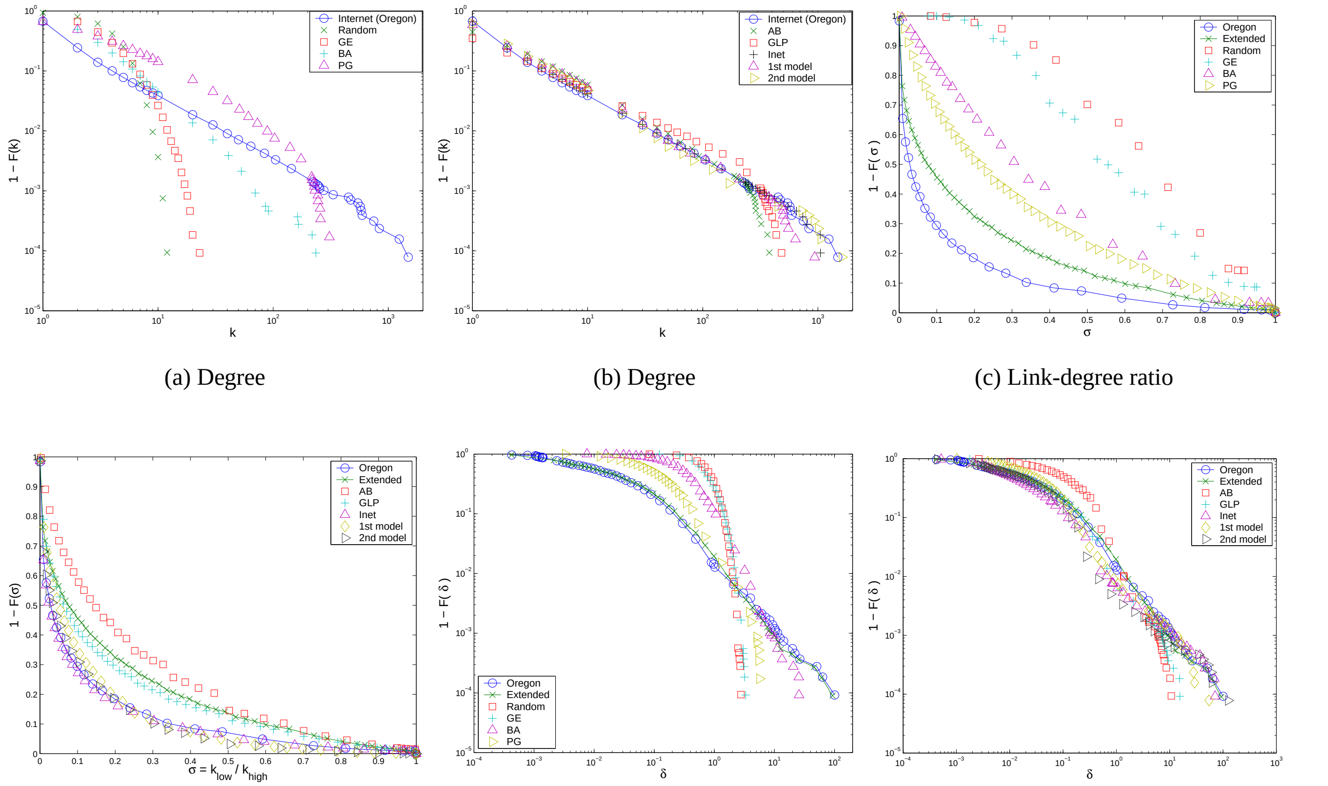
Next, we trace the behaviors of the models while the number of nodes in the networks increases. For the experiments, the Internet AS topologies from *Oregon* over a four year period from November 1997 to February 2002 were used. In each month, random graphs generated by network models include the same number of nodes with the Internet AS topologies. In Figure 5(a), only three network models (*GLP*, *PG*, and *Inet*) show continuous skewness decrement. With average diameter, only *AB*, *GLP*, and *Inet* shows decrement of the average diameter. With clustering coefficient, only *Inet* shows the continuous increment. Except *Inet*, all network models fail to follow the dynamic characteristics of the Internet: significant decrement of skewness and average diameter and significant increment of clustering coefficient.

