On the Origin of Power Laws in Internet Topologies*

Alberto Medina Ibrahim Matta John Byers
Computer Science Department
Boston University
Boston, MA 02215
{amedina, matta, byers}@cs.bu.edu

ABSTRACT

Recent empirical studies [6] have shown that Internet topologies exhibit power laws of the form $y = x^{\alpha}$ for the following relationships: (P1) outdegree of node (domain or router) versus rank; (P2) number of nodes versus outdegree; (P3) number of node pairs within a neighborhood versus neighborhood size (in hops); and (P4) eigenvalues of the adjacency matrix versus rank. However, causes for the appearance of such power laws have not been convincingly given. In this paper, we examine four factors in the formation of Internet topologies. These factors are (F1) preferential connectivity of a new node to existing nodes; (F2) incremental growth of the network; (F3) distribution of nodes in space; and (F4) locality of edge connections. In synthetically generated network topologies, we study the relevance of each factor in causing the aforementioned power laws as well as other properties, namely diameter, average path length and clustering coefficient. Different kinds of network topologies are generated: (T1) topologies generated using our parametrized generator, we call BRITE[1]; (T2) random topologies generated using the well-known Waxman model [12]; (T3) Transit-Stub topologies generated using GT-ITM tool [3]; and (T4) regular grid topologies. We observe that some generated topologies may not obey power laws P1 and P2. Thus, the existence of these power laws can be used to validate the accuracy of a given tool in generating representative Internet topologies. Power laws P3 and P4 were observed in nearly all considered topologies, *but* different topologies showed different values of the power exponent α. Thus, while the presence of power laws P3 and P4 do not give strong evidence for the representativeness of a generated topology, the value of α in P3 and P4 can be used as a litmus test for the representativeness of a generated topology. We also find that factors F1 and F2 are the key contributors in our study which provide the resemblance of our generated topologies to that of the Internet.

*This work was supported in part by NSF grants CAREER ANIR-9701988 and MRI EIA-9871022.

[1]for Boston university Representative Internet Topology gEnerator. Available at `http://www.cs.bu.edu/fac/matta/software.html`

ACM SIGCOMM Computer Communication Review, Vol. 30, No. 2, April 2000.

1. INTRODUCTION

The accurate characterization of Internet topologies is key to the successful modeling and analysis of the Internet and its protocols [9]. Internet studies usually assume certain topologies or use synthetically generated topologies. These topologies *must* reflect properties (or invariants) empirically found in the actual existing structure of the Internet; otherwise, correct conclusions cannot be drawn.

Recently [6], it was observed that actual Internet topologies exhibit power laws of the form $y = x^{\alpha}$ for the following relationships: (P1) outdegree of node (domain or router) versus rank; (P2) number of nodes versus outdegree; (P3) number of node pairs within a neighborhood versus neighborhood size (in hops); and (P4) eigenvalues of the adjacency matrix versus rank. However, causes for the appearance of such power laws have not been convincingly given.

In this paper, we consider four factors in the formation of Internet topologies. These factors are (F1) preferential connectivity; (F2) incremental growth; (F3) geographical distribution of nodes; and (F4) locality of edge connections. F1 dictates the tendency of a new node to connect to those existing nodes that have higher outdegrees. F2 dictates that new nodes join the Internet in an incremental way. F3 determines how nodes are distributed in space — in this paper, we consider random and heavy-tailed distribution of nodes. Finally, F4 dictates the tendency of a new node to connect to nearby nodes instead of far-away nodes.

In synthetically generated network topologies, we study the relevance of each factor in causing the aforementioned power laws as well as other properties, namely diameter, average path length and clustering coefficient. Different kinds of network topologies are generated: (T1) topologies generated using our parametrized generator BRITE; (T2) random topologies generated using the well-known Waxman model [12]; (T3) Transit-Stub topologies generated using GT-ITM tool [3]; and (T4) regular grid topologies. We observe that some generated topologies do not obey power laws P1 and P2. Thus, the existence of these power laws can be used to validate the accuracy of a given tool in generating representative Internet topologies. Power laws P3 and P4 were observed in nearly all considered topologies, *but* different topologies showed different values of the power exponent α. Thus, while the presence of power laws P3 and P4 do not give strong evidence for the representativeness of a generated topology, the value of α in P3 and P4 can be used as a litmus test for the representativeness of a generated topology. We also find that factors F1 and F2 are the key contributors in our study which provide the resemblance of our generated topologies to that of the Internet. Although some recent studies (e.g. [2]) have examined such factors for *some* power laws, in this paper, we

examine their relevance with respect to these as well as other power laws and metrics on a *large number* of topologies synthetically generated using commonly used tools.

The rest of the paper is organized as follows. In Section 2, we summarize three topology generation methods commonly used in Internet studies. In Section 3, we briefly describe the power laws found in [6]. Section 4 illustrates the absence or weak presence of power laws in topologies synthetically generated using existing tools. In Section 5, we briefly discuss preferential connectivity and incremental growth (factors F1 and F2 above) and their role in generating topologies that exhibit power laws. Section 6 argues for the geographical distribution of nodes and locality of edge connections (factors F3 and F4 above) and their role in making synthetically generated topologies more realistic. Section 7 describes BRITE, a parametrizable topology generator we built. Using BRITE, we present in Sections 8 and 9 our experiments and comparative analysis of generated topologies in order to assess the relevance of each factor. Section 10 concludes the paper with future work.

2. COMMONLY USED TOPOLOGY GENERATION METHODS

In this section, we briefly describe three commonly used topology generation methods [14].[2] We use these methods in Section 9 to analyze different kinds of synthetically generated topologies.

Random Method: In this method, a (fixed) set of nodes is distributed in a plane uniformly at random. A link is added between each pair of nodes with a certain probability. The Waxman method is an instantiation of this method where the probability of adding a link is given by:

$$P(u,v) = \alpha\, e^{-d/(\beta L)} \quad (1)$$

where $0 < \alpha, \beta \leq 1$, d is the Euclidean distance from node u to node v, and L is the maximum distance between any two nodes.

A limitation of this method is that it does not impose any large-scale structure among nodes. In particular, it is difficult to control the configurations so as to generate large, sparse, but connected Internet-like topologies.

Regular Method: This method generates regular topologies with a specific and rigid structure and thus have no randomness at all. Regular topologies such as grids are often used in analytic studies of algorithm performance to make the analysis tractable.

Hierarchical Method: In this method, hierarchical topologies are created by connecting small random graphs together according to a larger-scale structure. Thus this method attempts to provide a balance between randomness and structure.

The Transit-Stub method [3] tries to impose a more Internet-oriented hierarchical structure as follows. A connected random graph is first generated (e.g. using the Waxman method described above). Each node in that graph represents an entire *Transit domain*. Each Transit domain node is expanded to form another connected random graph, representing the backbone topology of that transit domain. Next, for each node in each transit domain, a number of connected random graphs are generated, representing *Stub domains* that are attached to that transit node. Finally, some extra connectivity is added, in the form of "back-door" links between pairs of nodes, where a pair consists of a node from a transit domain and another from a stub domain, or one node from each of two different stub domains. By having nodes of different types, it is possible to generate large sparsely-connected Internet-like topologies with typically low node degrees.

