On Distinguishing between Internet Power Law Topology Generators

Tian Bu and Don Towsley
Department of Computer Science
University of Massachusetts
Amherst, MA 01003
{*tbu,towsley*}@*cs.umass.edu*

Abstract— **Recent work has shown that the node degree in the WWW induced graph and the AS-level Internet topology exhibit power laws. Since then several algorithms have been proposed to generate such power law graphs. In this paper we evaluate the effectiveness of these generators to generate representative AS-level topologies. Our conclusions are mixed. Although they (mostly) do a reasonable job at capturing the power law exponent, they do less well in capturing the clustering phenomena exhibited by the Internet topology. Based on these results we propose a variation of the recent incremental topology generator of [6] that is more successful at matching the power law exponent and the clustering behavior of the Internet. Last, we comment on the small world behavior of the Internet topology.**

I. Introduction

Recent work has shown that the node degree in the World Wide Web (WWW) induced graph, and the AS-level topology of the Internet exhibit power laws [3] [7] [14]. This work has stimulated considerable activity aimed at understanding the implications on web and network design [4] [8] [18] [19], and on the development of algorithms for generating graphs exhibiting such power laws [5] [6] [10] [20]. Past research has shown that the network topology can have a significant effect on the performance of network protocols. For example, the AS-level topology of the Internet has a major impact on the convergence of the inter-domain routing protocol BGP [16]. The growth in the AS-level topology is placing significant pressure on the scalability of BGP [15]. Thus there is a tremendous need for research on inter-domain routing. This research will benefit from topology generators that can produce Internet-like topologies.

Many topology generators have been proposed for modeling Internet. They fall into one of three classes, random graph generators, structural generators, and degree power law generators. It has been demonstrated in [17] [21] that the topology produced by power law topology generators resembles the AS-level Internet topology better than those produced by random graph generators or structural generators. In this paper, we investigate the effectiveness of several recently developed power law topology generators, [5] [6] [10] [20] for generating representative Internet topologies. All of these algorithms generate topologies with node degree power laws. However, not all produce the correct power law exponent. Furthermore, even when they do, they differ from each other in more subtle ways. To distinguish among these various algorithms, we revisit the small world work of [22] and adopt two metrics, the characteristic path length and the cluster coefficient. We observe that all of the algorithms have difficulty generating topologies whose values for these metrics match that of the AS-level Internet topology. With this in mind we modify the incremental algorithm first proposed in [10] and refined in [6] by extending the "linear preference model" to include a translation. We observe from simulations that this algorithm does a better job matching these additional metrics than previous generators.

Our paper makes several contributions. First, it proposes the use of clustering coefficient and characteristic path length (both introduced in [22]) to distinguish power law topology generators from one another. Second, it presents a generalized linear preference model that, coupled with the incremental algorithm of [6], generates topologies that more closely model the Internet. Third, in proposing these metrics, we also make the not surprising observation that the Internet exhibits the small world properties introduced in [22]. Fourth, in studying the node degree power law, we point out the advantage of working with the empirical complementary distribution rather than the node degree histogram as was done in [14]. Last, except for the case of the Inet topology generator [20], which sets the maximum node degree in a deterministic manner, all of the generators (including ours) show great variability in the graphs that are generated. This is not surprising as the nature of a power law distributions are characterized by infinite variance.

In addition to the works empirically establishing the node degree power law, [14], and subsequent works concerned with generating power laws, [5] [6] [10] [20], several other works deserve mention. There is the important work of Zegura et al. [24] that presented the first Internet specific topology generator. Although this generator may be appropriate for router-level topologies, the subsequent work of [14] showed that it does not match the AS-level Internet topology characteristics. Another interesting work by Chen, et al., [13] examines the effectiveness of the incremental model of [6] to model the evolution of the AS-level topology of the Internet. Their analysis of BGP routing table data corroborates the premises underlying the model in [6], namely that new nodes are connected preferentially to existing nodes with a large degree, that links are incrementally added and that rewiring occurs. However, they conclude that the specifics of the model in [6] differ from the evolution of the Internet, namely that rewiring occurs infrequently and that new nodes express a greater preference for nodes with large degree

This work was supported in part by DARPA under contract F30602-00-2-0554, by National Science Foundation under Grant No. CDA-9502639, EIA0080119.

than is represented by the simple linear preference model introduced in [6]. Our new generator better matches these observations. It requires no rewiring and it gives greater preference to large degree nodes.

The remainder of the paper is organized as follows. Section II provides details about the AS-level topology that is the focus of the paper, reviews the node degree power law, demonstrates that the Internet is a small world graph. In Section III, we review existing power law topology generators and compare the topologies that they produce to the Internet topology using the clustering coefficient and characteristic path length as metrics as well as the power law metric. We propose, analyze, and evaluate a new power law topology generator in Section IV. Section V concludes the paper with some discussions.

II. Internet topology, power laws and small world phenomena

In this section, we first introduce the Internet AS-level topology that is the focus of our study. We then review the power law describing the node degree, [14], and introduce the clustering coefficient and characteristic path length. The latter two satisfy the small world conditions described in [23].

A. Internet topology

The Internet consists of a large collection of hosts interconnected by networks of links and routers. The Internet is divided into thousands of administrative domains, each of which possesses one or several autonomous systems (ASs). The Internet can be considered as either a graph of interconnected routers or a graph of interconnected ASs. In this paper, we study the AS-level Internet topology where nodes represent ASs and links represent the relationship of exchanging traffic between ASs. The real AS-level Internet topology is unknown. However, since the Border Gateway Protocol (BGP), an inter-AS routing protocol, is a path-vector protocol, the AS-level Internet topologies can be inferred from BGP routing tables. The routing tables we used are archived by the oregon route view server and the RIPE server.

The goal of the oregon route view server [1], `route-views.oregon-ix.net`, is to obtain real-time information about the global routing system from the perspectives of several different backbones and locations around the Internet. By the end of year 2001, the oregon route view server collects routing information from up to 57 ASs. The connections to the operational routers are done in such a way that these routers advertise all their routes current being used. From November 1997, the routing tables of the oregon route view server have been achieved.

The RIPE organization runs a Routing Information Service (RIS) project [2] whose goal is also to collect routing information between ASs. RIS uses the same collecting strategy as the oregon route view server, but it peers with different ISPs. Moreover, RIS has introduced different collection points that peer with different ASs. However, the RRC00 is the major collection site and the first to archive routing tables. The RRC00 connected to 15 ASs as of the end of year 2001 and its has archived routing tables since September 1999.

Since each of the two routing information collectors peers with many ISPs at different locations, the Internet topologies derived from the union of their routing tables should be sufficiently close to the topologies of the real Internet. However, some ASs and connections may still not be captured by these topologies due to various reasons. For instance, a customer AS does not necessarily announce all its connections to its provider. Only part of its connections can be discovered from the routing table of its provider. It is still an active research topic to obtain complete AS-level Internet topology [11] [12]. It has been proposed in [12] to augment connections among ASs using the Internet Routing Registry (IRR). Up to 20% more connections are discovered from the IRR database. However, since the records in the database are published by individual ISPs voluntarily and updated less frequently than the BGP routing tables, the connections discovered from IRR database may be out of date. Hence we do not include them.

