Inet-3.0: Internet Topology Generator

Jared Winick   Sugih Jamin
{*jwinick,jamin*}*@eecs.umich.edu*

CSE-TR-456-02

Abstract

In this report we present version 3.0 of Inet, an Autonomous System (AS) level Internet topology generator. Our understanding of the Internet topology is quickly evolving, and thus, our understanding of how synthetic topologies should be generated is changing too. We document our analysis of Inet-2.2, which highlighted two shortcommings in its topologies. Inet-3.0 improves upon Inet-2.2's two main weaknesses by creating topologies with more accurate degree distributions and minimum vertex covers as compared to Internet topologies. We also examine numerous other metrics to show that Inet-3.0 better approximates the actual Internet AS topology than does Inet-2.2. Inet-3.0's topologies still do not well represent the Internet in terms of maximum clique size and clustering coefficient. These related problems stress a need for a better understanding of Internet connectivity and will be addressed in future work.

NAME

inet - an AS-level Internet topology generator.

SYNOPSIS

`inet` `-n` N `[-d` k`] [-p` n`] [-s sd] [-f of]`

DESCRIPTION

The "Inet" generator generates an AS-level representation of the Internet. It is important to note that Inet only provides the connectivity information; the generated topologies do not have any information pertaining to latency, bandwidth etc. It generates random networks with characteristics similar to those of the Internet from November 1997 to Feb 2002, and beyond. The generator should be used to generate network of no less than 3037 nodes, which is the number of ASs on the Internet in November 1997. The software package with source code for Unix, can be found at:
`http://topology.eecs.umich.edu/inet/`

OPTIONS

-n N**:** the total number of nodes in the topology.
-d k**:** the fraction of degree-one nodes. Default is 0.3.
-p n**:** the size of the plane used for node placement. Default is 10,000.
-s sd: the seed to initialize the random number generator. Default is 0.
-f of: the debugging output file name. Default is `stderr`.

EXAMPLES

To generate a 6,000-node network with default values:

```
example% inet -n 6000 > Inet.6000
```

To generate a 6,000-node network on a 10,000 by 10,000 plane with 30% of total nodes as degree-one nodes, random seed of 16, and debug file `debug`:

```
example% inet -n 6000 -f debug -d .3 -s 16 -p 10000 > Inet.6000
```

OUTPUT FORMAT

The output of `inet` follows the following format:

```
nodes links
...
id x y
...
id1 id2 weight
```

The first line of the output specifies the number of nodes with `nodes`, and the number of undirected links with `links`. The next section contains the location of each node. Each line contains the node id, `id`, and the xy-coordinate, `x` and `y`. The last section contains the list of links in the topology. Each line contains the ids of the two nodes incident to the link, `id1` and `id2`, and the link weight, `weight` (reserved).

1 Overview

In [5], we compared Inet-2.2 to several other popular Internet topology generators. Using the properties of Internet AS topologies identified in [3], it was shown that Inet-2.2 performed as well as or better than other Internet topology generators. Therefore in this report, we forgo the comparison to other topology generators, and focus on a comparison between Inet-2.2, Inet-3.0, and Internet topologies. We first examine two properties of Inet-2.2 topologies that are not faithful to those found in Internet topologies. This analysis lead us to two simple, yet powerful, modifications that form the basis of our update to Inet. To verify Inet-3.0, we extend the metrics from [5] to include several others. We show that in comparison to Inet-2.2, Inet-3.0 generates topologies that are more faithful to Internet topologies. While we show that Inet-3.0 is more successful than Inet-2.2 in generating Internet-like topologies, we also highlight and discuss known discrepancies between Inet-3.0 and Internet topologies.

2 Internet Data

For our Internet topologies, we obtained raw BGP table information from The National Laboratory for Applied Network Research (NLANR) [8] and The University of Oregon Route Views Project [11]. Both of these sites collect BGP routing tables from the route server `route-views.oregon-ix.net`. From this raw BGP table data, we extracted connectivity information into a simple adjacency list of ASs. Our data set consists of 51 Internet topologies from November 1997 to February 2002.

3 Analysis of Inet-2.2 Topologies

Looking only at the characteristics analyzed in [5], we would think that Inet-2.2 topologies were almost identical to Internet topologies. It is clear though, that there are many properties that characterize a topology that were not evaluated in [5]. Unfortunately, neither the networking community, nor the graph theory community, have been able to find one unifying metric which fully describes a topology. Therefore we are forced to analyze as many metrics as possible, in an attempt to characterize and understand the Internet AS level topology. Our analysis focuses on two properties of a graph, one relating to degree distribution, and the other to connectivity. In terms of degree distribution, we look at the complementary cumulative distribution function (ccdf) $\bar{F}(d) = \sum_{i>d}^{\infty} f(d)$, which is the fraction of nodes with degree greater than d. With respect to connectivity, we examine the ratio of the minimum vertex cover of a graph to the size of the graph, $\mid VC \mid /n$. In the next two subsections, we analyze Inet-2.2 in terms of these metrics, and then describe the modifications in these areas that we made for Inet-3.0.

3.1 Degree Distribution - ccdf

In 1999, Faloutsos *et al.* [3] found several power laws relating to the topology of the Internet. One of these power laws was $f(d) \propto d^O$, that is, the frequency of nodes with degree d, is proportional to d raised to a power of a constant O. Unfortunately, this power law holds only if 1.5 to 2% of the highest degree nodes are removed from the data, as was noted in [5]. While this small percent may seem unsubstantial initially, this percentage of top degree nodes accounts for over 1/3 of the total degrees in the Internet. Due to this weakness of the Frequency-Outdegree power law, we instead look at the complementry cumulative distribution function $\bar{F}(d) = \sum_{i>d}^{\infty} f(d)$ which exhibits the power law

$$\bar{F}(d) \propto d^{\alpha} \tag{1}$$

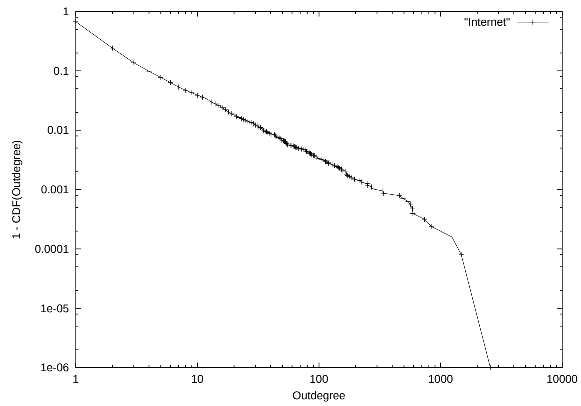


Figure 1: $\bar{F}(d)$ for all d. Top 3 nodes do not fit power law

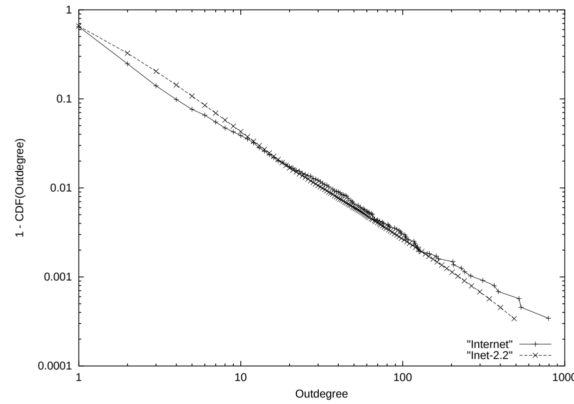


Figure 2: $\bar{F}(d)$ with top 3 nodes excluded.

This power law has been used by others [1, 2], as it describes the degrees of all but about the top 3 nodes as seen in Fig. 1. Fig. 2 shows $\bar{F}(d)$ for the Internet topology in February 2001, along with an equal sized Inet-2.2 topology.

