Small-World Characteristics of the Internet and Multicast Scaling

Shudong Jin
Computer Science Department
Boston University, Boston, MA 02115
jins@cs.bu.edu

Azer Bestavros
Computer Science Department
Boston University, Boston, MA 02115
best@cs.bu.edu

Abstract

Recent work has shown that the physical connectivity of the Internet exhibits small-world behavior. Characterizing such behavior is important not only for generating realistic Internet topologies, but also for the proper evaluation of network algorithms and protocols. Along this line, this paper tries to answer how small-world behavior arises in the Internet topologies and how it impacts the performance of multicast techniques. First, we attribute small-world behavior to two possible causes: the variability of vertex degree and the preference of vertices to have local connections. We found that both factors contribute with different relative degrees to the small-world behavior of the AS-level and router-level Internet topologies. For the AS-level topology, we have observed that extremely high variability of vertex degree is sufficient to cause small-world behavior, but for the router-level topology, preference for local connectivity plays a more important role. Second, we propose simple models to generate more realistic small-world Internet topologies. Our models consider both causes of small-world behavior. Third, we demonstrate the significance of our work by studying the scaling behavior of IP multicast tree size. We show that if topology generators capture only the variability of vertex degree, they are likely to underestimate the efficacy of multicast techniques.

1. Introduction

Because of its phenomenal growth in size, scope, and complexity, as well as its increasingly central role in society, the Internet has become an important object of study and evaluation. Moreover, as possibly the most complex and largest artifact of human engineering that was not deliberately designed, the Internet must be approached very much like a natural or physical phenomenon, whose emergent properties can not be understood by simple composition of well-understood behaviors. It is for these reasons that the last few years have witnessed a surge in research that attempts to empirically identify invariants about the topological properties of the Internet.

Characterizing Internet topological properties, while interesting simply for the sake of discovery, is crucial for the evaluation of new protocols and design choices. Indeed, many significant innovations in the networking community in recent years have resulted from a more accurate understanding of the fundamental properties of that complex system. Not only is the characterization of Internet emergent properties important, but also explaining *how* and *why* these properties emerge is extremely valuable for many reasons. Such an understanding would allow us to build models that could be used to generate synthetic artifacts (*e.g.*, large graphs) that resemble the "real" Internet well. Such synthetic artifacts are necessary for the simulation and proper evaluation of various network algorithms and protocols.

This paper focuses on one aspect of Internet topology characterization, which attracted much attention in recent years—namely the prevalence of the small-world phenomenon in the Internet routing graphs. Using a careful analysis of real datasets, we investigate what cause the small-world phenomenon. We illustrate how and to what extent the small-world phenomenon is caused by the variability of vertex degree; how and to what extent such phenomenon is caused by the preference of vertices to have local connections. We show that both causes contribute with different relative degrees to the small-world behavior of the autonomous system (AS) level and router-level topologies. This finding provides a basis for promising approaches to the development of more accurate Internet topology generators beyond degree-only models such as the Barabási-Albert model [3, 2] and the power-law random graph model [1]. We describe a new model to generate Internet topologies that exhibit both skewed vertex degree distributions and the preference for local connectivity, and show the small-world behavior in the synthetic graphs.

To demonstrate the significance of our characterization, we study the scaling behavior of IP multicast. We show how multicast tree size increases with the group size, by simulating multicast in different network topologies, including random graphs with highly variable degree, our small-world graphs, and real Internet graphs. The results show that us-

ing degree-based topology models, it is likely to underestimate the efficacy of multicast techniques; but using our small-world model, the scaling behavior is more accurately envisioned.

The paper is organized as follows. Section 2 reviews recent work on topology characterization. In Section 3, we describe the Internet routing graphs under study, and illustrate two possible causes of their small-world phenomenon. We describe and evaluate our new model to generate small-world topologies in Section 4. In Section 5, we use simulation to study the scaling behavior of multicast. Our findings and conclusions are summarized in Section 6.

2. Related Work

Recent work on Internet topology characterization has been related to one feature distinct from the early Erdős-Rényi random graph model [9]: highly variable vertex degree. In such networks, vertices have a non-uniform probability of being connected to others, with some vertices having extremely large number of neighbors. The degree distributions were often observed to follow approximately a power-law [10].

Power-law networks are particularly emphasized by the work of Barabási and Albert [3, 2] who explored a promising class of models that yield strict power-law vertex degree distributions. In their model (called the B-A model), three generic mechanisms are defined: (1) Incremental growth, which follows from the observation that networks develop by adding new vertices or new edges. (2) Preferential connectivity, which relies on an observation that highly popular vertices are more likely to be connected again in the process of incremental growth, *i.e.* a so called "rich-get-richer" phenomenon. (3) Re-wiring, which removes some links randomly and re-wires them according to the preferential connectivity mechanism. The combined use of these mechanisms drives the evolution of the network to a steady state, in which the vertex degree distribution follows a power-law (so called a scale-free property).