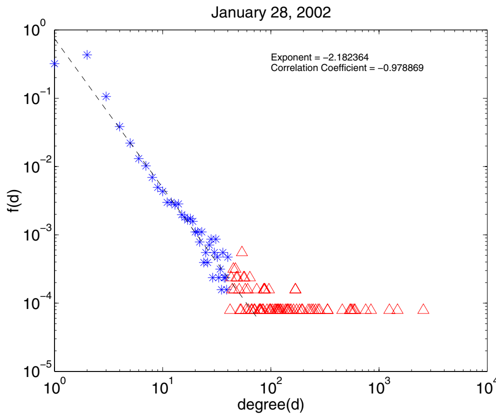
We represent the AS level topology of the Internet by a graph $G = < V, E >$. $\forall v \in V, v$ denotes an AS and $\forall e \in E$, e is an inter-AS inter-connection which we assume is an unordered pair of elements in V. Last, we will sometimes use $E(G)$ to refer to the set of edges in G.

B. Power laws

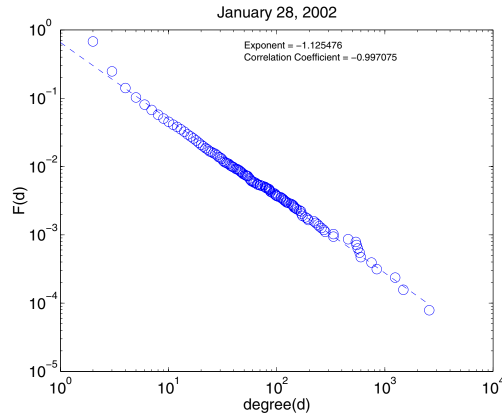
Recently Faloutsos et al. [14] showed empirically that certain properties of the AS-level Internet topology are well described by power laws. The most interesting of these regards the degree of a node. If we let $f(d)$ be the fraction of nodes with degree d, then [14] demonstrated that $f(d) \propto d^{\eta}$. The exponent η is obtained by performing a linear regression on $f(d)$ when plotted on a log log graph. Figure 1(a) shows such a plot for the January 2002 AS-level topology where $\eta \approx -2.18$. In order to claim the existence of a power law and to compute η, it is necessary to remove a small number of outliers, nodes with very high degree. Interestingly, the edges incident to these nodes correspond to approximately $1/3$ of the total number of edges in the AS-level Internet topology. Specifically, [14] plots the degrees starting from one until encountering a degree that has frequency of one. For instance, in Figure 1(a), the power law is obtained from a linear regression over all of the star ($*$) points; all of the triangle ($\triangle$) points are removed.

In this paper, we focus on the empirical complementary distribution (ecd) $F(d) = \sum_{i=d}^{\infty} f(d)$, i.e., the fraction of nodes with degree greater than or equal to d. Plotting it on a log log graph suggests that it is described by a power law $F(d) \propto d^{\alpha}$. Note that this representation allows us to utilize all of the data unlike the case for $f(d)$. For the case of the January 2002 AS-level topology, a linear regression produces $\alpha = -1.13$. Figure 1(b) illustrates the ecd for this topology. We note that, visually, the ecd more closely resembles a power law than the degree histogram. Moreover, the degree histogram can correspond to an infinite number of ecds (those in which the outliers are placed in other locations), most of which are not characterized by a power law.

It is not clear that the degree histogram, with the tail removed, exhibits heavy tail behaviors. On the other hand, the ecd plot demonstrate the heavy tails.



(a) degree histogram power law



(b) empirical complementary distr. power law

Fig. 1. The degree histogram power law vs. empirical complementary distribution power law.

We observe from Figure 1 that $\eta = \alpha - 1$ approximately. This can be readily explained. If we approximate a discrete variable d with its continuous version, we then have $f(d) = \frac{\partial F(d)}{\partial d}$. Assuming $F(d) = ad^{\alpha}$, we then have $f(d) = a\alpha d^{\alpha-1}$.

Last, note that the number of nodes with degree two is significantly higher than predicted by the power law in Figure 1. This is because many ASs tend to connect to at least two providers to assure increased reliability.

C. Small world graph

In [22], the notion of a small world graph was introduced. Intuitively, a graph is a small world graph if, for a given median shortest path length, it has a much larger clustering coefficient than a random graph with the same median shortest path length. In this section we will observe that the AS-level Internet topology exhibits small world characteristics. In doing so, we will introduce the characteristic path length and clustering coefficient of a graph, metrics that will be used in the next section to distinguish between power law graphs produced by different topology generators.

Let us first define, $\forall v \in V$, the degree of vertex v, k_v to be the number of edges incident to the vertex. $\forall u, v \in V$, the distance between u and v, $d(u,v)$, is the number of edges of the shortest path between these two vertices. Let $d(u)$ denote the average shortest path length over all possible destinations, i.e., $d(u) = (|V|-1)^{-1} \sum_{v \in V \setminus \{u\}} d(u,v)$.

We have following definitions from [23]:

Definition 1: A **characteristic path length**, L of a graph is the median of the means of the shortest path lengths connecting each vertex $v \in V$ to all other vertices. That is, L is the median of $\{d(u)\}$

The neighborhood $\Gamma(v)$ of a vertex v is the subgraph that consists of the vertices adjacent to v (not including v itself). The clustering coefficient $\gamma(v)$ of $\Gamma(v)$ characterizes the extent to which vertices adjacent to any vertex v are adjacent to each other. More precisely,

$$\gamma(v) = |E(\Gamma(v))| / \binom{k_v}{2}$$

where $|E(\Gamma(v))|$ is the number of edges in the neighborhood of v and $\binom{k_v}{2}$ is the total number of possible edges in $\Gamma(v)$.

Definition 2: The **clustering coefficient** of G is $\gamma = |V|^{-1} \sum_{v \in V} \gamma(v)$ provided that $k_v \geq 2, \forall v \in V$.

Let $V^{(1)} \subset V$ denote the set of vertices of degree one. $\forall v \in V^{(1)}$, note that $\gamma(v)$ is not defined when the degree of v is one. We will use a slightly different definition of clustering coefficient to account for this.

Definition 3: $\hat{\gamma} = (|V| - |V^{(1)}|)^{-1} \sum_{v \in V \setminus V^{(1)}} \gamma(v)$.

For the purpose of charactering whether a graph exhibits small world phenomena, it is necessary to define a random graph. A random graph of n vertices and characteristic path length L is constructed as follows. Begin with n vertices and no edges. At each step, two randomly chosen vertices are connected with an edge. Repeat this until the characteristic path length of the random graph is L.