While we see that Inet-2.2's degree distribution is close to that of the Internet, clearly there is room for improvement. As is described in [5], Inet-2.2 relies on exponential growth laws. As an extension of the power law $f(d) \propto d^O$, it is shown that $f(d) = e^{at+b}d^O$, where a,b, and O are know constants and t is the number of months since November 1997. To calculate a degree distribution for a given size topology N, Inet-2.2 simply uses the exponential growth rate of the number of ASs to calculate the number of months t it would take the Internet to grow from its size in November 1997 to N.[1] Recall that this power law can not be used to generate degrees for the top 2% of nodes. We would thus like to use Eq. 1 to generate degrees, as it can be used for all but the top 3 nodes. Eq. 1 would only be useful for this purpose if we could appropriately model it over time, as we did with the Frequency-Outdegree power law. For each of our Internet topologies from November 1997 to February 2002, we plotted $\bar{F}(d)$ for all degrees, except degree 1 nodes and the top 3 degrees. Using linear regression, we fit a line to each Internet data set as seen in Fig. 3. The slope of this line corresponds to α in $\bar{F}(d) \propto d^{\alpha}$. Fig. 4 shows the values of α over time. Fig. 5 shows the y-intercept over time.

Rather than a constant exponent and variable intercept as was seen over time for $f(d)$, we have a variable exponent and relatively constant intercept over time for $\bar{F}(d)$. The correlation coefficient for the exponents is 0.97, while the correlation coefficient for the intercepts is -0.65. This translates to having a relatively constant fraction of low degree nodes, with the degrees of the top nodes exponentially increasing. This leads us to a new exponential growth law:

$$\bar{F}(d) = e^c d^{at+b} \tag{2}$$

where c, a, and b are known constants and t is the number of months since November 1997.

In Inet-3.0, we use Eq. 2 to calculate the degree distribution for all but degree 1 nodes and the top 3 nodes. As seen in section 1, the fraction of degree 1 nodes is left up to the user. As the fraction of degree 1 ASs on the Internet has remained steady around 0.3, the default of this value is set to 0.3. We use the Outdegree-Rank exponential growth law defined in [5] to calculate the degrees of the top 3 nodes as is

[1] Internet growth rate has been less than exponential lately. However this does not completely invalidate the approach we adopted here. For a topology of a given size, Inet simply places it at an earlier time than it actually may be at. Since the "age" of the topology is neither a user controlled parameter nor an output of Inet, its value and use is transparent to the users.

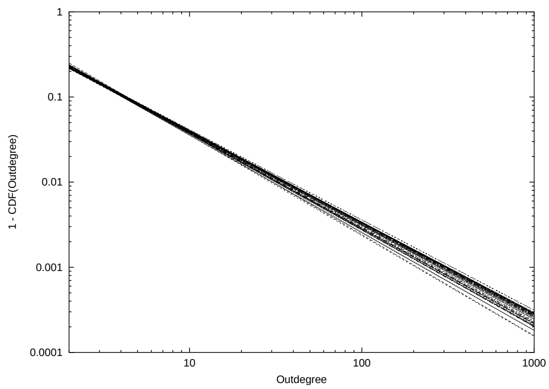


Figure 3: Lines fitting $\bar{F}(d)$ for each of 51 Internet topologies.

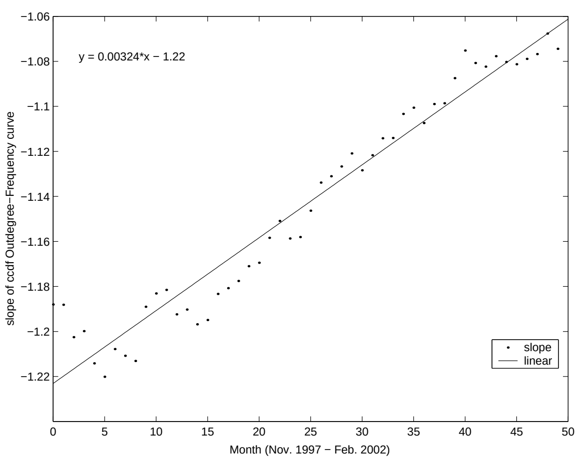


Figure 4: Relationship of ccdf power law exponent and time.

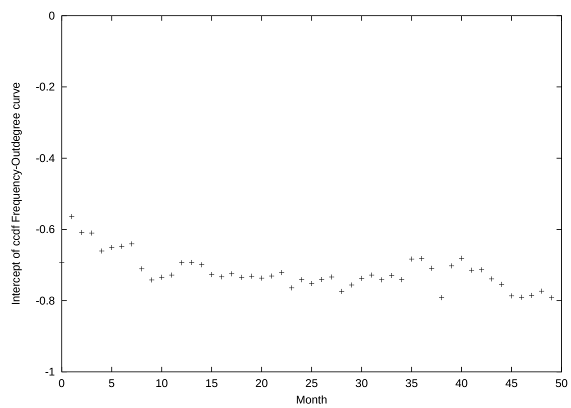


Figure 5: Relationship of ccdf power law intercept and time.

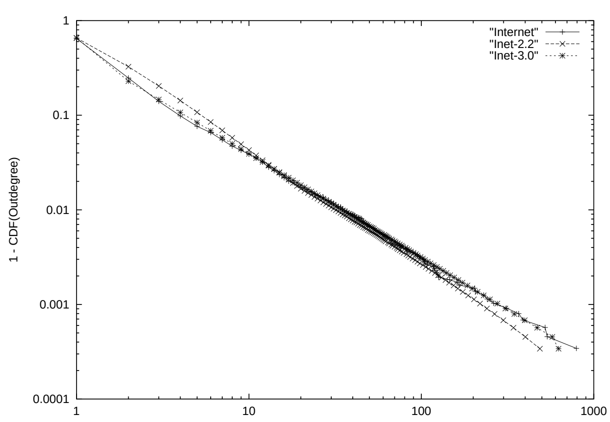


Figure 6: $\bar{F}(d)$ for February 2001 topologies.

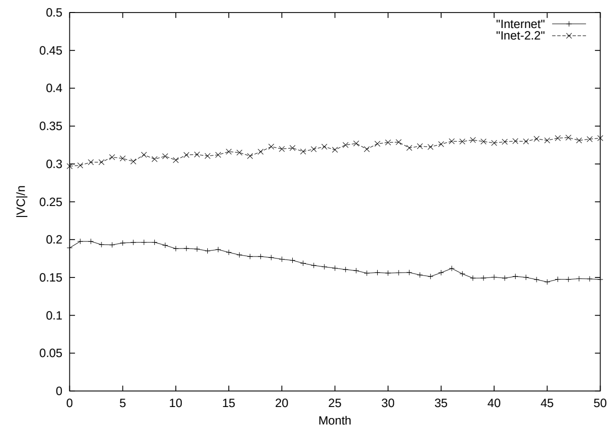


Figure 7: Vertex Cover of the Internet versus Inet-2.2 over time

done in Inet-2.2. In Fig. 6, we compare $\bar{F}(d)$ of a February 2001 Internet topology against Inet-3.0 and Inet-2.2 topologies of equal size.

As shown in Fig. 6, we see that Inet-3.0 matches the degree distribution of the Internet even more faithfully than does Inet-2.2. In Section 4 we compare the ccdf plots for two other size topologies, both showing similiar results.

3.2 Connectivity - vertex cover

Throughout this paper, we will refer to vertex cover as the ratio $\mid VC \mid /n$, where $\mid VC \mid$ is the number of nodes in the minimum vertex cover, and n is the number of nodes in the graph. This form allows us to compare graphs of different sizes. Finding the minimum vertex cover of a graph is an NP-hard problem. We use a well-known heuristic for approximating the minimum vertex cover of a graph. The heuristic is a greedy algorithm that builds a vertex cover by iteratively adding the node which covers the most remaining uncovered edges in the graph.

A discrepency between the vertex cover of Inet-2.2 topologies and Internet topologies was first noted by Park *et al.* [9]. They reported that the vertex cover of Inet-2.2 topologies was about 50% larger than the vertex cover of the Internet. This difference is clearly seen in Fig. 7 which plots the vertex cover of the Internet over the 51 months of our data set, versus Inet-2.2 topologies of equal sizes. Even more disturbing is that we see the difference in vertex cover values to be increasing over time to a difference of over 100% in early 2002.