There are debates on the ability of the B-A model to explain the evolution of the Internet. First, the mechanisms of the B-A model are found to be inconsistent with observations from the real Internet growth. For example, preferential connectivity was shown to be stronger in the AS-level graph growth and re-wiring was shown to be an insignificant factor [6]. Second, the strict power-law vertex degree distributions of the B-A model can not be confirmed [6]. Indeed, Internet object sizes are better captured by other distributions such as Weibull distribution [4]. This would imply that the high variability of vertex degrees in the AS-level graphs may be the result of mechanisms [19] other than those in the B-A model. A more recent study [15] revealed that even the observed power-law degree distributions can be the result of sampling bias in traceroute-based measurement.

One interesting work [5] made some connections between power-law networks and small-world graphs [17, 23]. Small-world graphs exhibit connectivity properties that are between random and regular graphs (*e.g.*, regular lattices). Like regular graphs, they are highly clustered; yet like random graphs, they have typically short distances between arbitrary pairs of vertices. It has been shown that many networks have similar small-world property. In [5], the authors observed that most Internet topology generators capture power-law vertex degree distributions well, but usually do not do as well in capturing the clustering phenomenon exhibited in the Internet topologies. They proposed a variant of the B-A model by using a different preferential connectivity mechanism. Nevertheless, this variation of the B-A model is still degree-based. It obtains stronger clustering by generating a more skewed vertex degree distribution. No other causes of clustering and small-world behavior were identified. Moreover, their new model targets the AS-level Internet topology, but not the router-level topology.

In addition to the topology generators [16, 5] inspired by the B-A model, there are other generators that have been proposed and used to model the Internet. The Waxman model [24] extends the classical Erdős-Rényi model by randomly distributing vertices on a plane and creating edges by considering the distance between the vertices. The widely used GT-ITM Internet-specific topology generator [27] takes a hierarchical approach, capable of creating large graphs by composing smaller random graphs. Inet [12] assigns degrees to the vertices, following a power-law distribution, and then uses a linear preferential model to realize the assigned vertex degrees. The power-law random graph model in [1] generates degree distributions that strictly follow a power-law. A recent study [20] compared degree-based topology generators such as Inet and structural topology generators such as GT-ITM.

3. Characterizing Small-World Internet

This section describes the AS-level and router-level graphs used in our characterization, demonstrates their small-world behavior, and reasons how such behavior arises.

3.1. Evidence of Small-world Behavior

Internet Graphs
Table 1 summarizes the AS-level graphs and router-level graphs used in this paper. The first two are AS-level Internet graphs. One was obtained from the routing tables at route-views.oregon-ix.net. Since 1997, the routing tables have been collected once a day by the National Laboratory for Applied Network Research (NLANR) [18]. The graph we are using is dated May 26, 2001. Hereafter, we call it the "AS-2001" graph. It was found that the Oregon route-views is incomplete [6]. We obtained the second AS-level

graph from [21], which incorporates not only the Oregon route-views, but also the Looking glass data and the Routing registry data. This graph is also dated May 26, 2001. Hereafter, we call it the "AS-2001+" graph. Although AS-2001+ has only a few more vertices, it has much more edges (it has about 40% more edges. Therefore, it has higher average vertex degree).

We use four router-level graphs. The first was obtained from traceroutes collected by the Internet Mapping project at Lucent Bell Laboratories around November 1999. Hereafter, this graph is called the "Lucent" graph. The second router-level graph was obtained by merging the Lucent graph and the SCAN graph obtained around October/November 1999 using the Mercator software [11]. Hereafter, this graph is called the "Scan+Lucent" graph. In preprocessing both graphs, we discarded a few edges with undefined vertices. The percentage of these edges was negligible. These two graphs are available at [22].

The other two recent router-level graphs were obtained from [8]. They were measured using the Skitter tool. We obtained a snapshot on October 15, 2002, recorded by 13 of the 19 Skitter monitors worldwide. We merged the corresponding 13 subgraphs and obtained the "Skitter" graph. In addition, we also use a subgraph recorded by the monitor located at Palo Alto, CA. This subgraph is called the "Skitter-CA" graph. The use of this subgraph is intent to indicate that the observed small-world invariants are not due to the incompleteness of the graphs.

Is the Internet A Power-law Network?

Figure 1 shows the complementary cumulative distribution function (CCDF) $1 - F(d)$ of vertex degrees for the six graphs under study. Here $F(d)$ is defined as the probability that a vertex has degree not higher than d. The CCDF quantifies the probability that a vertex has a degree d larger than a certain value. A common property of these graphs is their long tails. That is, vertex degree is highly variable. The level of variability appears to be different though.

Previous work [10] has shown that vertex degree distributions follow a power-law. This is best illustrated by the AS-2001 graph. With a power-law distribution, $1 - F(d) = cd^{-\alpha}$, where c and α are constants, and the log-log scale plot of CCDF is a straight line, as shown in Figure 1(a). Using a linear regression, we estimated that for the AS-2001 graph, the exponent α is close to 1.22. For the AS-2001+ graph in Figure 1(b), an observation is that the distribution follows a power-law less well [6].