[2]Reference [14] provides detailed explanation of each method.

3. POWER LAWS IN INTERNET TOPOLOGIES

Recently [6] it has been observed that certain properties of Internet topologies can be described using **power laws** of the form $y = x^{\alpha}$. This implies that those same distributions of interest in Internet topologies are *skewed*. As a result, Internet studies which assume that the distributions are not skewed, perhaps by taking into account only the average values from these distributions, can result in misleading results. For example, for a particular snapshot of the Internet topology in 1998, 85% of the nodes had outdegree less than the average. Power laws have been used to describe several characteristics of communication networks (e.g. traffic [4]), but not their topologies. In this section we summarize the main observations in [6] — see [6] for more details and how some power laws can be used to estimate other related metrics.

Power laws of the form $y = x^{\alpha}$ enable a compact characterization of topologies through their exponents. If such (x, y) relationships are plotted on a log-log scale, then the power-law exponent α defines the slope of the resulting linear plot. We can then use slopes of these linear plots to assess whether two topologies have similar properties. Indeed, Faloutsos $et\ al.$ [6] show that different Internet domain-level topologies of different size and observed at different times have almost equal slopes. On the other hand, their observations indicate that a router-level topology has slopes that significantly differ from those of domain-level topologies.

P1 (rank exponent): Consider the relationship between the outdegrees of nodes, sorted in decreasing order, and the ranks of nodes in the resulting order. The outdegree d_v of a node v, is proportional to the rank of the node, r_v, to the power of a constant R, that is, $d_v \propto r_v^R$.

P2 (outdegree exponent): The frequency (number of nodes), f_d, of an outdegree, d, is proportional to the outdegree to the power of a constant O, that is, $f_d \propto d^O$.

P3 (hop-plot exponent): The total number of pairs of nodes, $P(h)$, within h hops, is proportional to the number of hops to the power of a constant H, that is, $P(h) \propto h^H, h \ll \delta$, where δ is the diameter of the network. [3]

P4 (eigen exponent): The eigenvalues, λ_i, of the adjacency matrix of a topology, sorted in decreasing order, are related as follows: $\lambda_i \propto i^E$. [4]

4. LIMITATIONS OF EXISTING TOPOLOGY GENERATION TOOLS

Waxman [12] and Transit-Stub [3] are two of the most commonly used tools for generating network topologies. To test their accuracy in generating topologies that resemble the Internet (i.e. obey the same power laws empirically found in [6]), we show in Figure 1

[3]There is some controversy whether the neighborhood size, defined by $P(h)$ in [6], is a power law; the data in [10] show the neighborhood size to grow exponentially.

[4]Reference [6] considered the 20 largest eigenvalues.

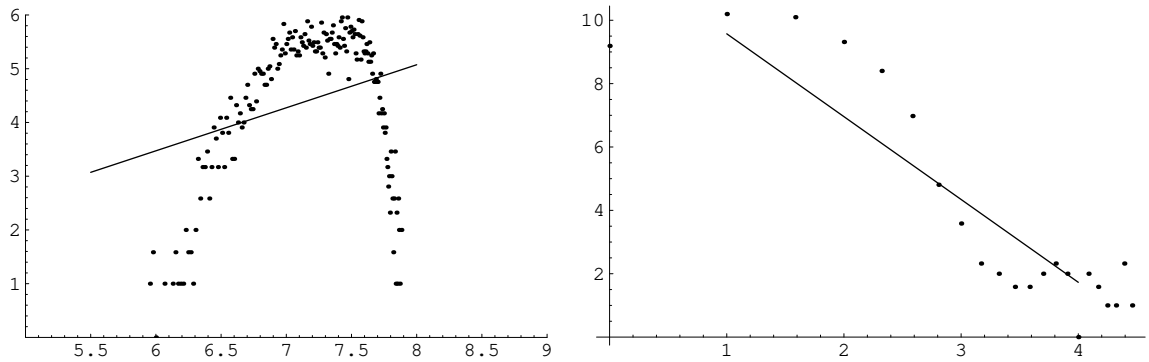

Figure 1: Log-log plot of frequency f_d vs. outdegree d for a 5000-node Waxman topology (left) and a 6660-node Transit-Stub topology (right). The correlation coefficient is 0.4 for the Waxman topology, and 0.9 for the Transit-Stub topology.

a fit for the log-log plot of the outdegree exponent (cf. power law P2 in Section 3). The linear fit is clearly not good for both Waxman and Transit-Stub, although it is much better for the Transit-Stub topology. It is not surprising that both topologies exhibit the absence or weak presence of the power law given that both tools are not based on a dynamical growth model of the Internet, which we subsequently discuss in Section 5. For example, both tools do not gradually add nodes and links to form a topology, rather they create *all* nodes and then start adding links to interconnect them. Section 9 presents an extensive comparative analysis of different topology generation models.

5. HOW POWER LAWS ARISE

In this section, we briefly discuss recent studies that suggest preferential connectivity and incremental growth to be possible causes for some power laws found in topologies. In this paper, we examine their relevance with respect to these as well as other power laws and metrics on a large number of topologies synthetically generated using commonly used tools.

Faloutsos $et\ al.$ [6] provide evidence but not convincing possible causes for the existence of power laws P1-P4 in Internet topologies. Recently in [1], power law P2 has also been observed in the topology of the World Wide Web. Here the nodes are *documents* and the links are *hyperlinks*. Huberman and Adamic [7] observe that the relationship between the number of web pages and the different web sites also follows a power law — many sites have only a few pages, while very few sites have hundreds of thousands of pages. (This may be viewed as equivalent to power law P1 in Internet topologies.) Using a stochastic dynamical growth model, power laws are argued to arise when sites grow at the same average rate, thus sites that are large become larger over time.

Barabási and Albert [2] suggest two possible causes for power law P2 in any network topology: incremental growth and preferential connectivity. *Incremental growth* refers to "open" networks that form by the continual addition of new nodes, and thus the gradual increase in the size of the network. *Preferential connectivity* refers to the tendency of a new node to connect to existing nodes that are highly connected or popular.

6. NODE PLACEMENT AND CONNECTION LOCALITY

In this section, we argue for two other possible causes for the existence of power laws in Internet topologies, in addition to incremental growth and preferential connectivity. The first is the way nodes of the network are distributed in space. Unlike random models, we conjecture that Internet topologies have a high degree of clustering. Thus, models that would generate topologies, where nodes are distributed in space according to a *skewed* (e.g. heavy-tailed) distribution, appear more realistic. This, for example, mirrors the skewed distribution of human population or web pages over web sites [7]. Another possible cause for power laws is the tendency of a new node to connect to existing node(s) that are close-by in distance.