As was seen above, the degree distribution of Inet-2.2 topologies is quite similiar to the degree distribution of the Internet, even before modifications made in Inet-3.0. For a set of nodes with specified degrees, there are many ways of connecting these nodes. This leads us to believe the discrepancy in vertex cover was a result of Inet-2.2's connectivity process. The process of connecting nodes in Inet-2.2 consists of three steps as follow:

1. Build a spanning tree with all nodes that have degrees greater than 1. Let G be the graph to be generated, initially empty. Randomly, with uniform probability, a node with degree larger than 1 that is not in G is connected to a node in G with linear preference, namely, the probability of connection to a node in G is d/D, where d is the degree of the node in G, and D is the sum of degrees of all nodes already in G that still have at least one free degree.

2. Next, we connect all nodes with degree 1 to nodes in G with linear preference.

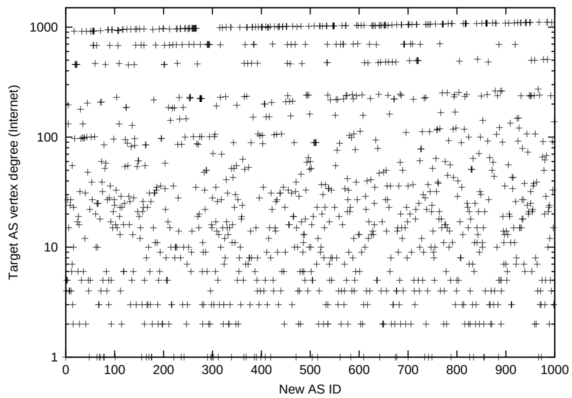


Figure 8: Actual target node's degree(s).

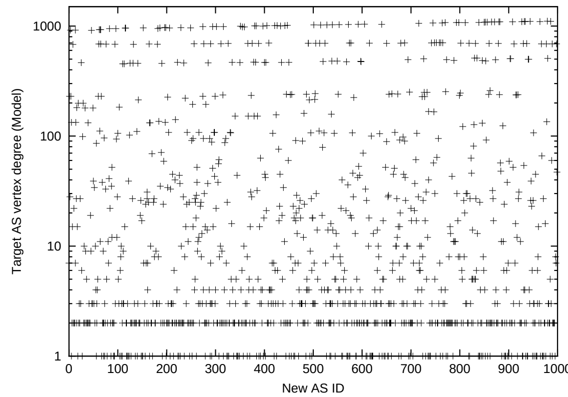


Figure 9: Target node's degrees(s) by the linear preferential model.

3. Finally, we connect the remaining free degrees in G, starting from the node with the largest degree first. In making these connections, we randomly pick nodes with free degrees using linear preference.

Before we attempted to analyze Inet-2.2's connectivity process, we first tested a few simple, intuitive *ad-hoc* modifications to the three steps listed above:

- It has been noted that the top 10-15 ASes on the Internet form a mesh, and that all other ASes are within an average of 3 hops of an AS in the mesh. Our modification consisted of first forming an initial mesh of the 10-15 nodes with the highest degrees. Then when building the spanning tree, attempt to insure that every node would be connected so that it was at most 3 hops away from at least one node in the mesh. This method, while feasible, showed to be ineffective in changing the value of the vertex cover.

- When connecting a node, attempt to connect it only to a node with higher degree. This proves to be infeasible in step 3 of the Inet connectivity process, as we describe below in more detail.

As none of these heuristics effected the vertex cover property of Inet-2.2 topologies, a somewhat less *ad-hoc* approach was needed. In step 3 of the Inet connectivity process, free degrees are filled in, starting with the node with the largest degree. This implies that when the free degrees of a node with degree d are to be filled in, there will be no nodes with degree greater than d that have free degrees. What this translated to was that many low degree nodes were forced to connect to other low degree nodes. In a practical sense, this seems to be a problem, because on the Internet, smaller stub-ASs are more likely to be connected to larger transist-ASs. In [2] the authors show that new ASs in the Internet have a greater preference to connect to high degree ASs than is predicted by linear preference. Fig. 9, taken from [2], shows that 1,000 nodes (x-axis) added to a snapshot of the Internet sequentially using linear preferentiale model shows a higher probability of connecting to low degree target nodes (y-axis) than exhibited in the historical data, Fig. 8. Under the linear preference model, despite the preference for high degree target nodes, the sheer number of low degree nodes (two thirds of the ASs on the Internet have degree ≤ 2), means that a large number of new nodes ended up connecting to a low degree node.

To quantify this intuitive understanding of Internet connectivity, we ask, for a node with a certain degree, what is the degree distribution of its neighbors? For example, what degree ASs does the highest degree node on the Internet connect to, as well as, what degree ASs do degree 2 ASs connect to? We plotted the cumulative degree distribution of a node's neighbors $F(d) = \sum_{i=1}^{d} f(d)$ where $f(d)$ is simply the frequency of neighbors with degree d. We plotted $F(d)$ for nodes with degrees 1, 2, 3, and 4, as well as for the top 4 ASs. Figs. 10 and 11 show these distributions for the Internet on February 01, 2001 versus an Inet-2.2 topology of equal size (8765 nodes).

We quickly observe signifigant deviations between the Internet and Inet-2.2 distributions. Our initial suspicion that low degree nodes are connecting to too many other low degree nodes in Inet-2.2, is firmly substantiated by these plots. For example, in this Inet-2.2 topology, almost 35% of degree 2 nodes' neighbors have degree less than or equal to 3. This is in sharp contrast to the Internet topology where only 5% of degree 2 nodes' neighbors have degree less than or equal to 3. Conversely, when we examine the distributions of the top degree nodes, we see that top degree nodes in the Inet-2.2 topology do not connect to as many low degree nodes as in the Internet. For example, about 45% of the nodes neighboring the second largest node[2] in the Inet-2.2 topology has degree 3 or less. This again is in sharp contrast to the Internet topology where 75% of the second largest node's neighbors have degree 3 or less. Very similiar results were found for a variety of topology sizes. The lowest degree nodes and the highest degree nodes connect far more often in the Internet topology than they do in an Inet-2.2 topology.

Intuitively, if a topology has many small degree nodes connected to each other, these connections need to be covered by one of the small degree nodes in the vertex cover. Adding these small degree nodes increases the vertex cover as they do not cover many edges. Under linear preference, the probability of a node i connecting to a node j is simply a function of the degree of node j, whereas from our understanding of AS relationships in the Internet, and the empirical data shown in Figs. 10 and 11, it appears that the probability of a node i connecting to a node j is a function of the degrees of both nodes i and j. If the degrees of nodes i and j are very different, the probability that they connect to each other should be higher than linear preference. If the degrees of nodes i and j are close, we leave the preference linear. A natural way to achive this desired function would be to weight the original linear preference function with a factor of the distance between the degrees of the two nodes being considered. We use the Frequency-Outdegree power law to set this weight to be the Euclidian distance between the two degrees on the log-log plot. Let d_i and d_j be the degrees of nodes i and j respectively, and $f(d_i)$ and $f(d_j)$ be the frequency of those degrees. Then w_i^j is the weighted value of d_j with respect to d_i, and is defined as:

$$w_i^j = MAX\left(1, \sqrt{\left(\log\frac{d_i}{d_j}\right)^2 + \left(\log\frac{f(d_i)}{f(d_j)}\right)^2}\right) \cdot d_j. \tag{3}$$

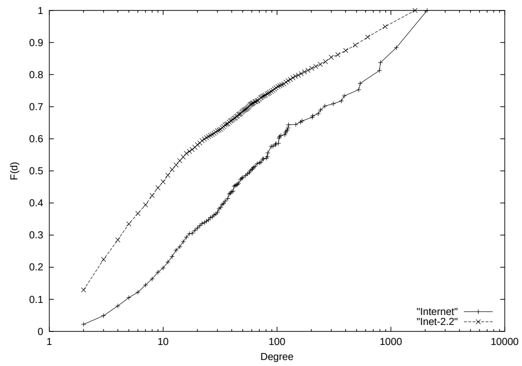
Then $P(i,j)$, the probability that a node i with degree d_i connects to a node j of degree d_j is:

$$P(i,j) = \frac{w_i^j}{\sum_{k\in G} w_i^k}. \tag{4}$$

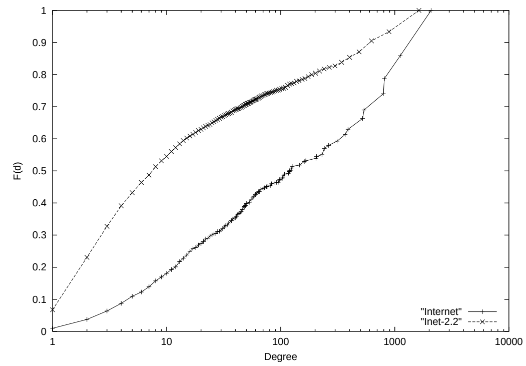
In Inet-3.0, whenever we need to connect a degree of a node i, we use $P(i,j)$ to calculate the probability that it connects to another node j. Guided by our observations of Internet connectivity, this simple change has profound changes on the connectivity of nodes in Inet-3.0 compared to Inet-2.2. Fig. 12 shows the plot of $\mid VC \mid /n$ over time. Here we see the value $\mid VC \mid /n$ of Inet-3.0 topologies closely matches that of Internet topologies for all sizes of topologies.