Date	No. of ASs	No. of Edges	L	$\hat{\gamma}$	L_{rand}	$\hat{\gamma}_{rand}$
Sept. 99	5764	11173	3.71	0.3886	3.7914	0.0023
Mar. 00	7012	14985	3.6367	0.4417	3.7035	0.0019
Sept. 00	8613	18346	3.6168	0.4531	3.5584	0.002
Mar. 01	10424	22488	3.6193	0.4621	3.5724	0.0016
Sept. 01	11867	25632	3.6205	0.4673	3.6144	0.0015
Jan. 02	12709	27384	3.6179	0.4597	3.638	0.0014

TABLE I

INTERNET IS A SMALL WORLD GRAPH

In order to check whether the Internet is a small world graph, we first measured the characteristic path length and clustering coefficient of a total of six Internet topologies. The earliest was derived from the routing tables collected on September 18, 1999 and the latest was derived from the routing tables collected on January 28, 2002 [1]. Table I shows the graph properties measured from these topologies. It is observed from Table I that the number of ASs in the Internet grew by more than 200%. On the other hand, neither the clustering coefficient nor the characteristic path length changed much. In fact, the characteristic path

[1] We infer the Internet topology by combining routing tables from the oregon route view server and the RIS server. The RIS server started to archive routing tables in September, 1999.

lengths decreased by 2% whereas the clustering coefficients increased by 18%.

We now show that the AS-level Internet topology is a small world graph. For each topology in Table I, we first construct 10 random graphs as defined earlier such that the random graphs include the same number of nodes and exhibit approximately the same characteristic path length as the Internet topology [2]. We then compute the clustering coefficient for all 10 random graphs. The median of both the characteristic path length and the clustering coefficient of the random graphs are recorded in Table I as L_{rand} and $\hat{\gamma}_{rand}$ respectively. We observe that the random graphs exhibit significantly smaller clustering coefficients than the Internet topologies even when the characteristic path lengths are close. Therefore, the Internet topology is a small world graph. In later sections, we will observe that the characteristic path length and clustering coefficient can be used to distinguish different topology generators.

III. EXISTING POWER LAW TOPOLOGY GENERATORS

In this section, we first briefly review several existing generators that produce topologies characterized by degree power laws. We will see that, although all generators are tuned to match the same Internet topology, these generated topologies differ widely from each other and the Internet with respect to the clustering coefficient and characteristic path length.

A. Description of generators

The discovery of power laws in the topology of the Internet and WWW has stimulated research on generating such topologies [5], [10], [6], [20]. Several such generators have appeared in the literature.

- PLRG (power law random graph) is a generator based on curve fitting proposed in [5]. Given a target number of nodes N, and a power law exponent η, PLRG first assigns degrees to N nodes drawn from a power law distribution with exponent η. It then randomly matches degrees among all nodes. Note that the graphs produced by PLRG may not be connected and may contain self loops and duplicate links. It is proved in [5] there is always a giant connected component for a large range of η. We choose the giant connected component, delete all self loops and merge duplicate links. We observe from our simulations we have done so far that we can always find a connected component containing $80\% \sim 90\%$ of nodes in the original topology.
- A model for generating power law graph was proposed in [10]. Starting with a small network core, it incrementally constructs a topology. At each step one of two operations is probabilistically chosen, *(i)* adding a new node along with m links or *(ii)* adding m new links without a node. In both cases, the links are connected to existing nodes with a probability that is proportional to the degree of the nodes, i.e., links are connected to existing nodes according to a linear preferential connectivity rule. We refer to this as the BA generator. It has been shown in [9] that the $f(d)$ power law exponents of topologies produced from BA generator is approximately $\eta = -3$ which is much less than the $f(d)$ power law exponents measured from the Internet (e.g., -2.18 in Figure 1).
- The model proposed in [10] was extended in [6] by adding a third rewiring operation consisting of choosing m links randomly and re-wiring each end of them according to the same linear preference rule as used in BA generator. We refer to this as the AB generator. Note that the rewiring operation may disconnect the graph and the generated graph may contain self loops and duplicated links. We choose the largest connected component, delete the self loops and merge the duplicated links. We observe from simulations we have done so far that we can always obtain a connected component which contains $75\% \sim 85\%$ of nodes in the original topology.
- The Inet (Internet Topology Generator) [20] makes use of both curve fitting and preferential attachment. It first calculates the frequency-degree and rank-degree distributions. It then assigns degrees to each node according to these distributions. Finally it matches these degrees according to the linear preferential model.

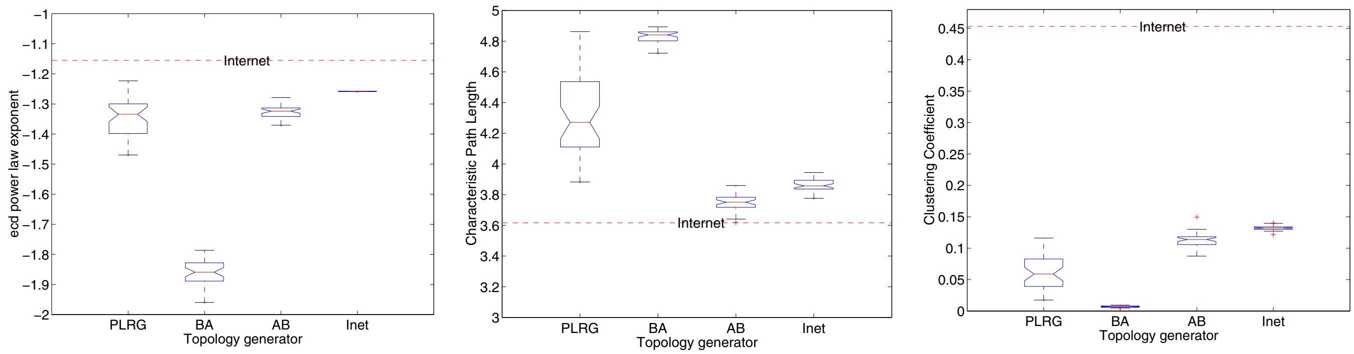
B. Comparison with the Internet

We begin by comparing the topologies produced by these generators to the Internet. We first choose an Internet topology observed in September, 2000 as the reference topology and measure the number of nodes $N_V = |V| = 8613$, the number of edges $N_E = |E| = 18346$, $F(d)$ power law exponent $\alpha = -1.155$, $f(d)$ power law exponent $\eta = -2.1824$, characteristic path length $L = 3.6168$ and clustering coefficient $\hat{\gamma} = 0.4531$. We then tune each generator to produce topologies that match the chosen Internet topology. Note that not all of these generators take the same inputs. Specifically, N_V and η are inputs to PLRG, N_V and N_E are inputs to BA, N_V, N_E and α are inputs to AB, and N_V is the input to Inet. By varying the random seeds while using the same inputs, we then produce 100 topologies using each generator and measure the power law exponents α, characteristic path lengths and clustering coefficients for all of generated topologies. All these measured graph properties are presented in Figure 2. Each column in the figure is a *boxplot*[3] of α, or L, or $\hat{\gamma}$ of all topologies generated by the same generator with the same inputs but different random seeds. The measured values from the AS-level Internet topology are shown as dashed lines.