Figure 2 shows a snapshot of a section of a topology (without links) in which nodes were placed randomly or according to a heavy-tailed distribution. The motivation behind using heavy-tailed placement of nodes together with locality of edge connections is to obtain a skewed distribution of outdegrees, where those few heavily populated areas have highly connected nodes, while the rest of the nodes are sparsely connected.

7. BRITE: A TOPOLOGY GENERATOR

To study the relevance of possible causes for power laws observed in Internet topologies [6], we built a parametrized topology generator, we call BRITE[5]. Different combinations of possible causes can be tested. We consider the possible causes described earlier, namely: preferential connectivity, incremental growth, node placement, and connection locality. For each combination, we analyze the generated topologies in terms of power laws and other metrics observed in real networks [6, 8]. In this section we describe the current version of BRITE. Sections 8 and 9 present our experiments and results, respectively. Table 1 lists the various parameters of BRITE. We describe each parameter next.

The Plane:

The nodes of the generated topology are distributed in a plane divided into `HS` $\times$ `HS` squares. Each one of these high-level squares is further subdivided into smaller `LS` $\times$ `LS` low-level squares. Each low-level square can be assigned at most one node.

Node Assignment:

A `Random` placement of nodes in the plane is achieved by simply selecting a low-level square randomly and dropping a node there

[5]for Boston university Representative Internet Topology gEnerator. Available at `http://www.cs.bu.edu/fac/matta/software.html`

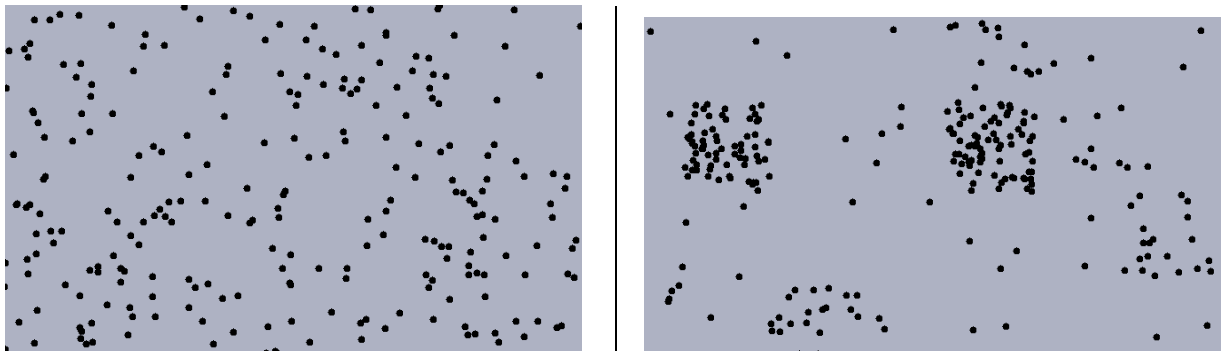

Figure 2: Snapshot of random node placement (left) and heavy-tailed node placement (right)

Parameter	Meaning	Values
HS	Size of one side of the plane	Integer > 1
LS	Size of one side of a high-level square	Integer ≥ 1
NP	Node Placement	0: Random, 1: Heavy-Tailed
m	Number of links added per new node	Integer ≥ 1
PC	Preferential Connectivity	0: NONE, 1: ONLY, 2: BOTH
IG	Incremental Growth	0: INACTIVE, 1: ACTIVE

Table 1: Parameters of BRITE

while avoiding collisions. To achieve a Heavy-Tailed distribution of nodes, for each one of the high-level squares, the generator picks a number of nodes n to be assigned to that square according to a *bounded Pareto distribution* given by [5]:

$$f(n) = \frac{a\ k^a\ n^{-a-1}}{1-(k/P)^a} \tag{2}$$

A node is then placed randomly in one of the LS $\times$ LS low-level squares while avoiding collisions.

Number of Links for a New Node:

The parameter m controls the number of neighbor nodes to which a new node connects when it joins the network (or in other words, the number of new links to be added to the topology). The greater the value of m, the denser the generated topology. We refer to the set of nodes from which a neighbor is selected for a newly considered node as the *candidate neighbor set*.

Incremental Growth:

This parameter controls incremental growth and can take one of two values:

- INACTIVE places all nodes at once in the plane before adding any link. At each step, a node is randomly selected and m links are used to connect it to m candidate neighbors from *all* other nodes.
- ACTIVE places nodes in the plane gradually one at a time as they join the network. In this case, a new node considers as candidate neighbors only those nodes that have already joined the network (i.e. nodes that are already connected to some other node(s)).

  Initially, before operating in either INACTIVE or ACTIVE mode, the generator generates a small randomly connected backbone of m_0 nodes. The remaining nodes are then connected.

Preferential Connectivity:

This parameter controls the activation or deactivation of both preferential connectivity and locality. There are three possible values for this parameter:

- NONE indicates that preferential connectivity is turned off. In this case, a newly considered node connects to a candidate neighbor node using Waxman's probability function [12] (cf. equation 1). This process is repeated to connect the new node to m nodes.
- ONLY means that preferential connectivity is turned on. In this case, a newly considered node v connects to a candidate neighbor node i with the following probability:

  $$\frac{d_i}{\sum_{j \in \mathcal{C}} d_j}$$

  where d_i is the current outdegree of node i, and $\mathcal{C}$ is the set of candidate neighbor nodes. This process implies that a new node joining the network selects with high probability those nodes with high outdegrees. This is repeated to connect v to m nodes.
- BOTH combines preferential connectivity and connection locality. In this case, for a newly considered node v, we compute for each candidate neighbor node i a Waxman's probability w_i (cf. equation 1). This gives preference to close-by nodes. Then, the final probability of connecting to node i is computed as follows:

  $$\frac{w_i\ d_i}{\sum_{j \in \mathcal{C}} w_j\ d_j}$$

  This process is repeated to connect v to m nodes.

8. EXPERIMENTS

In this section we describe our experiments. Section 9 presents our results and observations. We consider topologies of size ranging from 500 nodes to about 15000 nodes.[6] For each selected size and each combination of parameters described in Section 7, three BRITE topologies were generated using three different random seeds, and the average results are plotted.

Section 9 shows results for BRITE topologies generated starting from a backbone of size $m_0 = 100$ on a plane with $\texttt{HS} = 1000$ and $\texttt{LS} = 100$. In the heavy-tailed node placement experiments, we take the parameters of the bounded Pareto distribution (cf. equation 2) as follows: a minimum value k of 1, maximum value P of $10000 \times \texttt{LS} \times \texttt{LS}$ and shape parameter a of 1.[7] Also, for a newly considered node, we connect it with only one link, i.e. $\texttt{m} = 1$. We obtained similar results for different values of HS, LS and m, and thus are not shown. This includes the value of $\texttt{m} = 5$ considered in [2]. For network sizes of 500 to 15000, as m increases from 1 to 5, the average node degree ranges from 2 to 10 with incremental growth, and ranges from 30 to 100 without incremental growth. Note that with incremental growth, the average node degree is approximately 2m. We also do not show results for the case when both preferential connectivity and connection locality are activated (i.e. PC = 2) as it yields results similar to the case when only preferential connectivity is present (i.e. PC = 1).