Figs. 13 and 14 show the new neighbor degree distributions for Inet-3.0 compared to Inet-2.2 and the Internet for February 2001, and Figs. 15 and 16 show the distributions for February 2002. We see that the neighbor degree distributions of Inet-3.0 topologies are much closer to those of the Internet in Figs. 13 and 14. The results are not quite as dramatic when we examine topologies from February 2002 in Figs. 15 and 16. For the February 2002 topologies, which have about 12,700 nodes, we see in a few cases for low

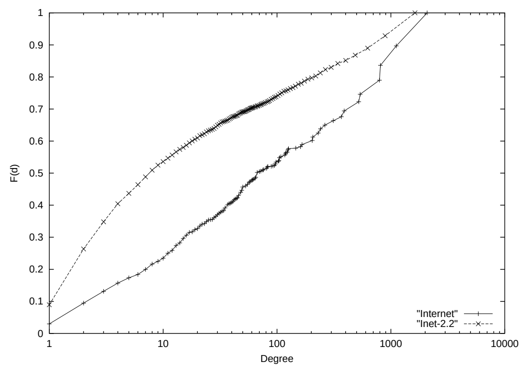
[2] In both topologies this node has a degree well over 1000.



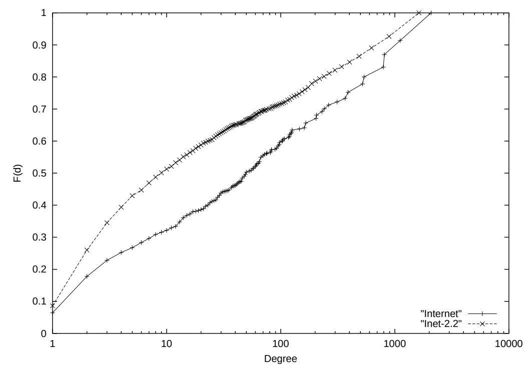
a. F(d) for nodes with degree 1.



b. F(d) for nodes with degree 2.

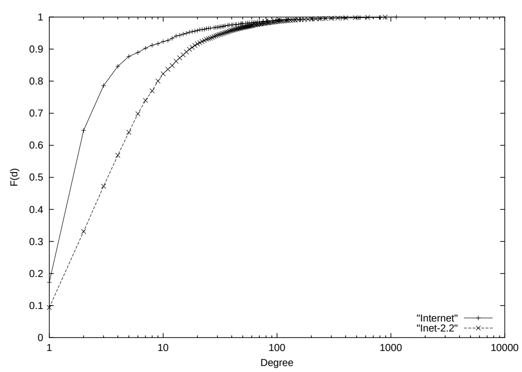


c. F(d) for nodes with degree 3.

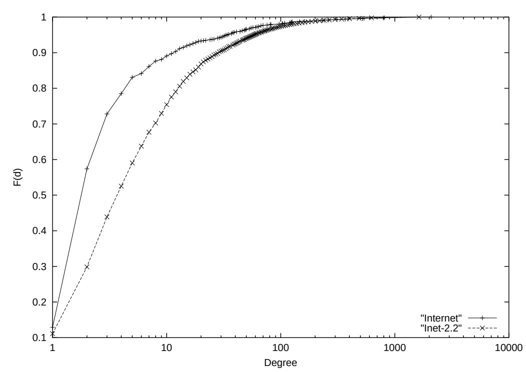


d. F(d) for nodes with degree 4.

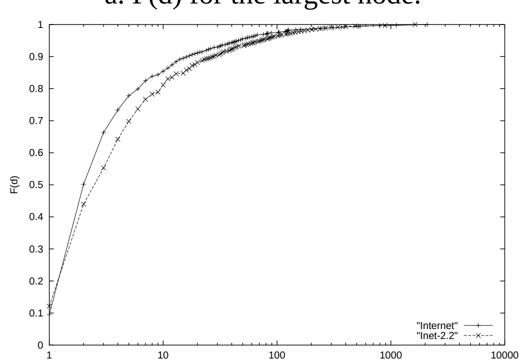
Figure 10: Cumulative Distribution of Neighbors Degree Frequency for Degree 1-4 Nodes



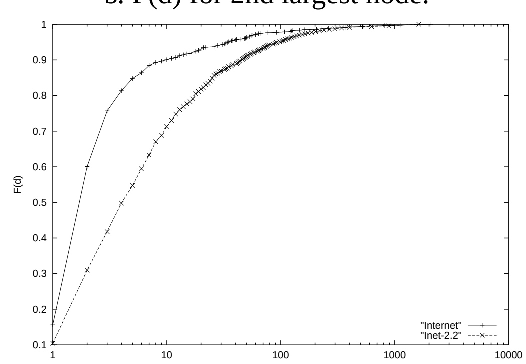
a. F(d) for the largest node.



b. F(d) for 2nd largest node.



c. F(d) for 3rd largest node.



d. F(d) for 4th largest node.

Figure 11: Cumulative Distribution of Neighbors Degree Frequency for Top 4 Nodes

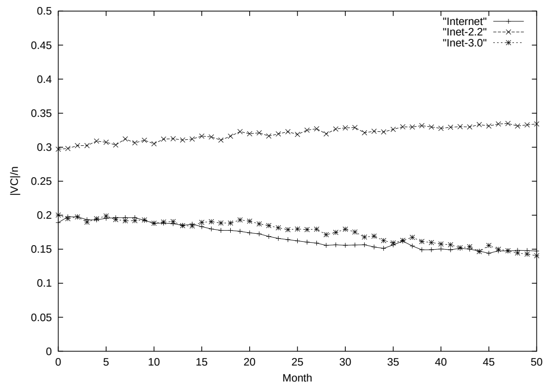


Figure 12: Vertex Cover over time

degree nodes that Inet-3.0 overcompansates for the problems seen in Inet-2.2. That is, low degree nodes are seen to connect to too few low degree nodes compared to on the Internet. This again stresses the need for a better understanding of Internet connectivity and metrics to fully describe the connectivity. In the next section, we examine several other metrics to show the improvements of Inet-3.0 over Inet-2.2.

4 Further Analysis

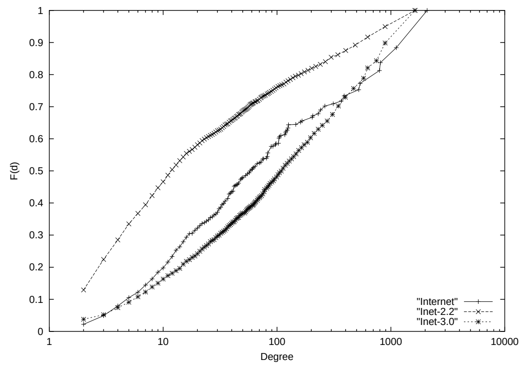
In this section we examine the topologies generated by Inet-3.0 along several additional metrics. When applicable, we compare Inet-3.0 to Inet-2.2 and the Internet over the 51 topology sizes of our Internet data set. For other metrics, this time based comparison is not feasible, in such cases, we use 3 snapshots of the Internet topologies, from February 2000, February 2001, and February 2002. These topologies have sizes of approximately 6700, 8880, and 12700 nodes, respectively. Below we describe the various characteristics we examined.