For the router-level graphs, we have also estimated their power-law exponents. However, as evident in Figure 1(c-f), it appears that the power-law distribution for the router-level graphs does not fit our empirical datasets. This is particularly the case for the Lucent graph and for the Skitter graphs, for which the tails of the distributions drop faster than any
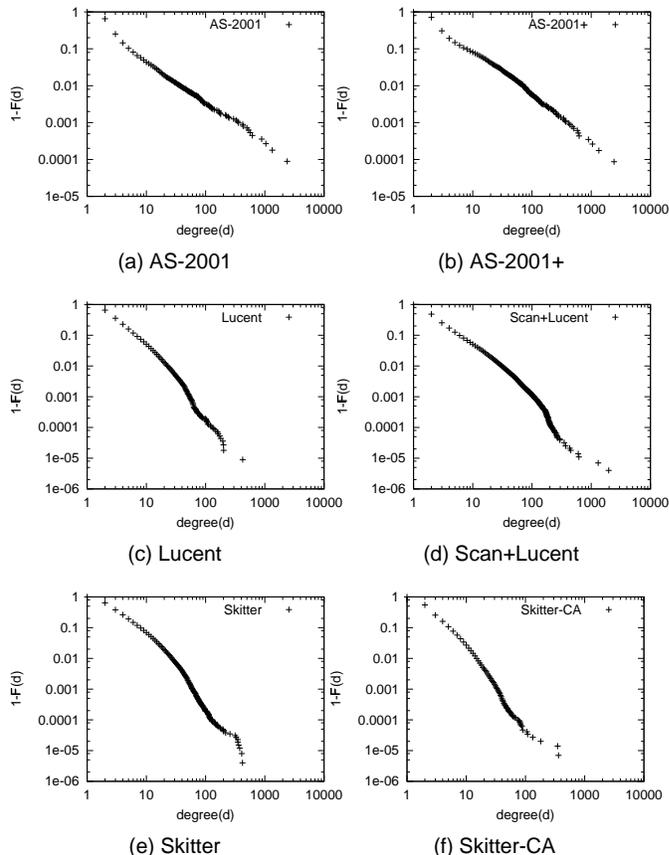
Figure 1. Vertex degree distributions of the AS-level and router-level Internet graphs.

power-law. In [4], Weibull distribution was found to provide a good fit to many Internet object size distributions. Its tail takes on the form $e^{-(x/\eta)^{\beta}}$, where η is the scale parameter and β is the shape parameter. Using rank regression on Y-axis, we estimated that the Weibull fits these empirical distributions better, though not perfectly.

Isn't the Internet A Small-world Graph?

In an influential paper [23], Watts and Strogatz defined a range of graphs termed *small-world graphs*. Small-world graphs are highly clustered, like regular graphs such as lattices; yet they have typically short distances between arbitrary pairs of vertices, like random graphs. The structural properties of these graphs are quantified by two metrics: the characteristic path length L and the clustering coefficient C. As in [23], we define

Definition 1 *Characteristic path length L is the number of edges in the shortest path between two vertices, averaged over all pairs of vertices.*

Definition 2 *Clustering coefficient C is defined as follows.*

Table 1. Internet graphs used in this study.

Graphs	AS-2001	AS-2001+	Lucent	Scan+Lucent	Skitter	Skitter-CA
Number of vertices	11174	11461	112669	282672	258329	145067
Mean degree	4.19	5.71	3.21	3.15	3.39	2.44
Maximum degree	2389	2432	423	1973	412	359
Fit distribution	Power law $(\alpha \approx 1.22)$	Power law $(\alpha \approx 1.22)$	Weibull $(\beta \approx 0.49)$	Power law $(\alpha \approx 1.7)$	Weibull $(\beta \approx 0.48)$	Weibull $(\beta \approx 0.48)$

Consider a vertex v which has k_v neighbors. There are at most $k_v(k_v - 1)/2$ edges among these k_v neighbors. Let C_v denote the fraction of these edges that actually exist. C is the average of C_v over all vertices with degree at least two.[1]

Following the original definition in [23], a small-world graph has two properties: (1) its L is not much larger than L_{random}, the characteristic path length of a random graph with the same number of vertices and edges, and (2) its C is much larger than C_{random}, the clustering coefficient of a random graph.[2] It is not difficult to see that for a connected random graph with N vertices and with an average degree of k, $L_{\text{random}} \sim \ln N / \ln k$ and $C_{\text{random}} \sim k/N$.

To show that the small-world phenomenon holds for both the AS-level and router-level Internet graphs, we computed their characteristic path lengths and clustering coefficients. Results are shown in Table 2. For comparison, we have also generated corresponding random graphs with approximately the same number of vertices and edges, and computed L_{random} and C_{random} by averaging over a number of random graph instances. The random graphs can be disconnected. In such cases, the largest connected components are used. Table 2 shows that the values of L for the Internet graphs are not much larger than L_{random}. Indeed, in most instances, L is smaller than L_{random}. Table 2 also shows that C is larger than C_{random} by 3-to-5 orders of magnitude (Note that this significant difference may not be the artifact due to the incompleteness of the Internet graphs under study). These two observations provide clear evidence of the small-world phenomenon in the Internet topologies.

In [5], AS-level graphs were found to exhibit small-world behavior. Our findings complement their work by showing that the router-level graphs are also small-world.