With dynamic metrics, our models show small-world effects [33]; that is, their average diameters are very small but their clustering coefficients are much larger than those of classical random graphs. Note that absolute metric values of our model are very similar to those of the Internet. However, our models still do not match the clear relative trends in the data, and this cannot explain our observed dynamic characteristics of the Internet. When p is generated dynamically in our second model, the resulting networks display higher clustering coefficients but lower average diameters. However, its dynamic behavior is quite similar to our first model and dynamic p does not affect these trends. We can only conclude that the Internet's average degree change over time is not the main factor for determining network structure according to the metrics we examined.

According to our analysis, *Inet* is the best Internet topology generator in terms of matching the data, especially the *Oregon* data. However, *Inet* has several weaknesses. First, as it is effectively a static model rather than a growth model, it is limited in its ability to explain how the Internet grows. The model's complex heuristics designed to mimic *Oregon* data may in effect be overfitting or overtuning to that particular data source, making the model considerably less flexible in matching other data sources or in generalizing toward the future evolution of the Internet, even if that future topology is a relatively slight variant of what is seen today. Among growth models, our two new models appear to perform best, with *GLP* the best among the seven existing models tested.

VI. Limitation and Future Work

One major limitation of our models is that, like other growth models, they do not consider node/link deaths, for reasons of simplicity. However, Figure 6 shows that death events are another important factor that can greatly affect Internet topologies.



(a) Degree (b) Degree (c) Link-degree ratio

(d) Link-degree ratio (e) Average node degree ratio (f) Average node degree ratio

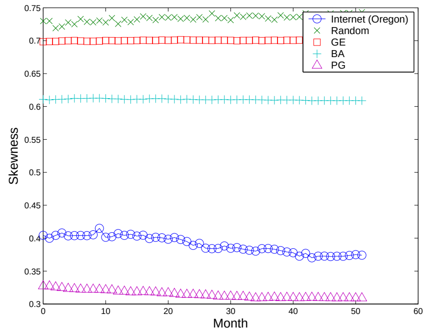
Fig. 4. *Static metric performance of nine Internet generating models.* (a,b) Degree-frequency distribution: *AB*,*GLP*, and *Inet* seem to be better model the Internet according to this metric; our models also show excellent performance with this metric. (c,d) Link degree ratio: *Inet* generates distributions similar to *Oregon*, but *GLP* generates distributions similar to *Extended*. This distribution clearly differentiates the models. Our models show better performance than other models except *Inet*. (e,f) Average node degree ratio: *AB*, *Inet* and *GLP* still show good performance with this metric. Our second model show slightly worse performance than the first model. In general, our two models generate good matching distributions over all three metrics.

One may argue that it is somewhat strange that our second model shows poor metric performance with average node-degree ratio even though it resembles real Internet topologies more. However, the current slow expansion of the Internet is due to the rapid increment of death events coupled with a slower increment of birth events. So, the actual internal-link probability p should be larger than our model [31]. These differences may affect the Internet's topology and be a source for the poor performance of our second model according to dynamic metrics.

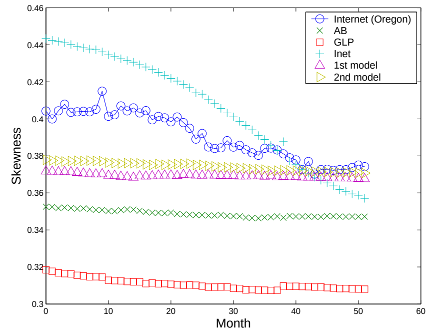
We also built a third model to explain Internet's dynamic characteristics. This model increases γ continuously according to the number of nodes to make a network more preferential while it grows. This model shows good dynamic metric performance, but does not work well with static metrics. We believe that death events in the Internet affect the growth pattern of the Internet significantly, and we need a closer analysis of death events to explain the dynamic characteristics of the Internet.