Outdegree versus rank power law. The first power law identified in [3] was $d_v \propto r_v^R$, where d_v is the degree of a node v, r_v is the rank of node v on a sorted list in decreasing order of node degree, and R is a constant. Fig. 19 shows that Inet-3.0 plot is almost identical to that of the Internet, an improvement over Inet-2.2.

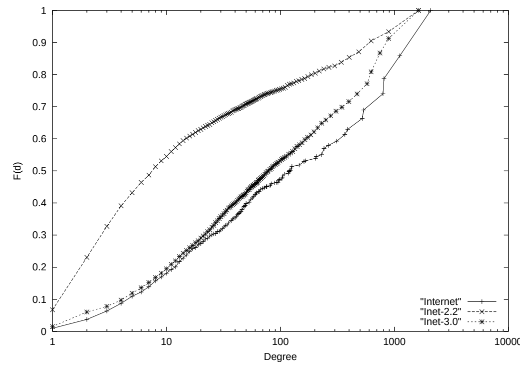
Pair size within h hops. The neighborhood size, $A_v(h)$, of an AS v within h hops is the number of ASs reachable within h hops from v. Pair size within h hops, $P(h)$, is then the sum of all ASs' neighborhood sizes within h hops. It was shown in [5] that $P(h)$ follows an exponential growth law defined as: $P_t(h) = e^{s_h t} P_0(h)$ where $P_0(h)$ is the pair size within h hops in November 1997, s_h is the pair size growth rate, and t is the number of months since 1997. Fig. 18 show the pairsize within 1, 2, 3, and 4 hops. We see both Inet-2.2 and Inet-3.0 follow the Internet data closely.

Resilience. Resilience $R(n)$ is defined in [12] to be the average minimum cut-set size within an n-node ball around any node in the topology. Simply, resilience characterizes the robustness of a graph by measuring the average number of alternate paths between all pairs of nodes in the topology. Fig. 20 shows that Inet-2.2, Inet-3.0, and Internet topologies all show similar resilience.

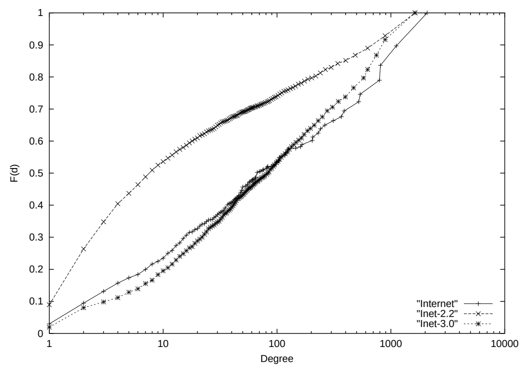
Normalized Laplacian spectrum of a graph. It is shown in [13] that the normalized laplacian spectrum (nls) is a much more discriminating and descriptive metric than the largest eigenvalues of the adjacency matrix of a graph. Coming from the field of spectral graph theory, the (nls) is related to



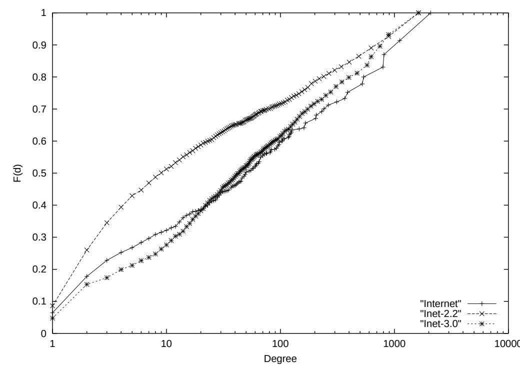
a. F(d) for nodes with degree 1.



b. F(d) for nodes with degree 2.

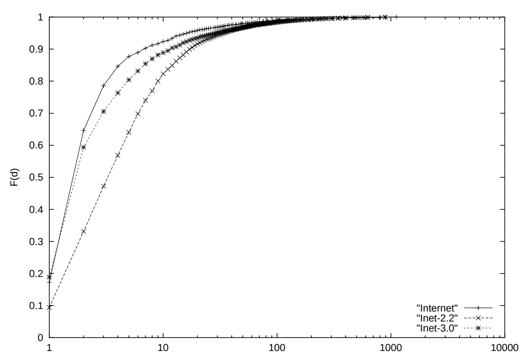


c. F(d) for nodes with degree 3.

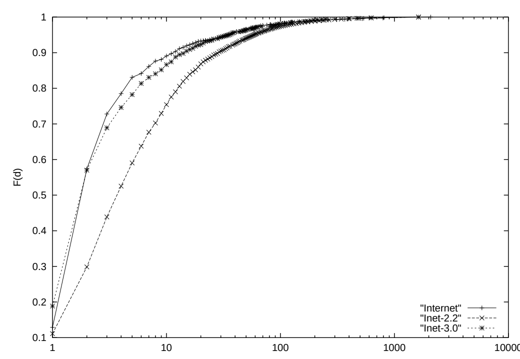


d. F(d) for nodes with degree 4.

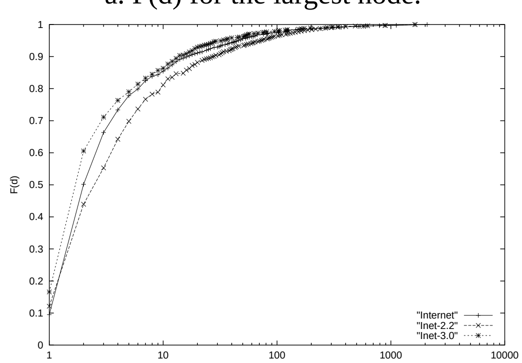
Figure 13: Cumulative Distribution of Neighbors Degree Frequency for Degree 1-4 Nodes, Feb. 2001



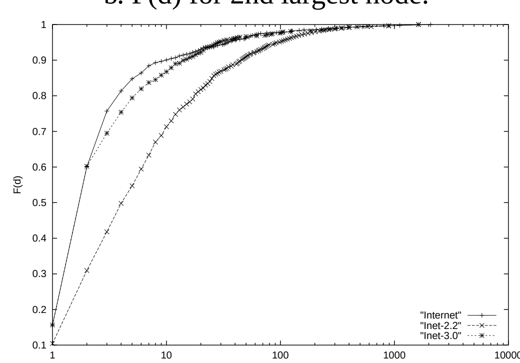
a. F(d) for the largest node.



b. F(d) for 2nd largest node.

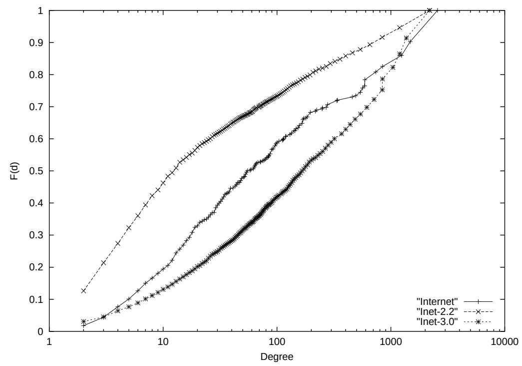


c. F(d) for 3rd largest node.

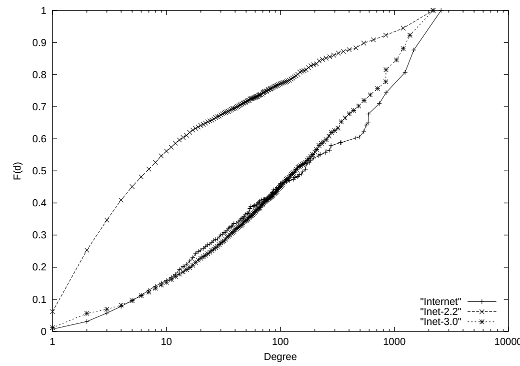


d. F(d) for 4th largest node.