3.2. Cause One: High Variability of Vertex Degree

One question is whether the variability of vertex degree can introduce small-world behavior, given that all Internet

---

[1]Since C_v is undefined when $k_v = 1$, this averaging excludes vertices with only one neighbor. If a graph has many vertices with degree one (especially when vertex degree is highly skewed), then all of them are ignored. We believe it is a limitation of the original metrics.

[2]In the original definition, it is not clear what is *not much larger* L value and what is *much larger* C value. We consider L not much larger if it is only a fraction larger than that of a random graph. We consider C much larger if it is orders of magnitude larger than that of a random graph.

graphs under study have highly variable degree.

To study the effect of high variability of vertex degree distributions on small-world behavior, we generated random graphs whose vertex degree distribution follows a power-law. For such a distribution, the CCDF is defined as $1 - F(d) = cd^{-\alpha}$, where c is a constant. We use a random matching of the degrees to generate graphs. In addition, if the resulting graph is disconnected, its largest connected component is used. We realized that there can be different random matching algorithms such as the one in [1]. Nevertheless, even if we use different random matching algorithms, the results are still very similar.

We generated graphs with about 10000 vertices and 100000 vertices, respectively. The average vertex degree is fixed at 4.2. The value of α varies over a wide range. The constant factor c is determined such that the average vertex degree is roughly equal to 4.2. We ensured that neither the number of vertices nor the average vertex degree of the generated graphs departs from their target values by more than 2%. We computed the characteristic path length and clustering coefficient of these graphs, and plotted them in Figure 2. Each point represents one graph instance.

Figure 2 indicates that smaller α values are consistent with shorter characteristic path lengths and much larger clustering coefficients. Note that a smaller α means higher variability of vertex degree distribution. The presence of both short characteristic path length and high clustering coefficient is the signature of small-world graphs. Thus, we conjecture that a highly skewed power-law vertex degree distribution is a possible cause of small-world behavior.

It has been often observed that vertex degree distributions do not fit power-law distributions well [4, 6, 19]. Nevertheless, we found that as long as the vertex degree exhibits high variability, other distributions can also give rise to small-world behavior. To show this, we have generated random graphs whose vertex degree follows Weibull distribution. The CCDF of Weibull distribution is $e^{-(x/\eta)^\beta}$. The value of β varies from 0.2 to 2.0, and the value of η is determined such that the average vertex degree is still close to 4.2. For each generated graph, we computed its L value and C value. The results are plotted in Figure 3. Notice that the results we obtained here using Weibull distribution are similar to those we obtained using power-law distribution. Moreover, a smaller value of β for Weibull distribution (*i.e.*,

Table 2. Characteristic path length and clustering coefficient of the Internet graphs.

Graphs	AS-2001	AS-2001+	Lucent	Scan+Lucent	Skitter	Skitter-CA
L	3.637	3.535	10.02	8.803	11.36	16.26
C	0.432	0.4943	0.1001	0.0996	0.0233	0.0109
L_{random}	6.797	5.547	10.49	11.38	10.12	15.73
C_{random}	0.0003	0.0003	0.000022	0.000007	0.000014	0.000012

(a) Characteristic path length   (b) Clustering coefficient
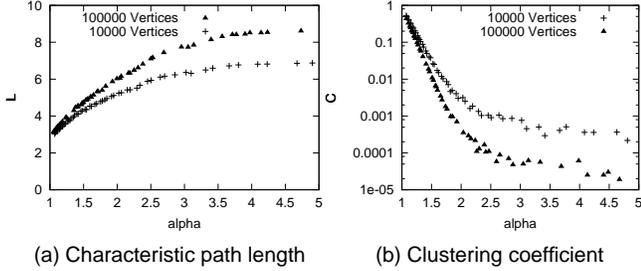
Figure 2. Small-world behavior as the result of power-law vertex degree distribution.

(a) Characteristic path length   (b) Clustering coefficient
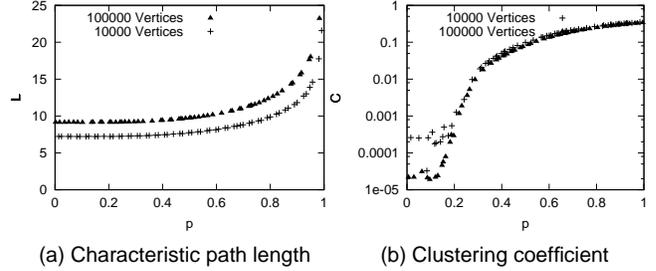
Figure 4. Small-world behavior as the result of preference for local connectivity.

(a) Characteristic path length   (b) Clustering coefficient
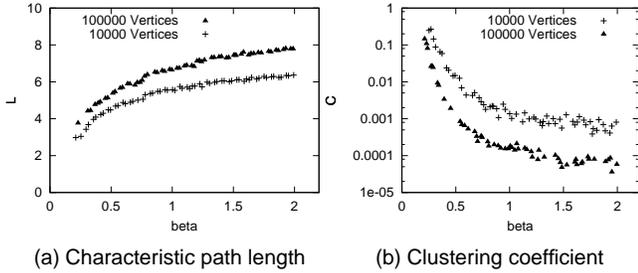
Figure 3. Small-world behavior as the result of Weibull vertex degree distribution.

a heavier tail and higher variability of vertex degree) results in smaller L and in much larger C. Thus, we conjecture that, high variability of vertex degree distributions, whether it is power-law or other distributions, can cause small-world behavior.

