Modeling and Generating Internet Topology
Can we characterize the Internet’s Topology?
How to generate realistic Internet topology for simulations?
Model Internet as a Graph
Router-Level, 
node = router 
edge = 1-hop link
AS-Level,
node = AS domain
edge = Peering
Generating Random Graph
Randomly generate points on a plane
Connects two nodes with fixed probability $p$
Waxman’s Method
Randomly generate points on a plane
Connects two points with probability $P(u, v)$

$L$: maximum distance
$d(u, v)$: distance between $u$ and $v$
Model locality but not the structure of Internet
Transit-Stub Method
Randomly generate a graph using Waxman’s method
Each node is expanded to form a random graph (transit domain)
Connect stub domains to the transit domain.
Looks good, but is it close to the real thing?
The Faloutsos brothers, SIGCOMM ‘99
Use four traces of Internet topology collected between 97-98
### AS-Level Topology

<table>
<thead>
<tr>
<th>Time</th>
<th>Num of Nodes</th>
<th>Num of Edges</th>
<th>Max outdegree</th>
<th>Average outdegree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov 97</td>
<td>3015</td>
<td>5156</td>
<td>590</td>
<td>3.42</td>
</tr>
<tr>
<td>Apr 98</td>
<td>3520</td>
<td>6432</td>
<td>745</td>
<td>3.65</td>
</tr>
<tr>
<td>Dec 98</td>
<td>4398</td>
<td>8256</td>
<td>979</td>
<td>3.76</td>
</tr>
</tbody>
</table>

### Router-Level Topology

<table>
<thead>
<tr>
<th>Year</th>
<th>Num of Nodes</th>
<th>Num of Edges</th>
<th>Num of Routers</th>
<th>Average Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>3888</td>
<td>5012</td>
<td>2.57</td>
<td></td>
</tr>
</tbody>
</table>
Observations: the graphs can be decomposed into two components: trees and core.
40-50\% of the nodes are in trees
maximum depth of trees
1

depth of >80% of the trees
<table>
<thead>
<tr>
<th>Time</th>
<th>Num of Nodes</th>
<th>Num of Edges</th>
<th>Max outdegree</th>
<th>Average outdegree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov 97</td>
<td>3015</td>
<td>5156</td>
<td>590</td>
<td>3.42</td>
</tr>
<tr>
<td>Apr 98</td>
<td>3520</td>
<td>6432</td>
<td>745</td>
<td>3.65</td>
</tr>
<tr>
<td>Dec 98</td>
<td>4398</td>
<td>8256</td>
<td>979</td>
<td>3.76</td>
</tr>
</tbody>
</table>

out-degree is highly skewed!
Let $d_v$ be the out-degree of a node, and $r_v$ be the rank of a node (i.e., index in the order of decreasing outdegree).
Plot $d_v$ versus $r_v$ on log-log scale
\[ \text{"971108.rank"} \]
\[ \exp(6.34763) \times x^{\text{-0.811392}} \]
"980410.rank"

\[ \exp(6.62082) \times x^{(-0.821274)} \]
"981205.rank"

\[ \exp(6.16576) \times x^{(-0.74496)} \]
\[ \log d_v = \mathcal{R} \log r_v + c \]

\[ d_v \propto r_v^{\mathcal{R}} \quad \text{Rank Exponent} \]
Lemma 1: \[ d_v = \frac{1}{N^R} r^R \]
Lemma 2: \[ E = \frac{N}{2(\mathcal{R} + 1)} \left( 1 - \frac{1}{N\mathcal{R} + 1} \right) \]
Let $f_d$ be the number of nodes with out-degree $d$.
Plot $f_d$ versus $d$ on log-log scale
"971108.out"

\[ \exp(7.68585) \times x^{(-2.15632)} \]

(a) Int-11-97
(b) Int-04-98

$\exp(7.89793) \times x^{(-2.16356)}$
\text{"routes.out"} \\
\exp(8.52124) \times x^{(-2.48626)}
"981205.out"

$\exp(8.11393) \times x^{-2.20288}$

(a) Int-12-98
\[ f_d \propto d^0 \]
“few connects to many, many connects to few”
On the Effectiveness of Route-Based Packet Filtering for Distributed DoS Attack Prevention in Power-Law Internets