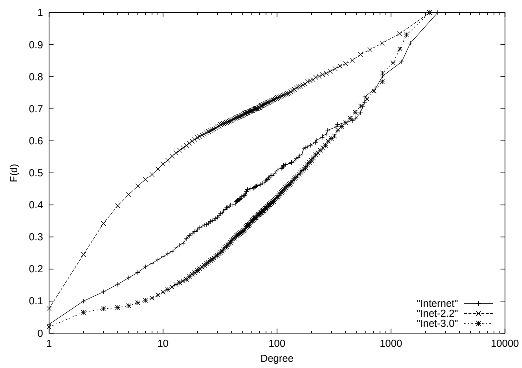
Figure 14: Cumulative Distribution of Neighbors Degree Frequency for Top 4 Nodes, Feb. 2001



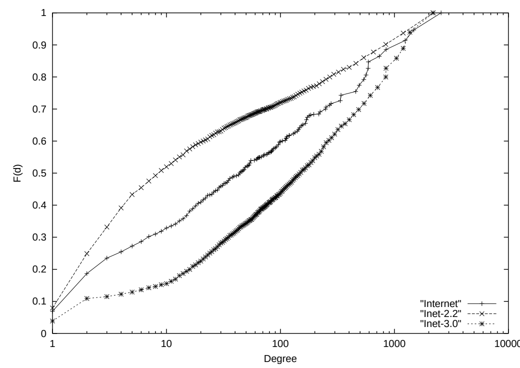
a. F(d) for nodes with degree 1.



b. F(d) for nodes with degree 2.

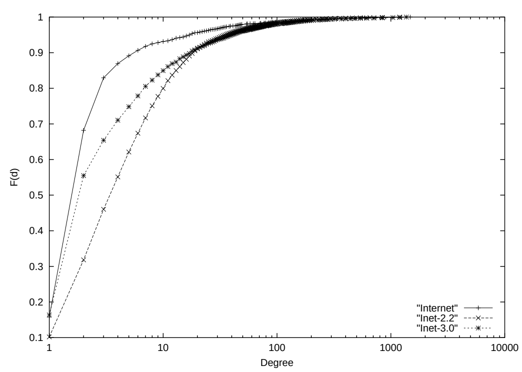


c. F(d) for nodes with degree 3.

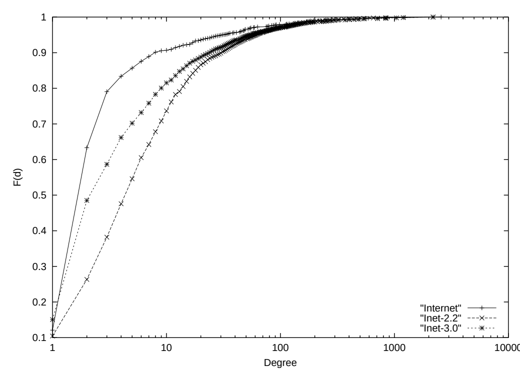


d. F(d) for nodes with degree 4.

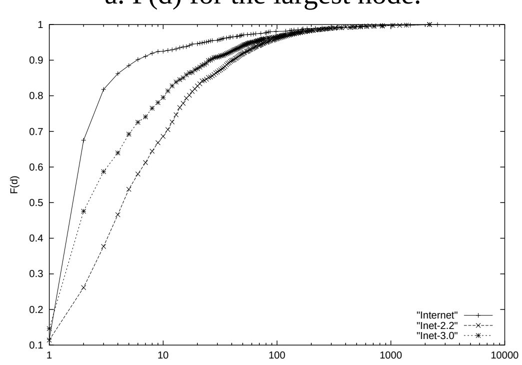
Figure 15: Cumulative Distribution of Neighbors Degree Frequency for Degree 1-4 Nodes, Feb. 2002



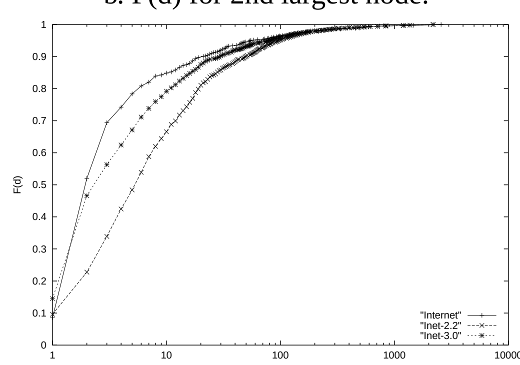
a. F(d) for the largest node.



b. F(d) for 2nd largest node.



c. F(d) for 3rd largest node.



d. F(d) for 4th largest node.

Figure 16: Cumulative Distribution of Neighbors Degree Frequency for Top 4 Nodes, Feb. 2002

Internet graphs in [13], where they define the normalized Laplacian of the graph G to be the matrix $\mathcal{L}(G)$ where:

$$\mathcal{L}(G)(u,v) = \begin{cases} 1 & \text{if } u = v \text{ and } d_v \neq 0, \\ -\frac{1}{\sqrt{d_u d_v}} & \text{if } u \text{ and } v \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

the nls is then just the set of eigenvalues of $\mathcal{L}(G)$. Vukadinoviè *et al.* show a large discrepancy between the nls plot of Inet-2.2 and Inetnet topologies in [13]. Much to our appreciation, the authors ran their spectral analysis on Inet-3.0 topologies. This analysis showed the nls plot of Inet-3.0 to be very similiar to that of the Internet, a large inprovement over Inet-2.2.

Characteristic (or Average) Path Length. For each node, calculate its average shortest path distance to all other nodes in the topology. Average path length is then the average of these distances for all nodes. This metric has been described in [14], [1], and [6]. Fig. 21 shows that both Inet-2.2 and Inet-3.0 have similiar characteristic path lengths over time compared to the Internet.

Average Eccentricity. This metric measures the diameter of a graph. The eccentricity of a node is the maximum distance from the node to all other nodes in the graph. Average eccentricity is the average of all nodes' eccentricity. Fig. 22 shows that Inet-3.0 topologies have an average eccentricity that is more consistant and much closer to the average eccentricity of the Internet compared to Inet-2.2.

Distortion. This metric comes from graph theory literature [4], but was related to Internet topologies in [12]. Distortion measures how much a spanning tree distorts the distances between adjacent nodes in the original graph. It measures how tree-like a topology is. Fig. 23 shows that Inet-2.2, Inet-3.0, and Internet topologies all show similar distortion.

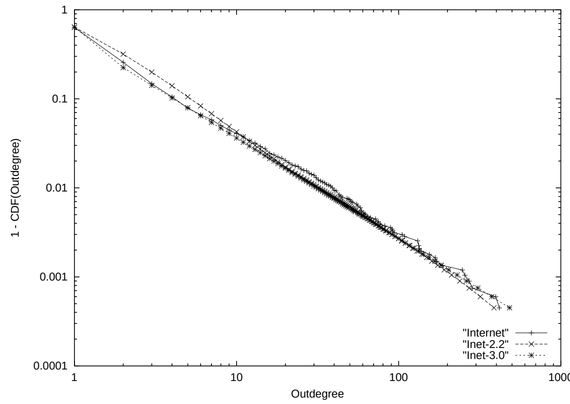
Maximum Clique Size. A clique C is a set of nodes that are all pairwise adjacent. The maximum clique size of a graph is just the size of the largest clique in the graph. Finding the maximum clique in a graph is an NP-hard problem, but there are many well-known approximation algorithms [10]. Fig. 24 highlights the most signifigant known deficiency with Inet-3.0 topologies. The maximum clique size in Inet-3.0 is consistently about half as large as the maximum clique size of the Internet. In this one case, we also see Inet-3.0 performs worse than Inet-2.2 in matching the characteristics of the Internet. As problems with vertex cover were a primary motivation for investigation into Internet connectivity and modification for Inet-3.0, these limitations with maximum clique size will drive further analysis of Internet connectivity and will be discussed in future research.