VII. Conclusion

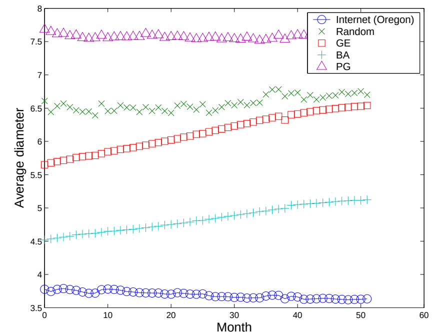
Recent studies have reported differing aggregate characteristics of the Internet's topology depending on the methodology used for sampling the Internet's true underlying structure. We examine two different data sets using six metrics (three of our own), showing that one static metric does a particularly good job at differentiating the data sets, one static metric appears invariant across the data sets, and all dynamic metrics exhibit a degree of invariance. We then compare nine generative models (two of our own). Among growth models, ours perform best, but all growth models (including our own) fail to capture the observed dynamic behavior of the Internet. A particular static model (Inet) does match the data well, but also is lacking in terms of an explanation for the Internet's growth pattern. We eagerly await any breakthroughs—perhaps incorporating a model of node/link deaths—that might yield plausible explanations for this striking behavior.



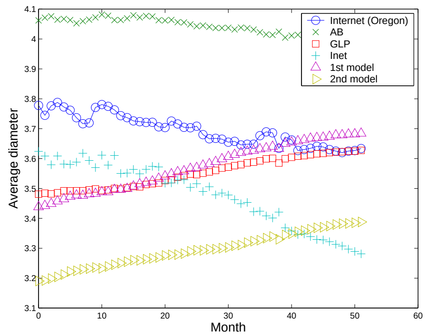
(a) Skewness



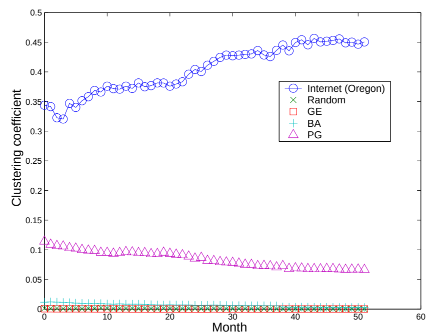
(b) Skewness



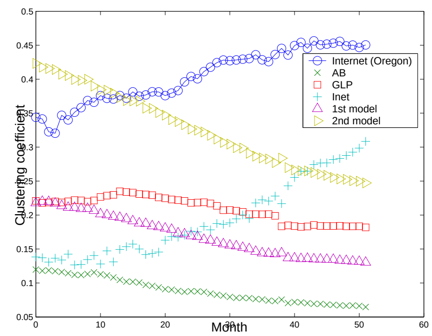
(c) Average diameter



(d) Average diameter

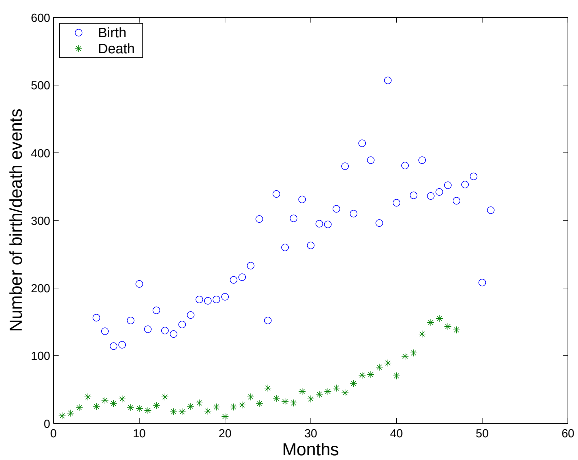


(e) Clustering coefficient

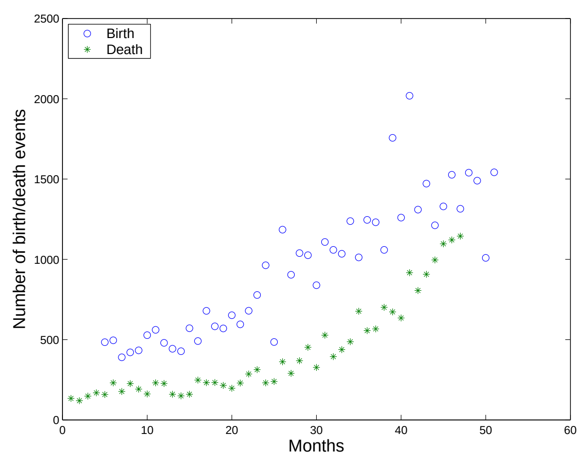


(f) Clustering coefficient

Fig. 5. Dynamic characteristics of models. (a,b) Skewness: only the random model shows a slight increment of skewness. *BA*, *AB*, and *GE* model do not show noticeable changes over time. *Inet* shows relatively similar behavior with that of the Internet, but its decrement rate is faster. Our models show good absolute values for this metric but show a slight *decrement* of skewness as they grow. (c,d) Diameter: only *Inet* and *AB* show a decrement in average diameter over time. Our first model shows similar average diameters to the those of *GLP*. (e,f) Clustering coefficient: only *Inet* shows an increment in the clustering coefficient over time. However, differences of clustering coefficient between the Internet and *Inet* are still high.



(a) Node



(b) Link