We observe from Figure 2(a) that the median of the power law exponents α obtained from the topologies produced by BA generator is significantly less than that measured from the Internet. This is consistent with its derived approximation for the power law exponent, $\eta \approx -3.0$ in [9]. The median of α obtained from Inet is lower than the α measured from the Internet. Inet uses the Internet in November 1997 as a reference for constructing topologies. The power law exponent has changed slightly since then. Hence there is a slight difference in α as we compare the generated topologies with the Internet topology in September

[2]The difference of characteristic path length between the Internet topology and the random graph is no more than 5%

[3]The *boxplot* produces a box and whisker plot for each variable(e.g., power law exponent, characteristic path lengths or cluster coefficient). The box has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extend of the rest of the data. Outliers are data with values beyond the ends of the whiskers.



(a) ecd power law exponents (b) Characteristic Path Length (c) Clustering Coefficient

Fig. 2. Comparison of various power law topology generators

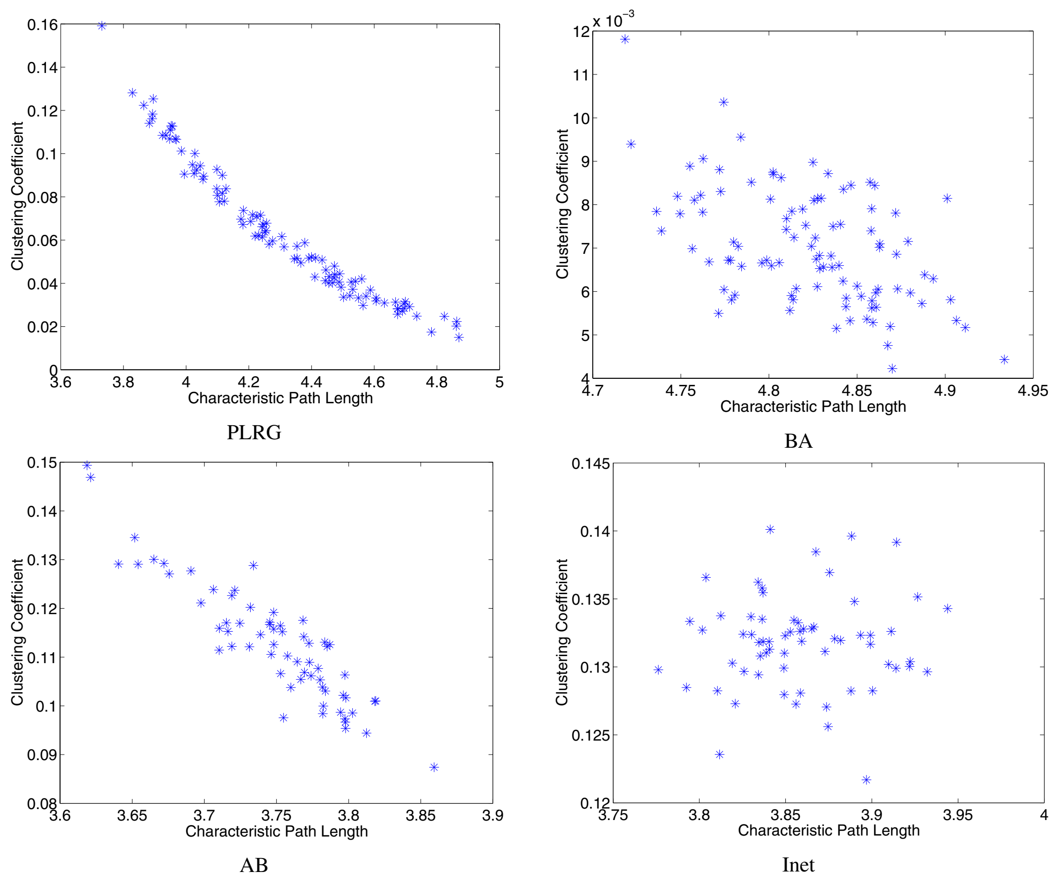


Fig. 3. The correlation between clustering coefficient and characteristic path length for all generators.

2000. We also observe that the median of the power law exponents α obtained from topologies produced by PLRG and AB are slightly lower than that for the Internet. We believe that this is due to the fact that we delete self loops and merge duplicate edges. Last, we observe that there is a greater variability in the power law exponent α obtained from PLRG than from AB and Inet.

Figure 2(b) suggests that the median of characteristic path lengths, L obtained from the topologies produced by Inet and AB are only slightly higher than the characteristic path length measured from the Internet. The median of the characteristic path lengths obtained from the topologies produced by BA generator are much larger than the characteristic path lengths measured from the Internet. The characteristic path lengths of the topologies produced by PLRG exhibit the greatest variability.

According to Figure 2(c), the clustering coefficients of topologies produced by all four generators are significantly less than the clustering coefficient of the Internet. In addition, the clustering coefficients of the topologies produced by BA model are so small that we can not easily claim these topologies are small world graphs. In addtion, the clustering coefficients of topologies generated by PLRG shows very large variability.

We examine the correlation between clustering coefficient and characteristic path length for different topology generators in Figure 3. It can be observed that the clustering coefficient and the characteristic path length show strong negative correlation in topologies produced by PLRG and AB but weak negative correlation in topologies produced by BA. We observe almost no correlation for these topologies produced by Inet.

We see from Figure 2 that the clustering coefficient and the characteristic path length can distinguish different topology generators very well. In addition, the results suggest that none of the topology generators produces topologies close to that of the Internet if the clustering coefficient is taken as the criterion.

We have observed large variability in the clustering coefficients and characteristic path lengths of topologies produced by PLRG, even though they were produced using the same inputs. We now show that both the clustering coefficient and characteristic path length are sensitive to the sum of the degrees of a small number of nodes with the largest degrees. For each topology, we first compute the average over the largest k degrees. We then show the *boxplot* of these averages obtained from 100 topologies for each topology generator in Figure 4. The measured values from the AS-level Internet topology are shown as dashed lines. We vary k from 1 to 7 for each topology. We observe that BA produces topologies with much smaller averages over the largest k degrees than the other generators. The average over the largest k degrees obtained from the topologies generated by PLRG and AB exhibit very high variability. The topologies produced by Inet have a maximum degree closer to that of the Internet than the topologies produced by any other generators. We next plot the maximum degree versus the clustering coefficient in Figure 5 and the maximum degree versus the characteristic path length in Figure 6. We observe from Figures 5 and 6 that the clustering coefficient increases whereas the characteristic path length decreases as the maximum degree increases for topologies produced by AB, PLRG and BA. Such correlations are not obvious for topologies produced by Inet. If one wants to generate topologies with constant clustering coefficients and characteristic path lengths, a generator such as Inet, which exhibits little variability in the average over the largest degrees is preferred.