We compare the results from **BRITE** topologies with topologies generated using other models and tools. We generate topologies according to the **Waxman** method, using the GeorgiaTech topology generator, GT-ITM [14]. In our experiments, we set the parameters of Waxman's function to $\alpha = 0.2$ and $\beta = 0.15$.[8] Also using the GT-ITM tool, we generate **Transit-Stub** topologies. Table 2 lists the parameters we used to generate Transit-Stub topologies of various sizes. For the considered network sizes, the average node degree for Waxman topologies ranges from 15 to 500, whereas Transit-Stub topologies have an average node degree of about 3. We also generate **regular grid** topologies.

For each generated topology, we verify the presence or absence of the power laws observed in [6] (cf. Section 3). For the log-log plots obtained, we test the existence or absence of power laws using *Pearson's* correlation coefficient[9] — the closer the absolute value of the correlation coefficient is to one, the closer the data follows a power law, that is, we have a good linear fit on a log-log plot. For the different kinds of topologies we generate, we obtain graphs of the correlation coefficient as a function of the topology size. A good linear fit should give a correlation coefficient of 0.95 or higher. If a given power law is found across different kinds of topology, then we analyze the slopes of the linear fits to determine which is closer to the slopes observed in [6].

We also measure for each generated topology other characteristics, namely diameter, average path length, and clustering coefficient.

[6] Due to the high computational cost of computing eigenvalues, we only consider topology sizes up to 7000 nodes for the eigenvalue exponent results.

[7] To obtain a sufficiently skewed distribution, we take P greater than $\texttt{LS} \times \texttt{LS}$, which is the maximum number of nodes that can be assigned to a high-level square. Thus we take this maximum to be the number of nodes we assign to a high-level square if the random number generated turns out greater.

[8] These are the same parameter values selected in [14].

[9] We obtain correlation coefficients using *Mathematica* [13].

The *average path length* in a topology is defined as the average path length taken over all pairs of nodes. The *clustering coefficient* is defined as follows [11]: consider any node v and its set of k_v neighbors. At most $\binom{k_v+1}{2}$ edges can appear between these nodes. The clustering coefficient C_v for this node is the fraction of that set of edges which are actually present in the graph. The clustering coefficient of the graph is then defined to be the average value of C_v, averaging over all nodes.

9. RESULTS

Before presenting our results, we first list general observations:

- The rank and outdegree power laws (cf. P1 and P2 in Section 3) are the most effective in distinguishing different kinds of topologies.
- The hopplot and eigenvalue power laws (cf. P3 and P4 in Section 3) are observed in nearly all topologies that we studied, however, different topologies differ in their H and E exponents.
- Preferential connectivity and incremental growth are found to be the main causes for all power laws in our simulations.

9.1 Rank Exponent

For different topologies, we consider the log-log relationship between outdegree and rank of a node for varying topology sizes. Then we plot the correlation coefficient obtained when we perform a linear fit of those plots as the topology size grows.

9.1.1 Waxman, Transit-Stub and Grid Topologies

Figure 3 shows the results for Waxman, Transit-Stub and grid topologies. Waxman and Transit-Stub topologies give correlation coefficients that are around 0.8 and 0.85, respectively. This indicates poor linear fits. For all Internet topologies studied in [6], the correlation coefficients were higher than 0.95. Thus, Waxman and Transit-Stub topologies do not seem to be representative of Internet topologies, at least with respect to the rank exponent.

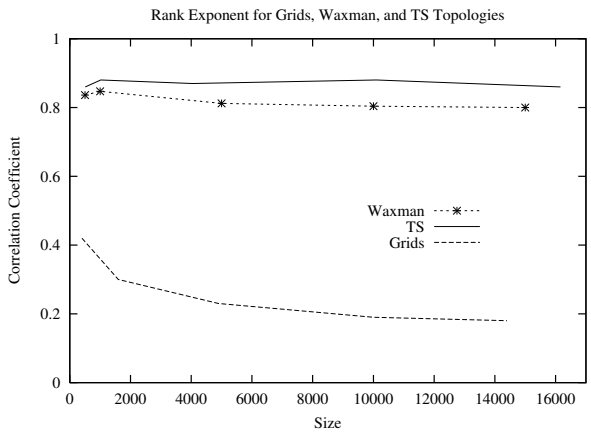

Figure 3: Correlation Coefficient vs. Size (Rank Exponent for Waxman, Transit-Stub and grid topologies)

As expected, grid topologies show very poor linear fits especially for large sizes. This is because nodes have only degrees of 2, 3 and 4, where only the four corner nodes have degree of 2, other nodes at the periphery have degree of 3 and the rest have degree of 4.

Size	Transit domains	Nodes/Transit domain	Stubs/Transit node	Nodes/Stub
510	10	1	5	10
1020	20	1	5	10
4040	20	2	10	10
10100	100	1	10	10
16160	20	8	10	10

Table 2: Parameters of Transit-Stub topologies

9.1.2 BRITE Topologies

Figures 4 and 5 show results for topologies obtained using our generator. Figure 4 corresponds to random node placement while Figure 5 corresponds to heavy-tailed (bounded Pareto) node placement. In Figure 4, for different combinations of incremental growth and preferential connectivity, we observe the following:

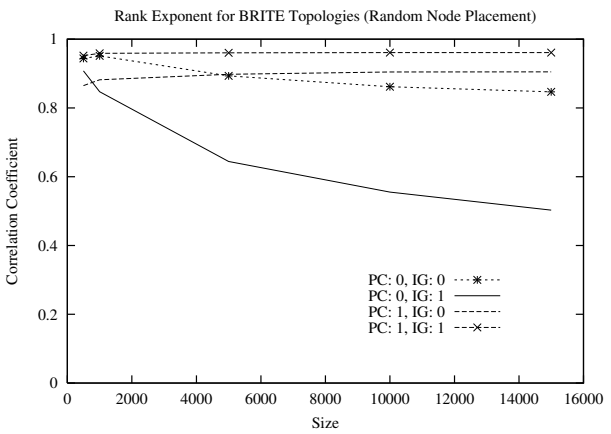

Figure 4: Rank Exponent for BRITE topologies (Random Node Placement)

- `PC` **set to** `1` **and** `IG` **set to** `1`**:** Topologies generated with *both* preferential connectivity and incremental growth show excellent correlation coefficients approaching 1. In this case, the slopes of the linear fits are around -0.6 for all generated topologies, which is very close to the value of -0.5 observed in [6] for actual Internet topologies. This suggests that if the rank power law is a fundamental property of Internet topologies, then using a topology generation model that includes *both* preferential connectivity and incremental growth will produce topologies that highly resemble Internet toplogies with respect to the rank power law.
- `PC` **set to** `1` **and** `IG` **set to** `0`**:** Topologies generated with preferential connectivity but without incremental growth show relatively high correlation coefficients (approaching 0.9). This indicates that, for the rank exponent, preferential connectivity plays a more important role than the incremental growth property.
- `PC` **set to** `0` **and** `IG` **set to** `1`**:** Topologies generated with incremental growth but without preferential connectivity show the lowest correlation coefficients in terms of the rank exponent. This suggests that preferential connectivity is a necessary condition for the rank exponent power law to hold.
- `PC` **set to** `0` **and** `IG` **set to** `0`**:** Topologies generated without incremental growth nor preferential connectivity also show poor correlation. This setting corresponds to a variant of Waxman's random model.