Clustering Coefficient. Like characteristic path length, the clustering coefficient has been popularized by the interest in "small-world" graphs [14], and has been used to characterize Internet topologies in [1, 7]. The clustering coefficient is a measure of how many of a nodes' neighbors are adjacent to each other. This is calculated for all nodes and averaged. Fig. 25 shows the other known problem with Inet-3.0 topologies. We believe that this problem with clustering coefficient is closely related to the smaller maximum clique sizes mentioned above. Consider the following scenerio in the Internet; if a small, degree 2 AS is multi-homed to two large ASs that peer with each other, this degree 2 AS will have a clustering coefficient of 1. As a signifigant percentage of ASs are small ASs that are multi-homed to large, transist ASs who are peers of each other, there will be a signifigant percentage of ASs with a clustering coefficient of 1. This will clearly result in the large clustering coefficient we see on the Internet. In Inet-3.0, we lack the large, dense clique of the top degree nodes that we see on the Internet. With a smaller maximum clique of top degree nodes, it is less likely that a small node will be multi-homed to two large nodes that are peers of each other, and we therefore have a smaller clustering coefficient in Inet-3.0 topologies. Again, this issue will be investigated, and reported in future work.

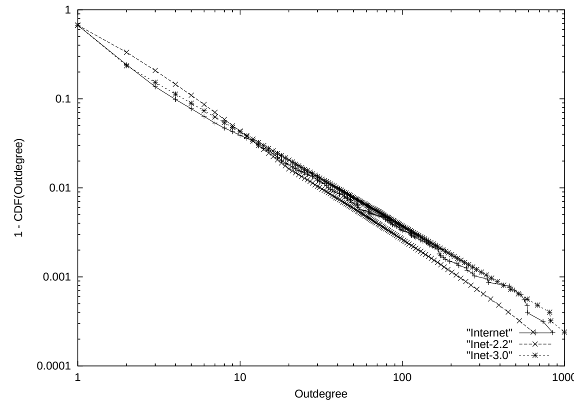
5 Conclusion and Future Work

As our analysis shows, Inet-3.0 performs as well as or better than Inet-2.2 in terms of matching the characteristics of Internet topologies. We have improved upon degree distribution and minimum vertex cover size, as was our intention initially. We have uncovered two (related) problems with the connectivity of Inet-3.0 topologies though, in terms of maximum clique size and clustering coefficient. This emphasizes the need for a better understanding of Internet connectivity, which has proven to be much harder to quantify than degree distribution. Our future work will address these limitations in density in Inet-3.0 topologies.

We thank Thomas Erlbach, Polly Huang, Damien Magoni, Hongsuda Tangmunarunkit, and Danica Vukadinoviȅ for their assistance in analyizing our topologies, and Amnon Shochat for discussion regarding vertex cover and a script to approximate the minimum vertex cover of Inet graphs.

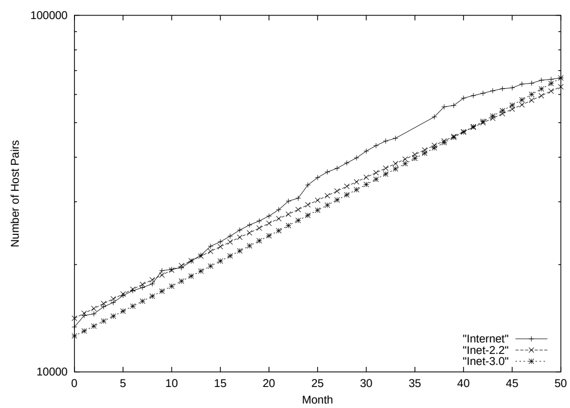


a. February 01, 2000.

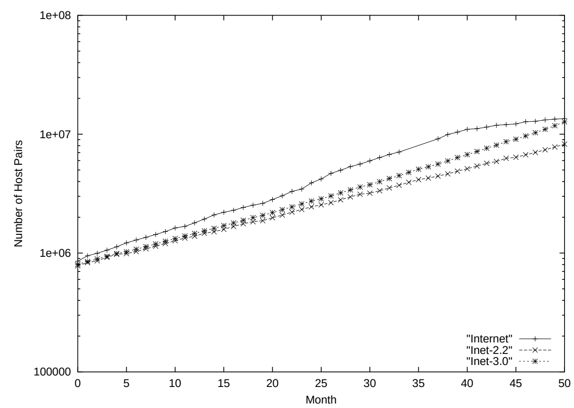


b. February 01, 2002.

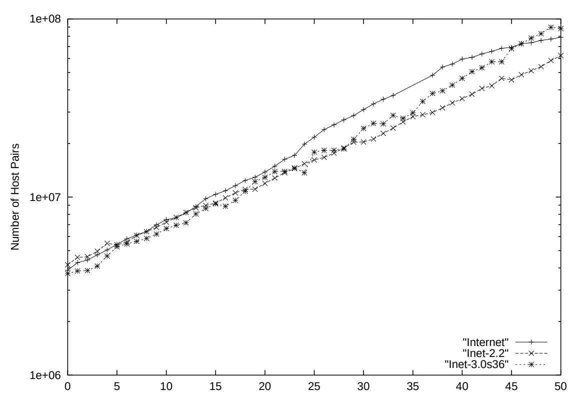
Figure 17: Additional ccdf of Frequency versus Outdegree



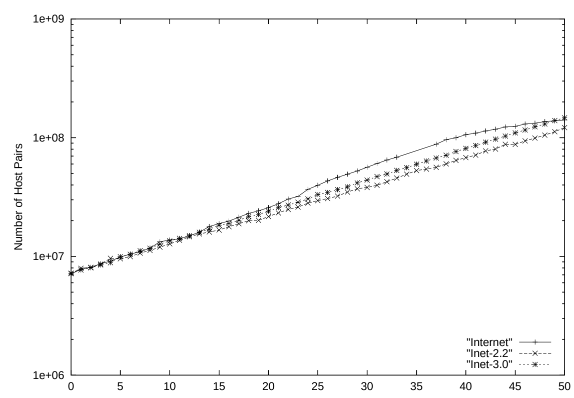
a. Pair Size within 1 hop.



b. Pair Size within 2 hops.

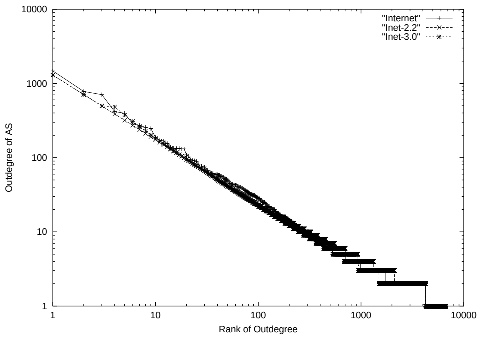


a. Pair Size within 3 hops.

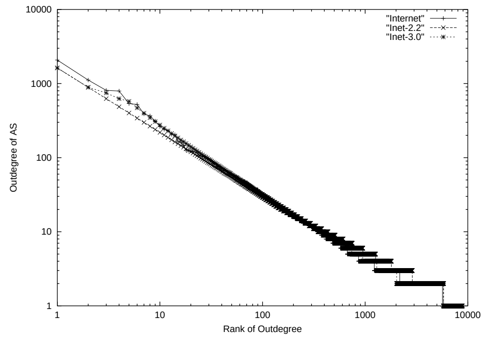


b. Pair Size within 4 hops.

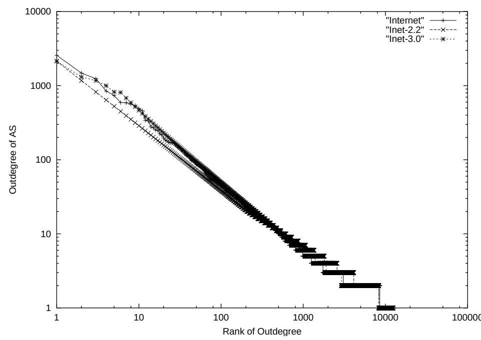
Figure 18: Pair Sizes within h hops over time.



a. February 01, 2000.

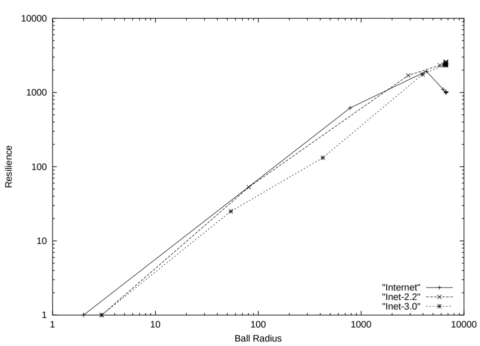


b. February 01, 2001.