Fig. 6. Birth/death events. To avoid mis-categorizing temporary node/link failures as node/link deaths, we only consider a node/link dead if it does not appear at any time in the future. Also, we consider a node/link new only if it does not appear in any previous months. To keep the number of false dead nodes low, we do not calculate death events for the final five months. Similarly, birth events are not calculated for first five months. Our results suggest that the effect of link death cannot be ignored.

ACKNOWLEDGMENT

We gratefully acknowledge partial support from Ford Motor Co. We thank Steve Lawrence.

REFERENCES

[1] Oregon RouteViews, "http://moat.nlanr.net/routing/rawdata/," .

[2] H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger., "Towards Capturing Representative AS-Level Internet Topologiesd," Tech. Rep., Technical Report CSE-TR-454-02, EECS Department, University of Michigan, 2002.

[3] 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 INFOCOM*, 2002.

[4] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-law Relationships of the Internet Topology," in *SIGCOMM*, 1999, pp. 251–262.

[5] B. Lowekamp, D. R. O'Hallaron, and Thomas Gross, "Topology discovery for large Ethernet networks," in *SIGCOMM*, 2001.

[6] D. S. Alexander, M. Shaw, S. Nettles, and J. M. Smith, "Active Bridging," in *SIGCOMM*, 1997, pp. 101–111.

[7] M. Allman and V. Paxson, "On Estimating End-to-End Network Path Properties," in *SIGCOMM*, 1999, pp. 263–274.

[8] E. Cohen, B. Krishnamurthy, and J. Rexford, "Improving End-to-End Performance of the Web Using Server Volumes and Proxy Filters," in *SIGCOMM*, 1998, pp. 241–253.

[9] A. Veres, Z. Kenesi, S. Molnár, and G. Vattay, "The Propagation of Long-Range Dependence in the Internet," in *SIGCOMM*, 2000.

[10] K. Lai and M. Baker, "Measuring link bandwidths using a deterministic model of packet delay," in *SIGCOMM*, 2000, pp. 283–294.

[11] A. B. Downey, "Using Pathchar to Estimate Internet Link Characteristics," in *SIGCOMM*, 1999, pp. 222–223.

[12] A. Medina, I. Matta, and J. Byers, "On the Origin of Power Laws in Internet Topologies," *ACM Computer Communication Review*, vol. 30, no. 2, 18–28 2000.

