Impact of Degree Correlations on Topology Generators

Priya Mahadevan, UCSD, pmahadevan@cs.ucsd.edu
Dmitri Krioukov, CAIDA, dima@caida.org
Brad Huffaker, CAIDA, brad@caida.org
KC Claffy, CAIDA, kc@caida.org
Amin Vahdat, UCSD, vahdat@cs.ucsd.edu

http://www.caida.org/analysis/topology/as_topo_comparisons

Internet topology analysis has attracted a lot of attention in recent years, since its relationship to a variety of networking technologies and phenomena is not only fundamental but generally poorly understood. Performance of applications and protocols, especially routing protocols, robustness of the network under attack, traffic engineering, etc., depend crucially on the topology of the network. In this work we focus on two cornerstones of Internet topology research: 1) knowledge of realistic network topologies; and 2) availability of tools to generate them.

The following three characteristics relate directly to practically important network properties: 1) *distance* (shortest path length) distribution, critical to scalability of routing protocols; 2) *spectrum* (set of eigenvalues of graph's adjacency matrix), important for network robustness and performance analysis; and 3) *betweenness* (the number of shortest paths passing through a node or link) distribution, important for analyzing the hierarchical structure of the network and for estimating the accuracy of traceroute-like explorations of the network.

Unfortunately, the brute force approach to generating synthetic topologies reproducing any of the above three metrics does not work: there are no known algorithms for constructing a graph with a target distance distribution, betweenness distribution, or spectrum. Furthermore, this brute force approach suffers from a daunting methodological problem: as soon as a new metric is discovered to be important, one's topology generator algorithm must reproduce this newly discovered metric.

There are at least two ways to solve this problem. One is to discover the true laws driving Internet evolution, so that topology generators based on them reproduce *all* important topological properties. This remains a difficult and open research challenge, made more difficult by the fact that the driving forces behind Internet evolution are likely to change over time.

An alternative way is to focus on interdependencies among the network topology metrics. By reproducing most simple characteristics that are not necessarily practically important, can we also capture all other properties including practically important ones? The main contribution of this work is the exploration of a single graph theoretic property that appears to define a wide range of known topology metrics.

Tag	Name	Degree correlations of nodes at distance
0K	Avg. degree	None
1K	Degree distribution	0
2K	Joint node degree distribution	1
3K	Joint edge degree distribution	2
...	...	...
(D+1)K	Full degree distribution	D

Table 1: Connectivity characteristics of a network with maximum distance D.

The most basic topology metrics are those characterizing connectivity in a network. We can list these characteristics in the order of increasing amount of information about local connectivity structure of the network (cf. Table 1).

The average node degree in the network (0K) is the coarsest connectivity characteristic. We can easily reproduce it: the classical (Erdős-Rényi) random graphs are random graphs having a given average node degree. We can construct them by connecting every pair of nodes with a certain fixed probability. We call such graphs *0K-random*. The 0K-random graphs have little in common with realistic topologies. In particular, they have Poisson node degree distributions, while the node degree distribution of the AS-level topology follows a power law.

The node degree distribution (1K) is a more informative characteristic—the probability that a randomly selected node is of degree k. There are known algorithms for constructing random graphs having a given form of the degree distribution. We call such graphs *1K-random*. The most popular topology generators today (PLRG, Inet, BRITE, etc.) incorporate the idea of reproducing the node degree distribution observed in the real network. It is known that 1K-random graphs are much closer to reality than 0K-random, but they also have a set of critical incongruities with real topologies.

The joint node degree distribution (2K) is an even more informative connectivity characteristic. It is the probability that a randomly selected edge connects nodes of degrees k_1 and k_2. It tells us how nodes of different degrees connect to each other. In other words, it describes correlations of degrees of nodes located at distance 1 from each other. The joint node degree distribution and topology generators trying to reproduce it have not been yet systematically considered in networking research.

As we increase the value of s in our notation sK in Table 1, $s = 0 \ldots D$, where D is the graph diameter, we define distributions of degrees of nodes located at a maximum distance s from each other. In doing so we gain information about the local structure of the topology, which allows a more accurate description of a node's (increasingly wider) neighborhood. The important question is: do we need to go all the way up to $(D + 1)K$ or can we obtain reasonable accuracy at a lower value of s, such as 2?

To this end, we implemented an efficient topology generator that creates random graphs with a given form of the joint node degree distribution. We call such graphs *2K-random*. In analyzing the accuracy of our 2K-random graph generator, we compared its output with observed AS-level topologies. We found that *all* important topology characteristics outlined above exhibited a close match with the measured topology. Figure 1 plots AS hop distance distributions for three data sets: our 2K generator; the popular Inet generator; and RouteViews BGP data from March 2004. The distance distribution for the 2K-generated topology matches the observed (RouteViews) data better than the Inet-generated topology does.

The only topology characteristic we are aware of that did not closely match between 2K-random and observed

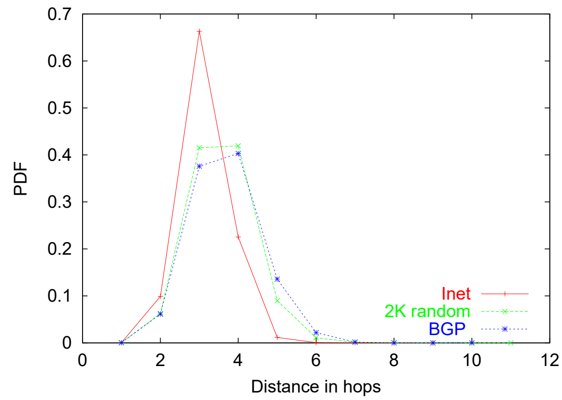

Figure 1: Distance distribution comparing original BGP table data for March 2004, 2K generator, and Inet

topologies was clustering. Clustering is a measure of how close the neighborhood of the average node of degree k is to a clique. One can show that clustering is not completely defined by the joint node degree distribution, and therefore it is not surprising that we did not have a close match. We expect that a 3K-random graph generator is likely to have more success in reproducing observed clustering. We note that it is unclear how important clustering is for practical networking applications.

To explore the variability among different executions of the generator, we generated 100 random topologies by specifying a different random seed for each execution. We found that the standard deviations for the metric values associated with distance, betweenness distributions, and spectrum were low in all cases. We checked that no pair of the 100 generated topologies was isomorphic.

We further explored the space of all 2K graphs by generating non-random graphs that preserved the node joint degree distribution. We found that all the important characteristics (distance, betweenness, spectrum) remained the same, meaning that 2K is a definitive metric and that a 2K generator is guaranteed to always match closely the observed AS-level topologies. We validated our 2K generator on AS graphs extracted from traceroute measurements and from the WHOIS database. We confirmed that our 2K generator reproduces all the important metrics in these cases, too.

Our current research agenda includes building a 3K generator to better match clustering. We are also trying to determine the minimal value of s in our sK notation required to closely match any given topology, such as the router-level Internet.