Figure 5 shows the results when nodes are distributed according to a Pareto distribution. We observe the following:

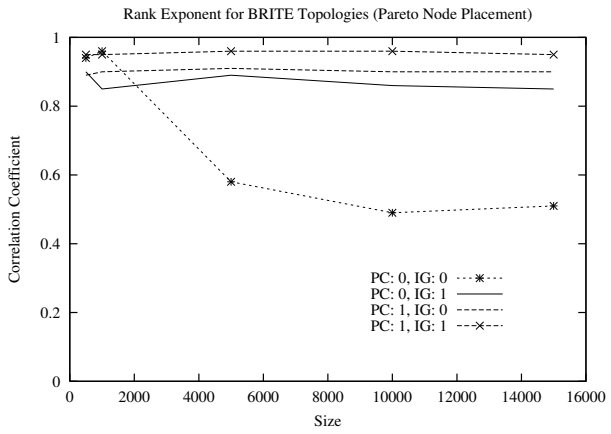

Figure 5: Rank Exponent for BRITE topologies (Pareto Node Placement)

- Topologies generated without preferential connectivity nor incremental growth show the worst correlation coefficients.
- In the presence of incremental growth, the skewed distribution of nodes makes the generation model less sensitive to the absence of preferential connectivity. The values of the correlation coefficient are relatively high (around 0.85), but the linear fits are still not good.
- The performance of topologies generated with preferential connectivity alone or with both preferential connectivity and incremental growth are comparable to those observed with random node placement. This suggests that skewed node placement may not be a fundamental cause for the appearance of power laws in Internet topologies.

Figure 6 shows a fit for the log-log plot of the rank exponent. The linear fit is clearly best for BRITE topologies with both preferential connectivity and incremental growth (under random node placement).

9.1.3 Conclusions on Rank Exponent

With respect to the rank exponent, preferential connectivity seems to be a necessary condition for the power law to hold. The presence of incremental growth increases the correlation coefficients,

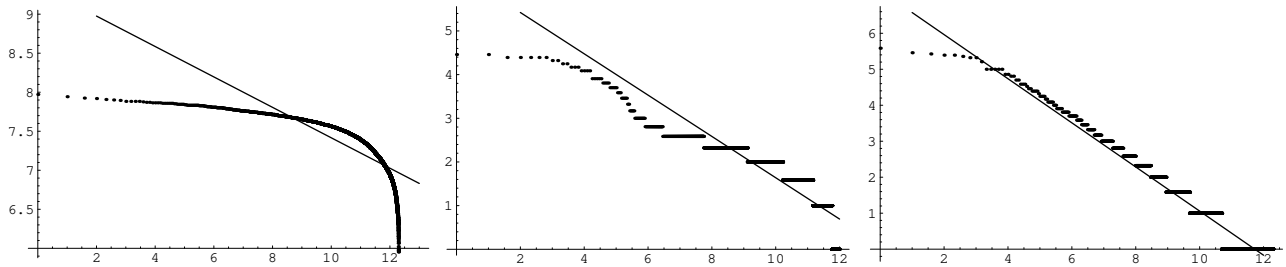

Figure 6: Log-log plot of outdegree d_v vs. rank for a 5000-node Waxman topology (left), a 4040-node Transit-Stub topology (middle) and a 5000-node BRITE topology with preferential connectivity and incremental growth (right). The correlation coefficient is 0.81 for the Waxman topology, 0.87 for the Transit-Stub topology, and 0.96 for the BRITE topology.

resulting in excellent linear fits. For Waxman, Transit-Stub and grid topologies, the rank exponent did not appear as the linear fits did not show high degrees of correlation. This suggests that the rank exponent can effectively distinguish different classes of topologies.

9.2 Outdegree Exponent

For different topologies, we plot for varying topology sizes the correlation coefficient of the log-log relationship between frequency and outdegree (cf. Section 3). As in [6], we plot outdegrees starting from one until an outdegree with exactly one node is found.

9.2.1 Waxman, Transit-Stub and Grid Topologies

Figure 7 shows the results for Waxman, Transit-Stub and grid topologies. Waxman topologies give very low correlation coefficients, indicating that the outdegree power law is not present. Transit-Stub topologies give correlation coefficients that are around 0.9, indicating that the linear fits are not good. Again, this suggests that, at least with respect to the outdegree exponent, Waxman and Transit-Stub topologies do not resemble Internet topologies.

For grid topologies, agreement with the outdegree power law seems to be present. However, this may be misleading because any grid has only three different degrees (2, 3 and 4), resulting in only three nearly colinear points in the log-log plot, and consequently the data can be easily fit to a straight line.

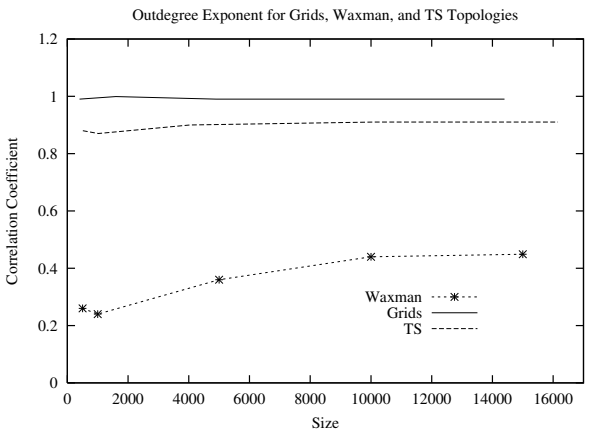

Figure 7: Correlation Coefficient vs. Size (Outdegree Exponent for Waxman, Transit-Stub and grid topologies)

9.2.2 BRITE Topologies

Barabási and Albert [2] show that *both* preferential connectivity and incremental growth are needed for the outdegree power law to exist. Our results are consistent with theirs as observed in Figures 8 and 9. Thus, any topology generation model should include both properties.

For the router-level topology analyzed in [6], the value of the outdegree exponent was -2.48. For BRITE topologies generated with both preferential connectivity and incremental growth, the slope of the linear fits was in the same range — from -2.03 for a 500-node network to -2.36 for a 15000-node network.