3.3. Cause Two: Preference for Local Connectivity

So far we have shown that the high variability of degree results in small-world behavior. But, are there other causes? To answer this question, we first conducted the following experiment. We generated graphs whose vertex degree distribution follows exactly the same distribution of the real Internet graphs. However, the edges were created randomly using a random matching. In this way, we preserved the high variability of vertex degree, but destroyed other topological properties that may exist in real Internet graphs. We call these synthetic graphs *randomized Internet graphs*.

For each real Internet graph, we generated a number of randomized instances, and for each we computed their L and C values, averaged over all randomized instances. A randomized instance is often disconnected. However, in our experiment, its largest connected component contained more than 90% of the vertices in the corresponding real graphs. The results are reported in Table 3. Comparing these results with those of the real graphs in Table 2 (the first two rows), we can make the following observations. First, for the AS-level graphs, the L values are very close to each other and the C values differ only by a small fraction (although the absolute differences are large). Second, for the router-level graphs, the L values are considerably different, and the C values differ by several orders of magnitude.

Our conclusions from this experiment are the follows. First, clearly there are other causes that contribute to the small-world behavior of the Internet topologies. Those other causes lead to larger clustering coefficient and longer path length. Second, it appears that those other causes are more pronounced when the variability of vertex degree distributions is not extremely high. In other words, when the variability of vertex degree distributions is extremely high, which is the case for the AS-level graphs, the effect of those other causes is overshadowed. When the variability of vertex degree distributions is moderate or low, which is the case for the router-level graphs, the effect of those other causes is evident.

So, what are these other causes of small-world behavior in the Internet graphs? In attempting to answer this question, we do not intend to provide a complete and exclusive

Table 3. Characteristic path length and clustering coefficient of randomized Internet graphs.

Graphs	Randomized AS-2001	Randomized AS-2001+	Randomized Lucent	Randomized Scan+Lucent	Randomized Skitter	Randomized Skitter-CA
L	3.406	3.346	6.944	5.971	6.704	8.748
C	0.2688	0.2507	0.00028	0.00076	0.00016	0.00007

explanation of small-world phenomenon, but to identify one plausible explanation—namely, the preference for local connectivity in the Internet. Indeed, this possible explanation was inspired by the work of Watts and Strogatz [23], who found that if only a portion of the edges of a regular lattice are reconnected randomly, but most edges are intact, then the resulting graph would exhibit small-world behavior. By doing so, the clustering coefficient remains large due to local connectivity, but the characteristic path becomes closer to that of random graphs due to remote connectivity.

To illustrate this possible cause of small-world phenomenon, we generated a set of 10000 and 100000 vertices, which we randomly placed on a two-dimensional plane. We set the average vertex degree to 4.2. The vertex distribution is the same as that of a random graph with the same number of vertices and edges. Specifically, the distribution has an exponentially decaying tail, *i.e.* low variability. Connections between vertices were made as follows. For each vertex v with degree k_v, on average v is connected to its pk_v nearest neighbors. The other edges of v are random. Here, $0 \leq p \leq 1$ is the probability of local connectivity, which we call *local probability*. We varied p to generate many graph instances and computed their L values and C values. The results are plotted in Figure 4, where each point represents one graph instance.

The results in Figure 4 reveal that preference for local connectivity leads to small-world graphs. First, the characteristic path length increases slowly when p is small or moderate. Second, the clustering coefficient increases drastically when p is small or moderate. In overall, there is a wide intermediate regime for p that yields the characteristic signature of small-world graphs.

3.4. Discussions

How does high variability of vertex degree distributions result in small-world behavior? With such high variability, it is likely that two interconnected vertices, say u and v, will have the same neighbor, say w. This occurs more frequently when w is a vertex with an extremely high degree, even though the edges are made random. An example is shown by Figure 5(a). It means that u, v, and w form a triangle. Such a pattern contributes directly to the computation of C_u, C_v, and C_w, and results in larger average clustering coefficient C, according to its definition. Intuitively, C grows with the variability of vertex degree. Also, notice that with
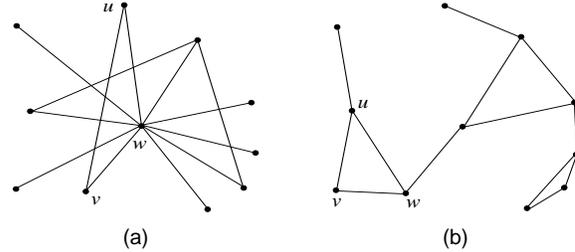
Figure 5. An illustration of strong clustering caused by (a) high variability of vertex degree, as vertices tend to have common neighbors (those with extremely high degrees) and hence tend to have large clustering coefficient, (b) preference for local connectivity, as nearby vertices tend to form triangles and hence tend to have large clustering coefficient.

highly variable vertex degree, the average distance between two vertices (L) is short. This is because the shortest path is usually through those extremely popular vertices. That is, highly popular vertices serve as good navigators in the graph.