[13] V. N. Padmanabhan and L. Qui, "The content and access dynamics of a busy web site: findings and implications," in *SIGCOMM*, 2000, pp. 111–123.

[14] C. Labovitz, A. Ahuja, R. Wattenhofer, and V. Srinivasan, "The Impact of Internet Policy and Topology on Delayed Routing convergence," in *INFOCOM*, 2001, pp. 537–546.

[15] C. R. Palmer and J. G. Steffan, "Generating network topologies that obey power laws," in *Proceedings of GLOBECOM '2000*, November 2000.

[16] L. Breslau and D. Estrin, "Design of inter-administrative domain routing protocols," in *Proceedings of ACM SIGCOMM*, 1990.

[17] D. Mitzel and S. Shenker, "Asymptotic resource consumption in multicast reservation styles," in *Proceedings of ACM SIGCOMM*, 1994.

[18] A. Feldman, A.C. Gilber, P. Huang, and W. Willinger, "Dynamics of ip traffic: A study of the role of variability and the impact of control," in *Proceedings of ACM SIGCOMM*, 1999.

[19] W.T. Zaumen, J.J Calvert, and M.J. Donahoo, "Dynamics of distributed shortest-path routing algorithm," in *Proceedings of ACM SIGCOMM*, 1991.

[20] B. M. Waxman, "Routing of Multipoint Connections," *IEEE Journal of Selected Areas in Communication*, vol. 6, no. 9, pp. 1617–1622, Dec. 1988.

[21] B. Bollobás, *Random Graphs*, Cambridge Mathematical Library. Cambridge University Press, 2001.

[22] K. L. Calvert, M. B. Doar, and E. W. Zegura, "Modeling internet topology," *IEEE Communications Magazine*, vol. 35, no. 6, pp. 160–163, June 1997.

[23] M. Doar, "A Better Model for Generating Test Networks," in *Globecom*, 1996.

[24] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger, "Network topology generators: Degree-based vs. structural," in *SIGCOMM*, 2002.

[25] A. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.

[26] C. Jin, Q. Chen, and S. Jamin, "Inet: Internet Topology Generator," Tech. Rep., CSE-TR-443-00, Department of EECS, University of Michigan., 2000.

[27] W. Aiello, F. Chung, and L. Lu, "A random graph model for massive graphs," in *Proceedings of the 32rd Annual ACM Symposium on Theroy of Computing*, 2000, pp. 171–180.

[28] R. Albert and A. Barabási, "Topology of evolving networks: local events and universality," *Physical Review Letters*, vol. 85, no. 24, pp. 5234–5237, December 2000.

[29] T. Bu and D. Towsley, "On Distinguishing between Internet Power Law Topology Generators," in *Proceedings of INFOCOM*, 2002.

[30] R. Albert, H. Jeong, and A. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, 2000.

[31] S-T. Park, A. Khrabrov, D. M. Pennock, S. Lawrence, C. L. Giles, and L. H. Ungar, "Static and dynamic analysis of the Internet's susceptibility to faults and attacks," in *Proceedings of INFOCOM*, 2003.

[32] R. Pastor-Satorras, A. Vazquez, and A. Vespignani, "Dynamical and correlation properties of the Internet," *Physics Review Letter*, vol. 87, 2001.

[33] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small world' networks," *Nature*, vol. 393, pp. 440–442, 1998.

[34] S.N. Dorogovtsev and J.F.F. Mendes, "Evolution of networks," arXiv:cond-mat/0106144, 2001, submitted to Adv. Phys.

[35] Topology Project, "http://topology.eecs.umich.edu/data.html," .

[36] D. M. Pennock, G. W. Flake, S. Lawrence, E. J. Glover, and C. L. Giles, "Winners don't take all: Characterizing the competition for links on the web.," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 99, no. 8, pp. 5207–5211, 2002.

[37] J. Winick and S. Jamin, "Inet-3.0: Internet topology generator," Tech. Rep., CSE-TR-456-02, Department of EECS, University of Michigan.