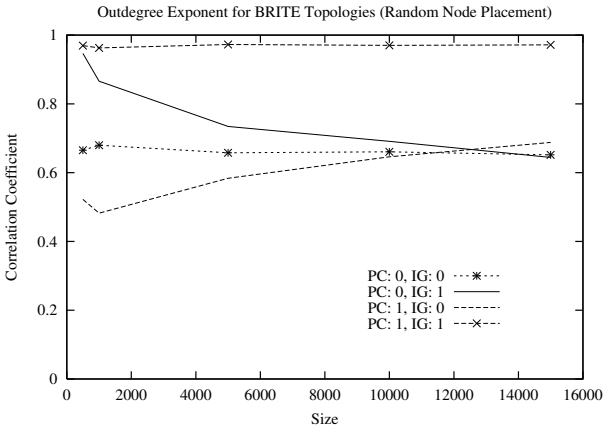

Figure 8: Outdegree Exponent for BRITE topologies (Random Node Placement)

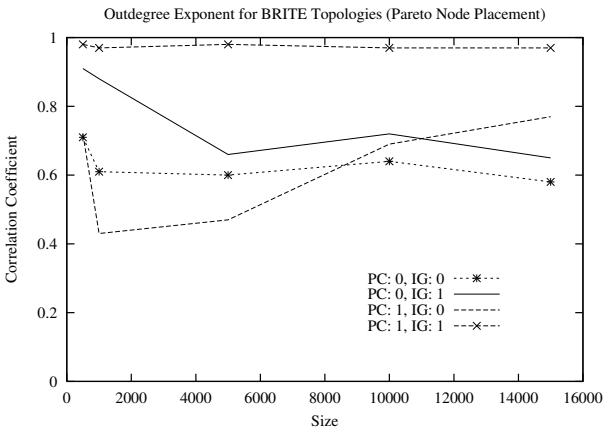

Figure 9: Outdegree Exponent for BRITE topologies (Pareto Node Placement)

9.2.3 Conclusions on Outdegree Exponent

Waxman and Transit-Stub topologies do not exhibit the outdegree power law. BRITE topologies generated with both preferential connectivity and incremental growth closely resemble Internet topologies reported in [6]. Thus, like the rank exponent, the outdegree exponent can effectively distinguish different classes of topologies. Since conclusions under Pareto node placement are consistent with those under random node placement, we henceforth show results only for the latter.

9.3 Hopplot Exponent

For different topologies, we plot for varying topology sizes the correlation coefficient of the log-log relationship of the number of node pairs within a certain distance versus the distance (cf. Section 3). Here we consider distance values up to the diameter of the topology.

9.3.1 Waxman, Transit-Stub and Grid Topologies

Figure 10 shows the results for Waxman, Transit-Stub and grid topologies. All Waxman topologies yield linear fit correlation coefficients close to 1, indicating that the hopplot power law holds. However, this may be misleading given that the number of data points was too low (around 3 or 4 data points). The reason for having so few data points has to do with a fundamental limitation of random methods for topology generation. In order to generate large *connected* topologies, the number of links needed is large. Consequently, the diameter of the network decreases significantly and the complete set of pairs of nodes is covered within few hops (around 3 or 4 hops).

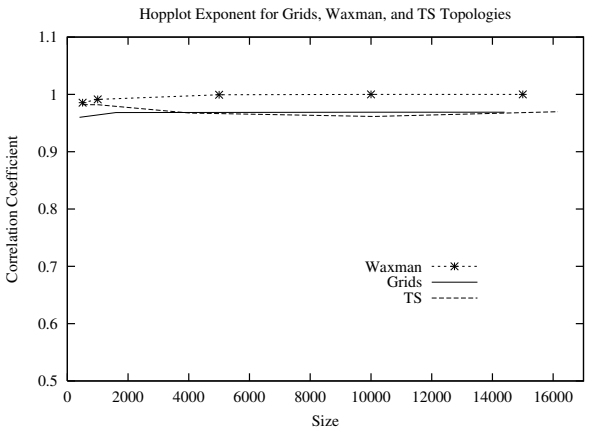

Figure 10: Correlation Coefficient vs. Size (Hopplot Exponent for Waxman, Transit-Stub and grid topologies)

Transit-Stub and grid topologies also yield correlation coefficients that are above 0.95. Thus, the hopplot exponent holds for different kinds of topologies, although they belong to three different classes (random, hierarchical and regular). As mentioned in Section 8, the correlation coefficients act as an initial test for assessing the resemblance to Internet topologies. Once this initial test is passed, we must look at the slopes of the linear fits to determine their closeness to those empirically found in [6]. We discuss this further in Subsection 9.3.3.

9.3.2 BRITE Topologies

Figure 11 shows the results for BRITE topologies. All generated topologies show the presence of the hopplot power law.

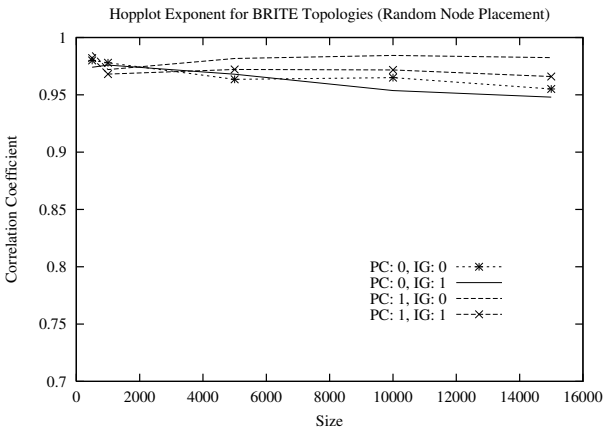

Figure 11: Hopplot Exponent for BRITE topologies

It is important to notice that when the incremental growth property is turned off, the generated topology "behaves" somewhat like a randomly generated topology with respect to a higher number of links needed to make the topology connected and, consequently, the generated topology has a smaller diameter (like Waxman topologies). Having a small diameter means that the log-log plot has few data points, which makes it easier to obtain a good linear fit. On the other hand, when incremental growth is present, the number of points in the plots increases significantly. However, the hopplot power law holds in both cases with correlation coefficients above 0.95.

9.3.3 Conclusions on Hopplot Exponent

Figure 12 shows a plot of the slopes of the linear fits for the hopplot exponent data for all kinds of topologies and for varying topology size.

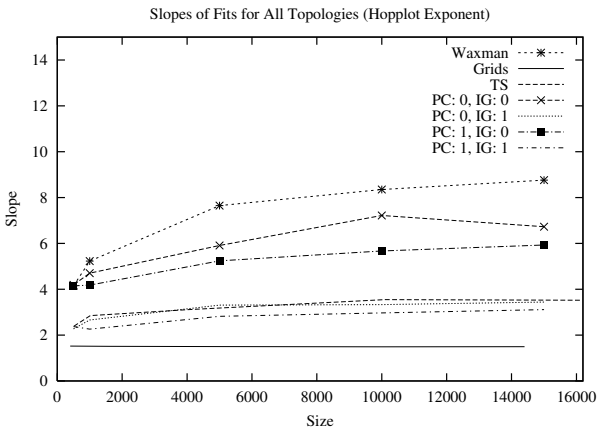

Figure 12: Linear Fit Slopes for Hopplot Exponent data (All topology classes)

Waxman topologies are generated with the highest degree of randomness, while grids are the most regular (no randomness at all). Thus, intuitively, the slopes for other kinds of topologies (including Transit-Stub) should lie between those of Waxman and grid topologies. This agrees with the results of Figure 12, where the slopes for Waxman topologies are the largest and they increase as topologies grow larger. For a 5000-node Waxman topology, the value of the slope is 7.65. At the other end of the spectrum, the slopes for the grid topologies were constant across all sizes with values always

around 1.5. These slopes are significantly different from the one measured in [6] for the router-level topology, which was 2.8.