How does preference for local connectivity result in small-world behavior? The answer to this question is straightforward. With a non-negligible probability of a local connection, if a vertex u is connected to v and w, then it is likely that v and w are also close to each other. See the example in Figure 5(b). As a result, there is a non-negligible probability that a triangle will form among these vertices, resulting in a higher clustering coefficient. Meanwhile, as long as there are still many long-range edges (short-cuts) in the graph, it is still easy to find a short path between two randomly chosen vertices. Those vertices with long-range edges serve as good navigators.

Both high variability of vertex degree distributions and the preference for local connectivity appear to be plausible causes of small-world behavior. In particular, the highly variable nature of vertex degree distributions is similar to the high-variable nature of many other Internet artifacts. Such distributions may be the result of some specific processes related to the evolution of these artifacts [3], or they may exist due to other reasons [19]. Preference for local connectivity may be explained for both the router-level and AS-level topologies as follows. At the router-level, links are created by considering the distances since there are tighter physical

constraints (there are certainly other relevant factors such as administrative considerations, which may cause hierarchy of the topology). At the AS-level, although the physical distance in a space is less important, there can be another kind of locality, for example Internet Service Providers (ISPs) may form cliques due to their business relationship.

To summarize our findings in this section:

- Higher variability of vertex degree distributions leads to shorter characteristic path length and larger clustering coefficient. Extremely high variability gives rise to small-world behavior, which is true for the AS-level topology.

- When the variability of vertex degree distributions is not high enough, it alone does not cause small-world behavior, which is true for the router-level topology.

- Preference for local connectivity is a possible cause of small-world behavior. This cause appears more pronounced for the router-level topology.

Our findings imply the failure of the Barabási-Albert model as an explanation of Internet growth and evolution [3]. Recently in [26], the same group recognized the limitations of their original model and combined distance dependence in a new hybrid model. Several other studies [6, 19, 4, 25] have also casted doubts on the adequacy of the Barabási-Albert model. However, none has examined the causes of small-world phenomenon as evidence. Specifically, the Barabási-Albert model targets power-law degree distributions. With only power-law degree distributions, the resulting graphs tend to have shorter characteristic path lengths and lower clustering coefficients. The comparisons between Table 2 and Table 3 provide the evidence. When the power-law exponent α departs much from unity, the Barabási-Albert model fails to generate small-world graphs at all. This result also calls for better models to generate more realistic Internet topologies.

4. Generating Small-world Internet

Based on our characterization, we have proposed better models to generate small-world Internet topologies. This section describes such a model. It not only captures variable vertex degree, but also captures the local preference cause of small-world behavior.

Previous studies [14, 26] have shown the probability of having an edge between two vertices can be explicitly expressed as a function of their distance. Let us consider the following model: the probability of having an edge between two vertices u and v is proportional to $l_{u,v}^{-r}$, where $l_{u,v}$ is the distance between u and v and r is a non-negative constant. It was originally used by Kleinberg [13] to generalize the Watts-Strogatz model. The key is the choice of r. If r is larger, $l_{u,v}^{-r}$ decays faster and the edge tends to be local;

Given a sequence of vertex degrees and the exponent of distance-dependence r, generate a graph with N vertices as follows:
  (1) Randomly place N vertices on a plane. A degree d_v is assigned to each vertex v, $1 \leq v \leq N$.
  (2) Create edges among the vertices. For each vertex v, in descending d_v order, repeat create edges (if not enough yet) as follows:
     – Choose u with probability proportional to $\frac{d_u}{l_{u,v}^r}$, where $l_{u,v}$ is the Euclidean distance between u and v, such that (i) $u \neq v$ and (ii) there is no edge (u, v) yet. Create edge (u, v).

Figure 6. A model for generating small-world graphs with variable degree.

(a) Characteristic path length   (b) Clustering coefficient
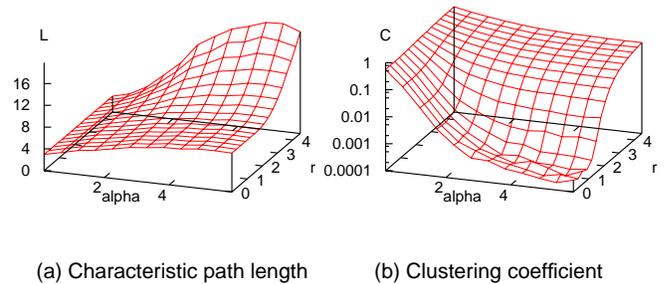
Figure 7. A range of small-world graphs generated by our small-world model.

but r can not be too large since otherwise the lack of long-range edges would lead to an excessively large L value. If r is smaller, $l_{u,v}^{-r}$ decays slowly and long-range edges become frequent; but r can not be too small since otherwise the graphs would be close to random (*i.e.*, the value of C would be excessively small). Using this model, Figure 6 describes an implementation.