Inet, BA and PLRG are all based on a linear preference connectivity model. It is reported in [13] that the evolution of the Internet AS-level topology is not well modeled by linear preference. A new node is more likely to connect to high degree nodes than can be explained by linear preference. Although the AB generator also uses the linear preference rule, it can assign larger preference to high degree nodes than other generators. This is done by introducing a rewiring operation where, with certain probability an edge is disconnected from low degree node and re-connect to high degree node. We observed that a re-wiring probability as high as 50% is needed for the purpose of matching the Internet power law exponent. However, according to the measurement reported in [13], rewiring may not be a significant factor in the evolution of the Internet AS topology. Note that except for the BA and Inet generators, none of the other generators can guarantee a connected graph.

In addition to the September 2000 Internet topology, we have also used AS-level Internet topologies collected at other dates as reference Internet topologies and have performed the same experiments as described in this section with similar results.

In the next section, we propose a new topology generator that uses a generalized linear preference model. The new generator allows us to produce topologies with clustering coefficients closer to that of the Internet.

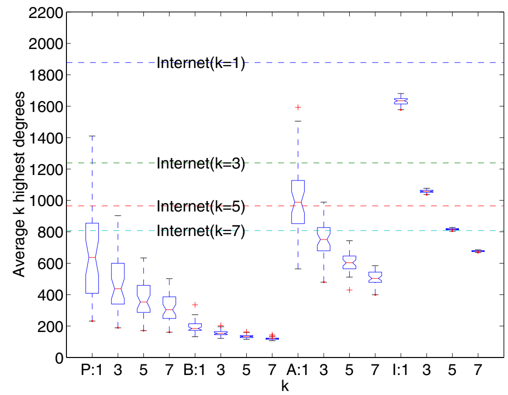


Fig. 4. The average k largest degrees for different topology generators. P:PLRG, B:BA, A:AB, I:Inet.

IV. A NEW TOPOLOGY GENERATOR

In this section, we modify the model proposed in [10] to allow more flexibility in specifying how nodes choose to connect to other nodes. The modification also allows us to delete the rewiring operation required by the AB model [6]. This is reasonable as the data presented in [13] suggests that rewiring rarely happens in the Internet.

Let $\Pi(d_i)$ denote the probability for choosing node i. In the linear preference model, $\Pi(d_i)$ is defined as,
$\Pi(d_i) = d_i / \sum_j d_j$ where d_i is the degree of node i. It has been reported that in the real Internet, new ASs have a much stronger preference to connect to high degree ASs than predicted by the linear preferential model [13]. A straightforward extension of the linear preferential model is to define $\Pi(d_i)$ as $\Pi(d_i) = d_i^a / \sum_j (d_j^a)$. Here, greater preference is given to high degree nodes when $a > 1$. However, it is difficult to analyze this model in order to determine whether it produces a power law distribution $F(d)$. On the other hand, we adapt a generalized linear preference model where $\Pi(d_i)$ is defined as

$$\Pi(d_i) = (d_i - \beta) / \sum_j (d_j - \beta) \quad (1)$$

where $\beta \in (-\infty, 1)$ is a tunable parameter that indicates the preference for a new node (edge) connecting to more popular nodes. The smaller the value of β is, the less preference gives to high degree nodes. Note that we choose $\beta < 1$ so that there is a nonzero probability that nodes of degree one acquire new links. We can mathematically show that this generalized linear preference model produces a power law distribution $F(d)$.

At the inter-domain level, the growth of the Internet is due mostly to two operations, the addition of new domains and the addition of new interconnections between existing domains. In order to have better connectivity to the Internet, the newly added

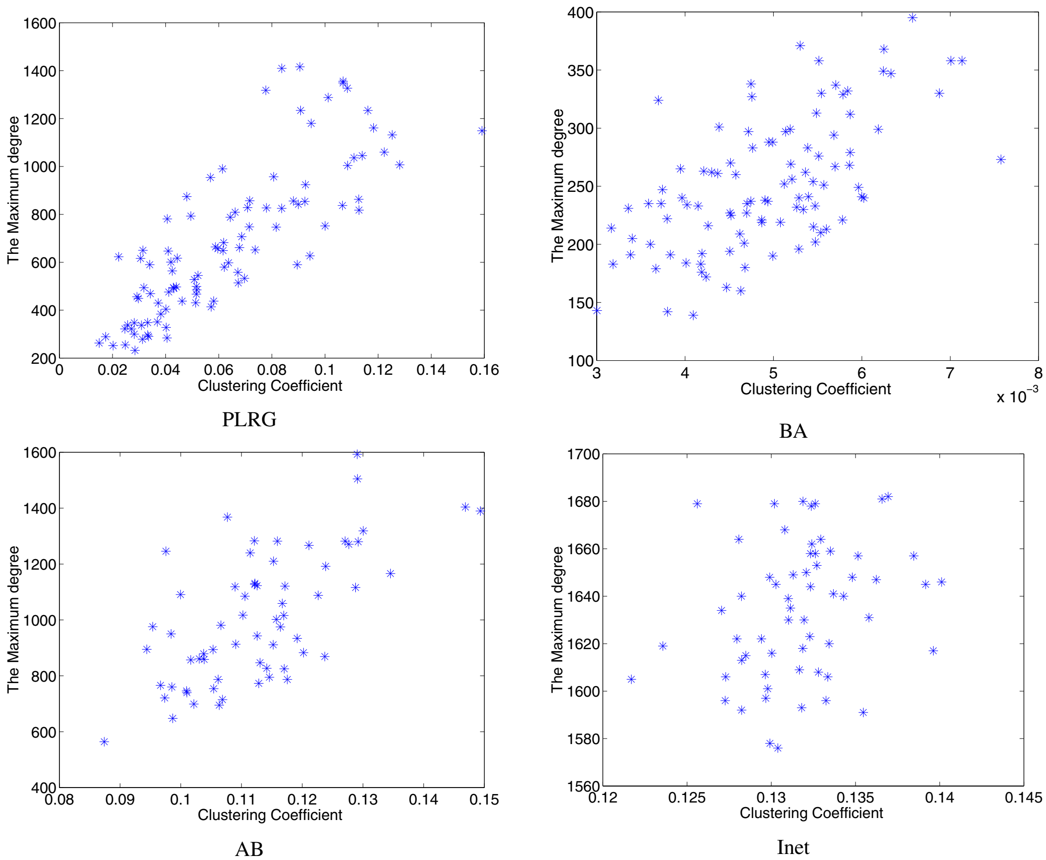


Fig. 5. The correlation between clustering coefficient and the maximum degree for all generators.

domains prefer to connect to a popular domain. Due to the fact that popular domains tend to peer with each other, the newly added edges are more likely to connect to popular domains.