BRITE topologies without preferential connectivity nor incremental growth show, as expected, slopes similar to that of Waxman topologies. The slopes for those topologies lie between 4.19 for 500-node network and 6.73 for 15000-node network. BRITE topologies with preferential connectivity but no incremental growth have less randomness and, as expected, show lower slope values. Finally, we observe that BRITE topologies with incremental growth alone or together with preferential connectivity behave very similarly to Transit-Stub topologies. Also, the values of the slopes lie between 2.39 and 3.52 for Transit-Stub topologies, and between 2.37 and 3.11 for BRITE topologies with preferential connectivity and incremental growth. Thus, BRITE topologies have slopes closer to the value of 2.8 measured in [6]. In summary, with respect to the hopplot exponent, both Transit-Stub and BRITE topologies with incremental growth resemble Internet topologies.

9.4 Eigenvalue Exponent

For different topologies, we plot for varying topology sizes the correlation coefficient of the log-log relationship of the eigenvalues of the topology adjacency matrix versus their rank in decreasing order (cf. Section 3).

9.4.1 Waxman, Transit-Stub and Grid Topologies

Figure 13 shows the results for Waxman, Transit-Stub and grid topologies. The eigenvalue power law tends to hold for Waxman topologies as they grow large. Similar to large Waxman topologies, all generated Transit-Stub topologies show correlation coefficients above 0.95, indicating good linear fits. Grid topologies show relatively low correlation coefficients.

Similar to the hopplot exponent, this result indicates that it is difficult to distinguish different topologies in terms of the eigenvalue exponent. We thus examine the slopes of the linear fits in Subsection 9.4.3.

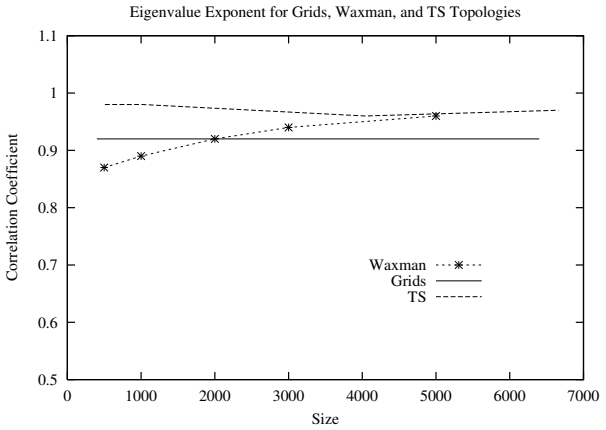

Figure 13: Correlation Coefficient vs. Size (Eigenvalue Exponent for Waxman, Transit-Stub and grid topologies)

9.4.2 BRITE Topologies

Figure 14 shows the results for BRITE topologies. The correlation coefficients are above 0.95 for nearly all topologies. Topologies generated with preferential connectivity but without incremental growth have clearly inferior correlation coefficients. It is worth mentioning that the value of the first (dominant) eigenvalue relative to that of other eigenvalues largely determines the linearity of the log-log plot. Since different kinds of topologies exhibit the eigenvalue power law (at least for certain network sizes), we examine next the slopes of the linear fits as a means to distinguish different classes of topologies.

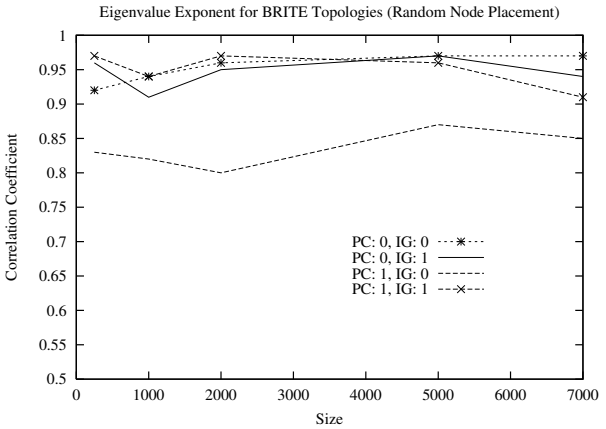

Figure 14: Eigenvalue Exponent for BRITE topologies

9.4.3 Conclusions on Eigenvalues

Figure 15 shows a plot of the slopes of the linear fits versus topology size for all considered topologies. For Waxman topologies, in which there is a high degree of randomness, the slope of the linear fits ranges from -0.21 for 500-node network to -0.65 for 5000-node network with correlation coefficients close to 0.9 in all cases. In contrast, for grid topologies, where there is no randomness at all, the eigenvalue log-log plots are flatter as the size of the grid increases. The slope of the fits for the grids ranges from -0.031 for 400-node grid to -0.002 for 6400-node grid. Thus, for both kinds of topologies, there is a clear difference in the value of the exponents/slopes. For these two classes of topologies, the slope values differ significantly from the value of -0.17 measured in [6] for a router-level topology.

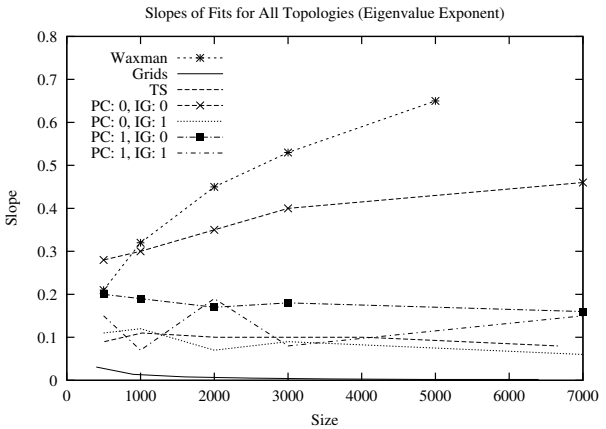

Figure 15: Linear Fit Slopes for Eigenvalue Exponent data (All topology classes)

Again, BRITE topologies generated without preferential connectivity nor incremental growth behave similar to Waxman topologies. The other four topology classes, Transit-Stub, BRITE with incremental growth only, BRITE with preferential connectivity only, and BRITE with both preferential connectivity and incremental growth,

lie between the extremes of complete randomness (Waxman) and complete regularity (grids). For these classes of topologies the slope values are not significantly different from the value of -0.17 reported in [6]. BRITE topologies with preferential connectivity show slopes closest to -0.17.

9.5 Path Length and Clustering Coefficient

In addition to the power laws, we also studied the diameter, the average path length and the clustering coefficient, defined earlier in Section 8, for all generated topologies.

Figure 16 shows the average path length for topologies of different sizes. We observe that Waxman (random) topologies and those BRITE topologies in which incremental growth was turned off, "behave" similarly in terms of the average path length. On the other hand, those topologies generated with incremental growth turned on are less random, resulting in a much higher average path length. This is consistent with results in [11].