This model produces graphs exhibiting a wide range of small-world behavior. For example, we generated graphs with 10000 vertices and with average degree 4.2. Vertex degree follows a power-law with exponent α. We varied parameters α from 1.05 to nearly 8.0 and r between 0 and 4.5. For each pair of α and r, we generated ten graphs, computed their average L value and C value, and plotted them in Figure 7. We observed that when α is small or r is moderate, both the characteristic path length and the clustering coefficient satisfy the requirements for small-world graphs.

We have also used this model to generate synthetic Internet graphs that closely resemble the real Internet graphs. We describe here our effort to use the described model to synthesize the Lucent router-level graph.

We first obtained the sequence of vertex degree from the real graph, and fed it into the graph generator. The other parameter r was set to 3.15. The output largest connected component has 111328 vertices and the mean degree is 3.24. We then computed its characteristic path length 9.814 and clustering coefficient 0.193. Comparing them to those of the real graph, we found that the path lengths are close but the clustering coefficients are still different. A possible explanation is as follows. The model assumes the probability of having an edge between two vertices u and v is proportional to $l_{u,v}^{-r}$. But the distance-dependence in the real graph may not follow a power-law well [14]. In other words, the power-law relationship may over-estimate local preference when $l_{u,v}$ is small. As a result, the model tends to produce higher clustering coefficient. In the next section, this graph is used in our simulation study of multicast scaling.

5. Implications on Multicast Scaling

In this section, we use simulation to study the scaling behavior of IP multicast, *i.e.*, how the multicast tree size increases with the group size. In our simulation, we used three topologies in comparison. The first is the original Lucent router-level graph in Table 1. The second is generated using the power-law random graph model in [1]. We found the generated largest connected component (after eliminating duplicated edges and self-loops) has 110986 vertices and mean degree 3.18. Notice that this graph preserves the original vertex degree distribution in the Lucent graph only. The third topology, as we described before, is generated using our small-world model in Figure 6. It has 111328 vertices and the mean degree is 3.24. These three graphs have approximately the same number of vertices and edges. However, their small-world behaviors are different. Our small-world graph resembles the real Lucent graph better.

To study the scaling behavior of IP multicast with shortest path tree, we proceeded as follows. We randomly chose a vertex as the server and n other vertices as clients in the graph. Then a shortest path tree is constructed. The size of this tree was computed and compared to the average distance from these clients to the server. The ratio is the normalized multicast tree size and it reflects how the tree size increases with n. For each value of n we repeated above simulation many times and computed the average multicast tree size. By varying n we plotted the scaling behavior of IP multicast tree. Figure 8 compares the results using the Lucent graph and using the degree-only graph. Figure 9 compares the results using the Lucent graph and using our small-world graph.

We observed that using the degree-only graph, IP multicast has worse scaling behavior than it using the real Lucent graph. On the contrary, using our small-world graph, the scaling behavior fits that using the real Lucent graph very well. It thus suggests the importance of capturing the small-

(a) Linear scale   (b) Log scale
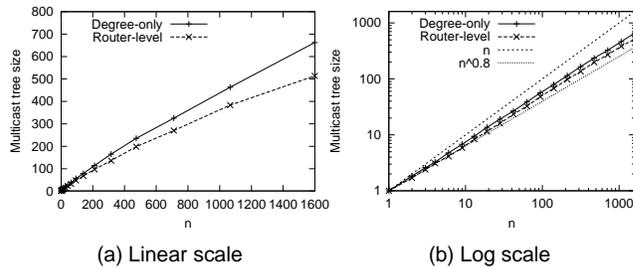
Figure 8. IP multicast scaling: degree-only topology vs Internet topology.

(a) Linear scale   (b) Log scale
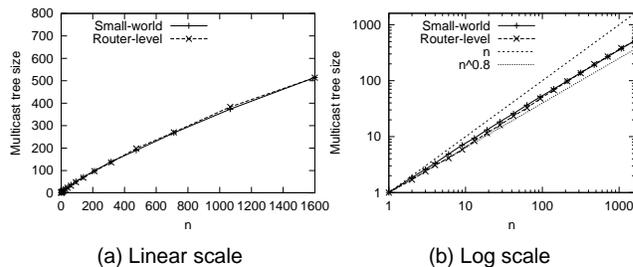
Figure 9. IP multicast scaling: small-world topology vs Internet topology.

world behavior in the Internet topologies. Without doing so, flawed topology generation models may underestimate the efficacy of IP multicast in reducing overall network cost.

We should also point out that accurate modeling of the Internet topology helps understand the so-called Chuang-Sirbu law [7] of IP multicast scaling. To better illustrate this point, we have also shown the log-scale plots of our simulation results in the figures. The Chuang-Sirbu claims that multicast tree size increases as $n^{0.8}$, a straight line in the log-log scale plots. From our simulation, we can make two observations. First, it appears multicast tree size increases faster than $n^{0.8}$, *i.e.*, the exponent is larger: we estimated the tree size scales as $n^{0.85}$. Second, we can see that with degree-only graphs, the scaling behavior is far from the Chuang-Sirbu law, but with small-world graph, the scaling behavior is closer to the law. It could suggest that, small-world behavior (coupled with the variable degree nature of the Internet topologies) may draw the true scaling behavior line toward, yet not close to, the Chuang-Sirbu law.