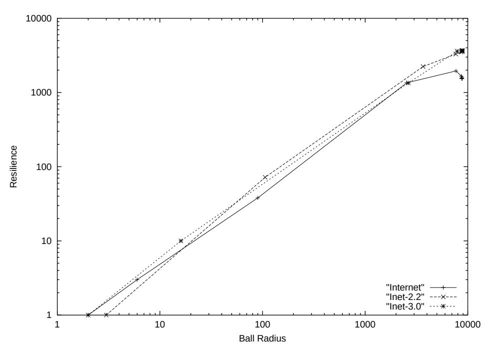


a. February 01, 2002.

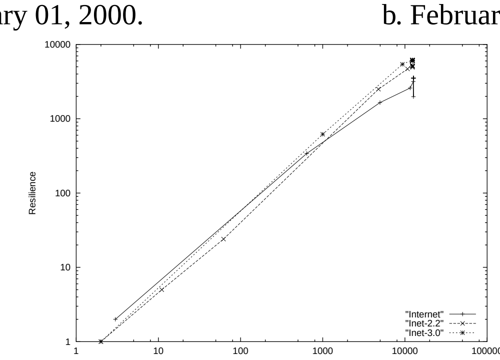
Figure 19: Outdegree versus Rank power law.



a. February 01, 2000.



b. February 01, 2001.



a. February 01, 2002.

Figure 20: Resilience.

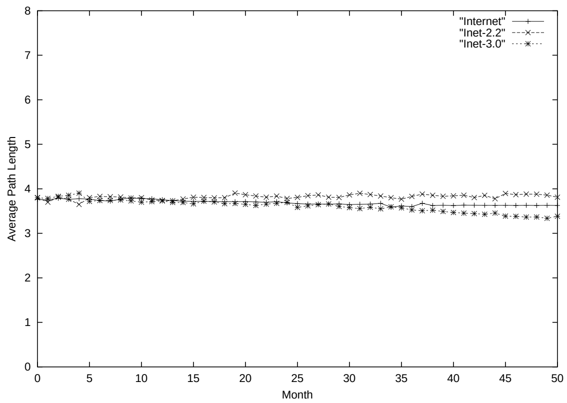


Figure 21: Characteristic path length over time.

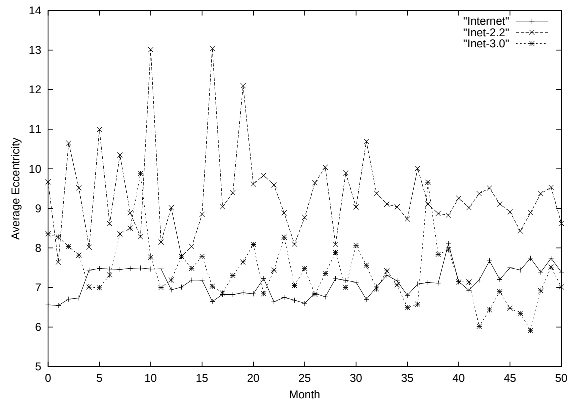
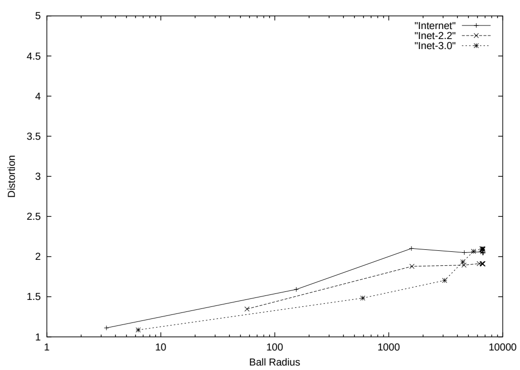
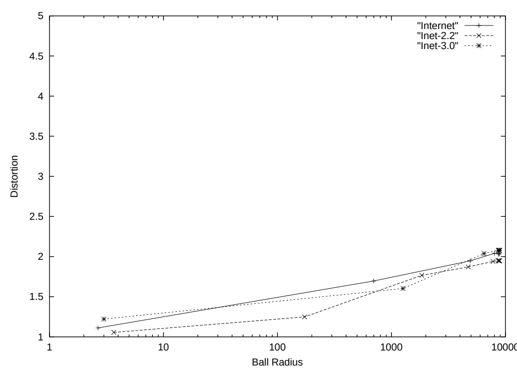


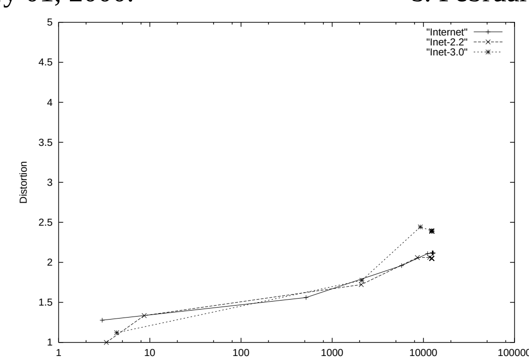
Figure 22: Average eccentricity over time.



a. February 01, 2000.



b. February 01, 2001.



a. February 01, 2002.

Figure 23: Distortion.

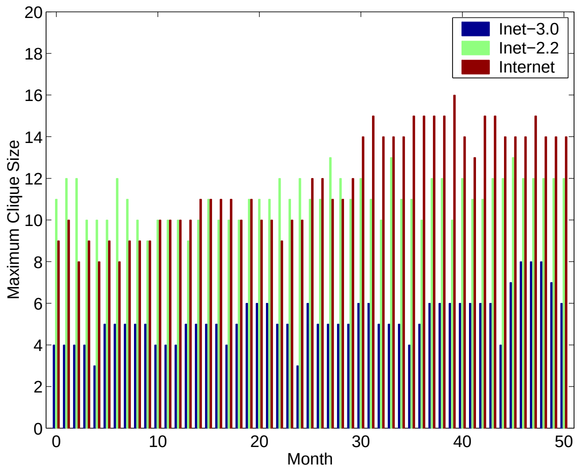


Figure 24: Maximum clique size over time.

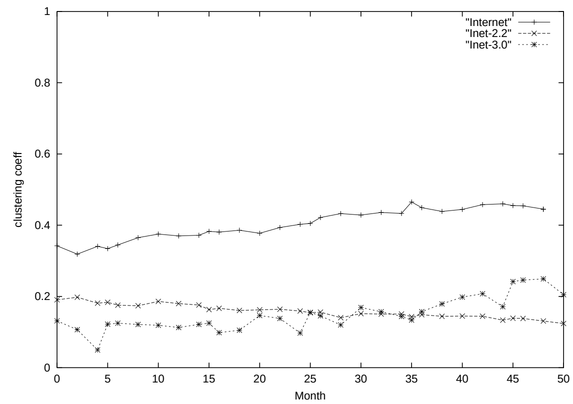


Figure 25: Clustering coefficient.

References

[1] T. Bu and D. Towsley. On distinguishing betweem internet power law topology generators. In *In. Proc. of IEEE INFOCOM*, 2002.

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

[3] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *Proc. of ACM SIGCOMM*, pages 251–262, 1999.

[4] T.C. Hu. Optimum communication spanning trees. *SIAM Journal of Computing 3*, pages 188–195, 1974.

[5] C. Jin, Q. Chen, and S. Jamin. Inet: Internet topology generator. Tech.rep.cse-tr-433-00, University of Michigan EECS Dept., 2000.

[6] D. Magoni and J.J. Pansiot. Analysis of the autonomous system network topology. *ACM SIGCOMM Computer Communication Review*, pages 26–37, July 2001.

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

[8] NLANR. National laboratory for applied network research.

[9] K. Park and H. Lee. On the effectiveness of route-based packet filtering for distributed dos attack prevention in power-law internets. In *In. Proc. of ACM SIGCOMM*, 2001.

[10] M. Pelillo. Heuristics for maximum clique and independent set.

[11] Routeviews.org. The university of oregon route views archive project.

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

[13] D. Vukadinovìc, P. Huang, and T. Erlebach. A spectral analysis of the internet topology. Eth tik-nr.118, Swiss Federal Institute of Technology (ETH), 2001.

[14] D. Watts and S. Strogatz. Collective dynamics of ’small-world’ networks. *Nature*, pages 440–442, June 1998.