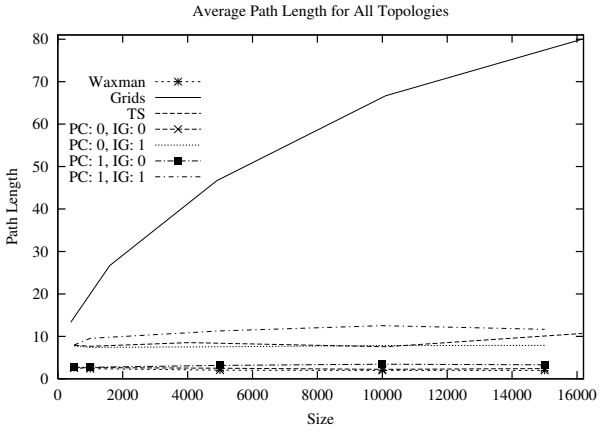

Figure 16: Average Path lengths for all topology classes

We observe that topologies generated with both preferential connectivity and incremental growth activated, have an average path length approaching 12 hops as the network size increases. In contrast, other topologies show lower average path lengths.

Figure 17 shows the clustering coefficients for all topology classes. Figure 18 shows the diameters. The diameter plot shows the same trends as the average path length. BRITE topologies generated with both preferential connectivity and incremental growth activated, show a diameter approaching 30 hops as the network size increases.[10]

Finally, we note that for Waxman topologies and BRITE topologies without incremental growth, as the network size increases, the diameter and average path length tend to decrease. This is because in a random network, as the number of nodes increases, the number of links needed to have a connected topology increases significantly, making "shortcuts" between many nodes and consequently decreasing the diameter and average path length of the network.

10. CONCLUSIONS

We studied possible causes for power laws [6] in Internet topologies. The outdegree and rank exponents are found to provide a powerful means to test the resemblance of a topology to that of the Internet. The hopplot and eigen exponents are found to hold for almost every topology we studied. However, the value of the exponent of the power law can provide a useful means to evaluate a topology. We also studied other metrics besides power laws. Topologies that strike a good balance between randomness and structure are found to have metrics close to those measured for the Internet. Consistent with other studies, we found that both preferential connectivity and incremental growth are the key contributors to the success of a topology to resemble that of the Internet.

Future work remains to build an even more realistic topology generator, which generates truly representative Internet topologies. This would include the assignment of capacities and propagation delays to links, the introduction of queues and the grouping of nodes into administrative domains. Further work also involves the study of other topological properties, how such properties may affect the design of Internet protocols, and refining and extending our topology generator.

[10] As a sanity check, Paxson [8] measures a mean Internet path length of around 16 hops and diameter of beyond 30 hops.

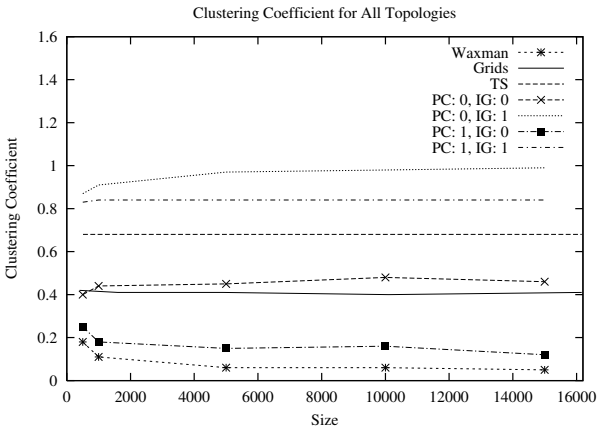

Figure 17: Clustering Coefficients for all topology classes

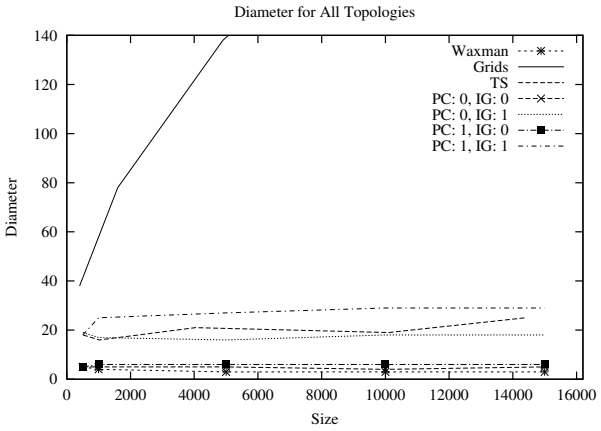

Figure 18: Diameters for all topology classes

11. ACKNOWLEDGMENTS

We would like to thank Mark Crovella for various discussions on aspects of this work and for his encouragement. Thanks also to the area editor and anonymous reviewers for their careful reading of the paper and their constructive comments.

12. REFERENCES

[1] R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the World-Wide Web. *Nature*, page 130, September 1999.

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

[3] K. Calvert, M. Doar, and E. Zegura. Modeling Internet Topology. *IEEE Transactions on Communications*, pages 160–163, December 1997.

[4] M. Crovella and A. Bestavros. Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes. *IEEE/ACM Transactions on Networking*, pages 835–846, December 1997.

[5] M. Crovella, M. Harchol-Balter, and C. Murta. Task Assignment in a Distributed System: Improving Performance by Unbalancing Load. In *Proceedings of ACM Sigmetrics '98 Conference on Measurement and Modeling of Computer Systems Poster Session*, Madison, WI, June 1998.

[6] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On Power-Law Relationships of the Internet Topology. In *ACM SIGCOMM*, Cambridge, MA, September 1999.

[7] B. A. Huberman and L. A. Adamic. Growth Dynamics of the World-Wide Web. *Nature*, page 131, September 1999.

[8] V. Paxson. End-to-End Routing Behavior in the Internet. *IEEE/ACM Transactions on Networking*, pages 601–615, December 1998.

[9] V. Paxson and S. Floyd. Why We Don't Know How To Simulate The Internet. In *Proceedings of the 1997 Winter Simulation Conference*, Atlanta, GA, January 1997.

[10] G. Phillips and S. Shenker. Scaling of Multicast Trees: Comments on the Chuang-Sirbu Scaling Law. In *ACM SIGCOMM '99*, Cambridge, MA, August 1999.

[11] D. J. Watts and S. H. Strogatz. Collective Dynamics of 'Small-World' Networks. *Nature*, pages 440–442, June 1998.

[12] B. Waxman. Routing of Multipoint Connections. *IEEE J. Select. Areas Commun.*, SAC-6(9):1617–1622, December 1988.

[13] S. Wolfram. *Mathematica: A System for Doing Mathematics by Computer*. Addison-Wesley, 1991.

[14] E. W. Zegura, K. Calvert, and M. J. Donahoo. A Quantitative Comparison of Graph-based Models for Internetworks. *IEEE/ACM Transactions on Networking*, pages 770–783, December 1997.