This simulation study strongly suggests that the scaling behavior of multicast depends on the underlying network topology. Previous work on IP multicast scaling does not capture the variable degree nature of the real Internet. Hence, the analysis does not apply here. We believe there are more intriguing properties of the Internet that alter the scaling behavior, such as clustering and hierarchy.

6. Conclusions

This paper has focused on the small-world aspect of the Internet topology. We illustrated the possible causes of such behavior and demonstrated its importance to evaluating the scaling behavior of multicast techniques. Our main findings and conclusions are:

- Two possible causes of small-world behavior in the Internet are (i) high variability of vertex degree and (ii) preference for local connectivity. Extremely high variability gives rise to the small-world behavior in the AS-level topology; but when the variability of vertex degree distributions is moderate, clustering in small-world graphs is mainly caused by local connectivity, which is true for the router-level topology.

- If Internet topology generators target vertex degree distributions only, they generate less realistic topologies. We proposed promising models that captures both vertex degree distributions and preference for local connectivity. By doing so, it is easy to generate more realistic small-world Internet topologies.

- IP Multicast tree size depends on the small-world behavior of the Internet topology. If Internet topology generators target variable vertex degree only, then it is likely to underestimate the efficacy of IP multicast.

References

[1] W. Aiello, F. R. K. Chung, and L. Lu. A random graph model for massive graphs. In *ACM Symposium on Theory of Computing (STOC)*, pages 171–180, 2000.

[2] R. Albert and A. Barabási. Topology of evolving networks: Local events and universality. *Physical Review Letters*, 85:5234–5237, 2000.

[3] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[4] A. Broido and kc claffy. Internet topology: Connectivity of IP graphs. In *Proceedings of SPIE ITCOM*, 2001.

[5] T. Bu and D. Towsley. On distinguishing between Internet power law topology generators. In *Proceedings of IEEE INFOCOM*, 2002.

[6] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. The origin of power laws in Internet topologies revisited. In *Proceedings of IEEE INFOCOM*, 2002.

[7] J. Chuang and M. Sirbu. Pricing multicast communications: A cost based approach. In *Proceedings of Internet Society INET*, 1998.

[8] Cooperative Association for Internet Data Analysis. The CAIDA Website. http://www.caida.org.

[9] P. Erdős and A. Rényi. The evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

[10] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. In *Proceedings of ACM SIGCOMM*, 1999.

[11] R. Govindan and H. Tangmunarunkit. Heuristics for Internet map discovery. In *Proceedings of IEEE INFOCOM*, 2000.

[12] C. Jin, Q. Chen, and S. Jamin. Inet: Internet topology generator. Technical Report CSE-TR-433-00, EECS Department, University of Michigan, 2000.

[13] J. Kleinberg. The small-world phenomenon: An algorithm perspective. In *ACM Symposium on Theory of Computing (STOC)*, pages 163–170, May 2000.

[14] A. Lakhina, J. Byers, M. Crovella, and I. Matta. On the geographic location of Internet resources. In *Proceedings of ACM SIGCOMM Internet Measurement Workshop (IMW)*, 2002.

[15] A. Lakhina, J. Byers, M. Crovella, and P. Xie. Sampling biases in IP topology measurements. In *Proceedings of IEEE INFOCOM*, 2003.

[16] A. Medina, I. Matta, and J. Byers. On the origin of power laws in Internet topologies. *ACM SIGCOMM Computer Communication Review*, April 2000.

[17] S. Milgram. The small world problem. *Psychology Today*, 2:60–67, 1967.

[18] National Lab for Applied Network Research. Global ISP interconnectivity by AS number. http://moat.nlanr.net/AS/.

[19] H. Tangmunarunkit, J. Doyle, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Does AS size determine degree in AS topology? *ACM SIGCOMM Computer Communication Review*, October 2001.

[20] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Network topology generators: Degree based vs. structural. In *Proceedings of ACM SIGCOMM*, 2002.

[21] Topology Project at University of Michigan. AS graph data sets. http://topology.eecs.umich.edu/data.html.

[22] USC Information Sciences Institute. Internet maps. http://www.isi.edu/div7/scan/mercator/maps.html.

[23] D. Watts and S. Strogatz. Collective dynamics of small-world networks. *Nature*, 363:202–204, 1998.

[24] B. M. Waxman. Routing of multipoint connections. *IEEE Journal on Selected Areas in Communications*, 6(9):1617–1622, 1988.

[25] W. Willinger, R. Govindan, S. Jamin, V. Paxson, and S. Shenker. Scaling phenomena in the Internet: Critically examining criticality. In *Proceedings of the National Academy of Sciences*, 2001.

[26] S. Yook, H. Jeong, and A. Barabási. Modeling the Internet's large-scale topology. *Proceedings of the National Academy of Sciences (submitted)*, 2002.

[27] E. W. Zegura, K. L. Calvert, and M. J. Donahoo. A quantitative comparison of graph-based models for Internet topology. *IEEE/ACM Transactions on Networking*, 5(6):770–783, 1997.