A. Model

Our model captures two events corresponding to the addition of a new node and the addition of a link. We start with m_0 nodes connected through $m_0 - 1$ edges, and at each time-step we perform one of the following two operations:

- With probability p we add $m \leq m_0$ new links. For each end of each link, node i is chosen with probability $\Pi(d_i)$ as defined in formula (1). This incorporates the fact that new links preferentially connect popular nodes.
- With probability $1-p$ we add a new node. The new node has m new links. Each link is connected to node i already present in the system with probability $\Pi(d_i)$ as defined in formula (1).

In the model, $p \in [0,1]$. We show next that this model can produce a graph with a degree power law. In addition, we demonstrate how to choose parameters so as to produce a desired power law exponent. We'll refer to the new model as the GLP (Generalized Linear Preference) model from now on.

B. Analysis

In the GLP model the probability that node i increases its degree d_i is a function of that degree. We adapt the mean-field approach of [9] to analyze our model. We begin by assuming that d_i changes in a continuous manner and, thus, the probability $\Pi(d_i)$ can be interpreted as the rate at which d_i changes.

- Addition of m new links with probability p.

$$\frac{\partial d_i}{\partial t} = 2m \frac{d_i - \beta}{\sum_j (d_j - \beta)} \tag{2}$$

- Addition of a new node with probability $1-p$:

$$\frac{\partial d_i}{\partial t} = m \frac{d_i - \beta}{\sum_j (d_j - \beta)} \tag{3}$$

Since these two processes take place simultaneously, we have to sum up their contributions.

$$\begin{aligned} \frac{\partial d_i}{\partial t} &= p\frac{2m(d_i - \beta)}{\sum_j (d_j - \beta)} + (1-p)\frac{m(d_i - \beta)}{\sum_j (d_j - \beta)} \\ &= (1+p)\frac{m(d_i - \beta)}{\sum_j (d_j - \beta)} \end{aligned} \tag{4}$$

Since the total number of edges $E(t)$ is a linear function of time, $E(t) = mt$ and the total number of vertices vary with time as $V(t) = m_0 + (1-p)t$, we have
$\sum_j (d_j - \beta) = 2E(t) - V(t)\beta = 2mt - (m_0 + (1-p)t)\beta$.
Let t_i denote the time at which node i is added. The initial condition for node i is $d_i(t_i) = m$, and the solution for $d_i(t)$ takes the form

$$d_i(t) = \frac{(A(m,p,\beta)t - B(m,p,\beta))^{1/A(m,p,\beta)}(m-\beta)}{(A(m,p,\beta)t_i - B(m,p,\beta))^{1/A(m,p,\beta)}} + \beta \tag{5}$$

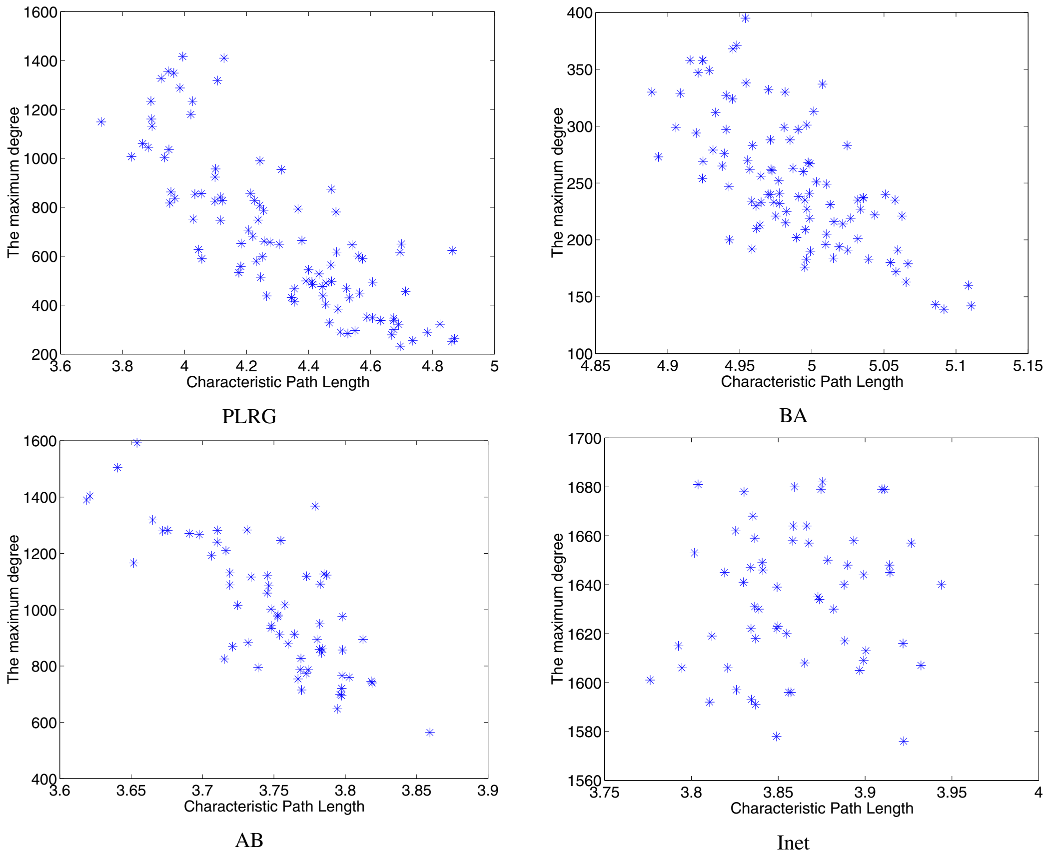


Fig. 6. The correlation between characteristic path length and the maximum degree for all generators.

where

$$A(m,p,\beta) = \frac{2m-\beta(1-p)}{(1+p)m} \quad (6)$$

$$B(m,p,\beta) = \frac{m_0\beta}{(1+p)m} \quad (7)$$

The probability that a node has a connectivity $d_i(t)$ smaller than d, $P(d_i(t) < d)$, can be written as

$$P(d_i(t) < d) = P(t_i > \frac{A(m,p,\beta)t - B(m,p,\beta)}{A(m,p,\beta)((d-\beta)/(m-\beta))^{A(m,p,\beta)}} + \frac{B(m,p,\beta)}{A(m,p,\beta)}) \quad (8)$$

Assuming that each operation of either adding a new node or a set of edges takes one unit of time, the probability density of t_i is $P_i(t_i) = 1/(m_0 + t)$. Thus

$$P(d_i(t) \geq d) = \frac{1}{m_0+t} \quad (\frac{A(m,p,\beta)t - B(m,p,\beta)}{A(m,p,\beta)((d-\beta)/(m-\beta))^{A(m,p,\beta)}} + \frac{B(m,p,\beta)}{A(m,p,\beta)}) \quad (9)$$

We have

$$\lim_{t\to\infty} P(d_i(t) \geq d) = (\frac{d-\beta}{m-\beta})^{-A(m,p,\beta)} \quad (10)$$

Thus as $t \to \infty$ the complementary cumulative connectivity distribution has the form

$$F(d) = P(d_i \geq d) \propto d^{-A(m,p,\beta)} \quad (11)$$

Since $\beta < 1, 0 \leq p < 1$ and $m \geq 1$, we have $1 < A(m,p,\beta) < \infty$. Thus, we can achieve any value of the power law exponent, $\alpha < -1$ through the proper choice of β, p, and m.

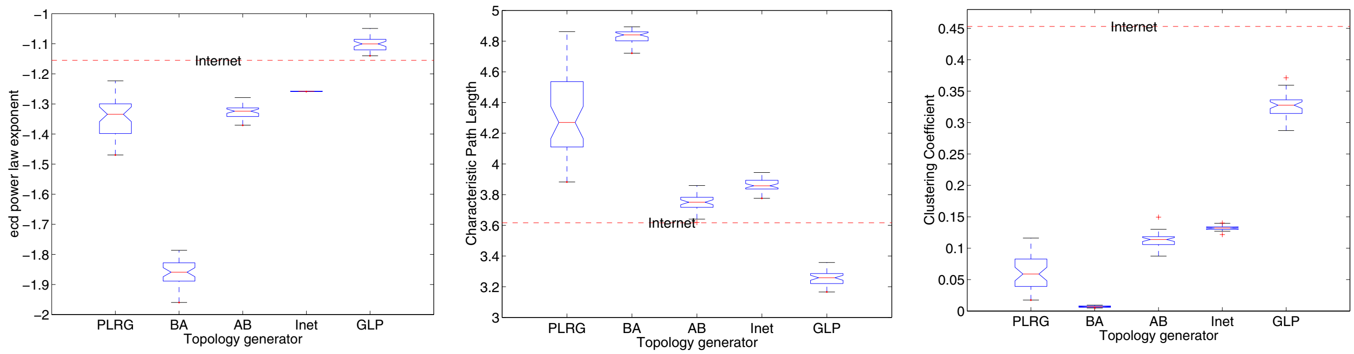
We now demonstrate the use of GLP to generate Internet-like topologies. As discussed, m, the initial degree of new nodes in the GLP model, is a constant integer. However, the initial degree can be a random variable with some distribution and the analysis of GLP model is still valid when m is interpreted as the average initial degree. It is reported in [13] that 86.82% of new ASs are born with degree one and 12.67% are born with degree two. That is, 99.49% are born with a degree less than or equal to two. We ignore the case where a node is born with a degree higher than two and let our generator produce a new node of initial degree one with probability .87 and a new node of initial degree two with probability .13. The average initial degree of new nodes is 1.13. This corresponds to let $m = 1.13$. We assume that the number of nodes N_V, the number of edges N_E and the exponent α of power law are given for now. p can be computed from

$$p = (N_E - mN_V)/N_E. \quad (12)$$

and β is obtained from

$$\frac{2m-\beta(1-p)}{(1+p)m} = -\alpha \quad (13)$$

Not that the graph produced by GLP may contain self loops and duplicate links. We delete self loops and merge duplicate links.



(a) ecd power law exponents

(b) Characteristic Path Length

(c) Clustering Coefficient

Fig. 7. Comparison of the GLP topology generator to existed topology generators

C. Model validation

We focus on the September 2000 Internet topology described in Section III. From Table I we know the number of node $N_V = 8613$, the number of edges $N_E = 18346$. And from Figure 1, $\alpha = -1.155$. We take $m = 1.13$ and obtain $p = 0.4695$ from Equation (12) and $\beta = 0.6447$ from Equation (13). Using these as inputs we produce 100 different topologies by using different initial random seeds. We then compute power law exponents α, clustering coefficients, and characteristic path lengths for these topologies. All of these results are box-plotted in Figure 7. The measured values from the AS-level Internet topology are shown as dashed lines. We observe that the value of α obtained from the generated graphs by GLP are very close to that of the Internet whereas the characteristic path length of the generated graph by GLP is smaller than that of the Internet. Although the clustering coefficients of the graph produced by the GLP model are also smaller than that of the Internet as is the case for other generators, they are much more accurate. In terms of all graph properties we measured, GLP exhibits less variability than PLRG but more variability than the other two generators. Note that by a careful choice of m, one can match the characteristic path length at the cost of obtaining a less accurate clustering coefficient.

We plot the correlation between the clustering coefficients and characteristic path lengths of the topologies produced by the GLP generator in Figure 8 and observe negative correlation.

The average over k largest degree(s) of the topologies produced by the GLP generator as well as the existing generators are plotted in Figure 9. The values of k are varied from 1 to 7. We observe from the figure that the median of the maximum degree of these topologies produced by the GLP generator is the closest to the maximum degree of our reference Internet topology. The variability of the average over largest degrees obtained from topologies produced by the GLP generator is close to that obtained from the topologies produced by PLRG and AB.

We investigate how the variability of the average largest degrees impact the clustering coefficients and characteristic path length by plotting the correlation between the maximum degree and clustering coefficients in Figure 10 and the correlation between the maximum degree and characteristic path length in Figure 11. We observe from these two figures that the clustering coefficient increases whereas the characteristic path length decreases as the maximum degree increases.

In addition to the September 2000 Internet topology, we have also performed the same experiments using other AS-level Internet topologies as reference and have observed similar behavior of the GLP topology generator.

We also examined the behavior of various topologies for producing larger topologies than the current Internet. To do so we double the number of nodes and edges required to be generated than in Section III while use the same power law exponents. We let each generator produce 50 topologies with same inputs but different random seeds. The results suggest that all of generators exhibit similar behaviors to what we demonstrated earlier.

V. Conclusions and Discussion

In this paper we examined the effectiveness of several recently proposed algorithms for generating power law topologies that model the Internet AS-level topology. Although they (mostly) match the power law exponent, they do less well in matching the characteristic path length and the clustering coefficient. Stimulated by the inadequacies of these algorithms, we proposed a variant of the incremental algorithm based on a "generalized linear preference" model and achieve better success.

In the process of conducting this work, we have also pointed out the benefits of using the empirical complementary distribution rather than the histogram for characterizing power law behavior. We also observed that the Internet AS-level topology exhibits the small world phenomena defined in [22]. We also observed considerable variability in the topologies produced by most generators. Thus, if a detailed study of a realistic Internet topology is required, then it may be best to use the actual topology.

Last, in a recent work, [13], the authors appear to argue that a topology generator is suspect if it does not *exactly* account for the evolution of the topology over time. We disagree with this and believe that the quality of a topology should be determined by its resulting graph rather than on the mechanics of the algorithm producing it. We consider this paper to be a step in this direction toward developing metrics with which to distinguish power law generating.

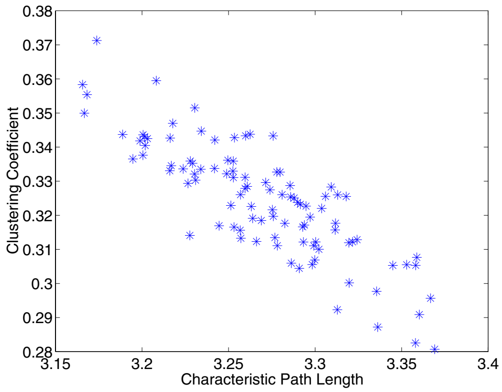


Fig. 8. The correlation between clustering coefficient and characteristic path length for the GLP generator.

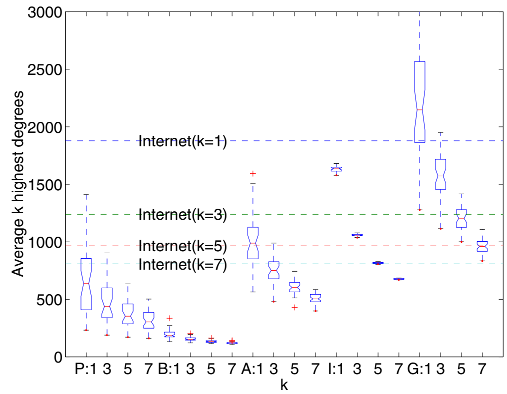


Fig. 9. Comparison of the average k largest degree of the GLP generator to these existing generators. P:PLRG, B:BA, A:AB, I:Inet,G:GLP

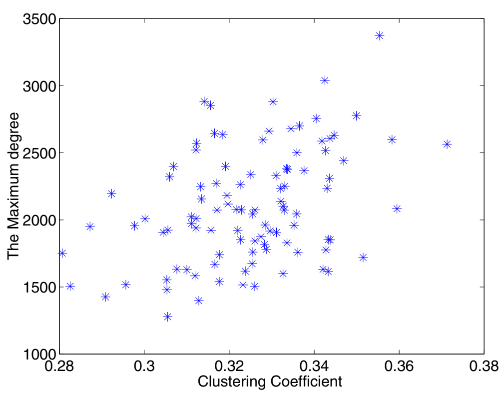


Fig. 10. The correlation between clustering coefficient and the maximum degree for the GLP generator.

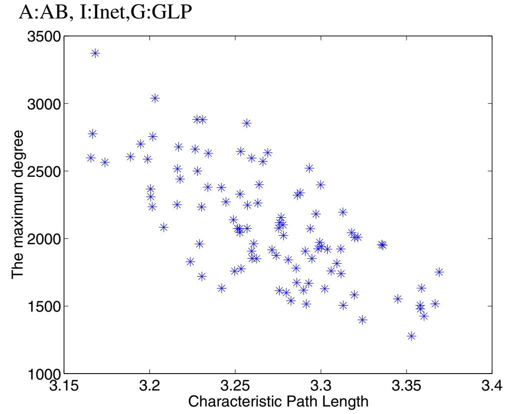


Fig. 11. The correlation between characteristic path length and the maximum degree for the GLP generator.

References

[1] University of Oregon Route Views Project. *http://www.antc.uoregon.edu/route-views/*

[2] Routing Information Service *http://www.ripe.net/ripencc/pub-services/np/ris-index.html*

[3] L. Adamic and B. Huberman "Scaling Behavior of the World Wide Web". *A comment in Science, 287 (2000), 2115*

[4] L. Adamic, R. Lukose, A. Puniyani, B. Huberman "Search in Power-Law Networks" *Preprint, 2001.*

[5] W. Aiello, F. Chung, and L. Lu. "A random graph model for massive graphs." *In proc. of the 32rd Annual ACM Symposium on Theory of Computing, 2000*

[6] R. Albert and A. Barabasi. "Topology of Evolving Network: Local Events and Universality." *Physica Review Letters, 85:5234-5237,2000*

[7] R. Albert, H. Jeong, and A. Barabasi "Diameter of the World Wide Web." *Nature 401 130-131 (1999)*

[8] R. Albert, H. Jeong, and A. Barabasi "Attack and error tolerance of complex networks." *Nature 406 378 (2000)*

[9] A. Barabasi, R. Albert, H. Jeong. "Mean-field Theory for Scale-free Random Networks." *Physica A 272 (1999)*

[10] A. Barabasi and R. Albert. "Emergence of scaling in random networks." *Science, pages 509-512, October 1999*

[11] A. Broido and K. Claffy. "Internet topology: connectivity of IP graphs" *proc. of SPIE International symposium on Convergence of IT and Communication*

[12] H. Chang, R. Govindan, S.Jamin, S. Shenker, and W. Willinger, "Towards Capturing Representative AS-level Internet Topologies", Submitted for publication, 2001

[13] Q. Chen, H, Chang, R. Govindan, S.Jamin, S. Shenker, and W. Willinger, "The Origin of Power Laws in Internet Topologies Revisited," *Proc. of INFOCOM 2002*

[14] M. Faloutsos, P. Faloutsos, C. Faloutsos. "On power-Law Relationships of the Internet Topology." *Proc. of ACM SIGCOMM'99, Aug. 1999*

[15] G. Huston. "Architectural Requirements for Inter-Domain Routing in the Internet," IETF Draft draft-iab-bgparch-01.txt.

[16] C. Labovitz, A. Ahuja, R. Wattenhofer, S. Venkatachary. "The Impact of Internet Policy and Topology on Delayed Routing Convergence." *Proc. of INFOCOM 2001*

[17] A. Medina, I. Matta, and J. Byers. "On the Origin of Power Laws in Internet Topologies." *ACM Computer Communications Review, April 2000.*

[18] R. Pastor-Satorras and A. Vespignant. "Epidemic Spreading in Scale-Free Networks" *Physica Review Letters,86,2001*

[19] K. Park and H. Lee. "On the effectiveness of route-based packet filtering for distributed DoS attack prevention in power-law Internet." *To appear in Proc. ACM SIGCOMM, 2001.*

[20] C.Jin, Q.Chen and S. Jamin "Inet: Internet Topology Generator." *Technique Report CSE-TR-433-00, University of Michigan, EECS dept. 2000. http://topology.eecs.umich.edu/inet.*

[21] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, W. Willinger "Network Topologies, Power Laws, and Hierarchy." *Technical Report 01-746, Computer Science Department, University of Southern California.*

[22] D.J. Watts, S.H. Strogatz "Collective dynamics of "small-world" networks." *Nature **393**, 440-442, 1998*

[23] D. J. Watts, *Small World*, Princeton University Press, 1999.

[24] E. Zegura, K. Calvert, and M. Donahoo. "A Quantitative Comparison of Graph-Based Model for Internet Topology." *ACM/IEEE Transactions on Networking, 5(6):770-783, Dec. 1997*