Kihong Park† Heejo Lee‡
Network Systems Lab
Department of Computer Sciences
Purdue University
West Lafayette, IN 47907
{park,hlee}@cs.purdue.edu

that cannot be proactively curtailed are extremely sparse so that their origin can be localized—i.e., IP traceback—to within a small, constant number of candidate sites. We show that the two proactive and reactive performance effects can be achieved by implementing route-based filtering on less than 20% of Internet autonomous system (AS) sites. Second, we show that the two complementary performance measures are dependent on the properties of the underlying AS graph. In particular, we show that the power-law structure of Internet AS topology leads to connectivity properties which are crucial in facilitating the observed performance effects.
Let $P(h)$ be the number of node pairs within $h$ hops of each other (include self-pairs, count every pair twice)
\[ P(0) = N \]
\[ P(1) = N + 2E \]
$P(\delta) = N^2$

where $\delta$ is the diameter of the graph
Plot $P(h)$ versus $h$ on log-log scale
Figure 7: The hop-plots: Log-log plots of the number of pairs of nodes $P(h)$ within $h$ hops versus the number of hops $h$.

Figure 8: The hop-plots: Log-log plots of the number of pairs of nodes $P(h)$ within $h$ hops versus the number of hops $h$. 
$P(h) \propto h^\mathcal{H}, h \ll \delta$
\[ P(h) = \begin{cases} 
(N + 2E)h^H & h \ll \delta \\
N^2 & h \geq \delta 
\end{cases} \]
\[ P(\delta_{ef}) = N^2 \]
MODELING PEER-TO-PEER NETWORK TOPOLOGIES THROUGH “SMALL-WORLD” MODELS AND POWER LAWS

Mihajlo Jovanović
ECECS Department, University of Cincinnati
Cincinnati, OH 45221

\[ \delta_{ef} = \left( \frac{N^2}{N + 2E} \right)^{1/\gamma} \]

Substituting the values for the Gnutella topology snapshot from December 28, 2000, we get that, during that time, a more cost-effective value for the maximum TTL would have been 4 (instead of 7, which is the default specified by the Gnutella protocol).

IV CRAWLER ARCHITECTURE
Gnutella is a highly dynamic network in which topology is constantly changing as hosts join and leave the network, establish new connections, and close the existing ones. Therefore, discovering topology of the Gnutella network...
Average Number of Nodes within $h$ Hops

$$\text{NN}(h) = \frac{P(h) - N}{N}$$
Average Number of Nodes within $h$ Hops (using average degree)

$$NN(h) = \bar{d}(\bar{d} - 1)^h$$
\[ d_v \propto r_v^R \]
\[ f_d \propto d^O \]
\[ P(h) \propto h^H, h \ll \delta \]
It holds in 97-98. What about later?
“Power-Laws and the AS-Level Internet Topology”
G. Siganos and the Faloutsos brothers, IEEE/ACM TON
Do topologies generated by Waxman and Transit-Stub exhibit Power Law?
Figure 1: Log-log plot of frequency $f_d$ vs. outdegree $d$ for a 5000-node Waxman topology (left) and a 6660-node Transit-Stub topology (right). The correlation coefficient is 0.4 for the Waxman topology, and 0.9 for the Transit-Stub topology.
How to generate topology that follows power laws?
Where does power law come from?
Internet

Diameter of the World-Wide Web

Despite its increasing role in communication, the World-Wide Web remains uncontrolled: any individual or institution can create a website with any number of documents and links. This unregulated growth leads to a huge and complex web, which becomes a large directed graph whose vertices are documents and whose edges are links (URLs) that point from one document to another. The topology of this graph determines the web’s connectivity and consequently how effectively we can locate information on it. But its enormous size (estimated to be at least $8 \times 10^8$ documents\(^1\)) and the continual changing of documents and links make it impossible to catalogue all the vertices and edges.

The extent of the challenge in obtaining a complete topological map of the web is illustrated by the limitations of the current.

Figure 1 Distribution of links on the World-Wide Web. a, Outgoing links (URLs found on an HTML document); b, incoming links (URLs pointing to a certain HTML document). Data were obtained from the complete map of the nd.edu domain, which contains 325,729 documents and 1,469,680 links. Dotted lines represent analytical fits to the data.
MAP OF INTERACTING PROTEINS in yeast highlights the discovery that highly linked, or hub, proteins tend to be crucial for a cell’s survival. Red denotes essential proteins (their removal will cause the cell to die). Orange represents proteins of some importance (their removal will slow cell growth). Green and yellow represent proteins of lesser or unknown significance, respectively.
Even the network of actors in Hollywood—popularized by the game Six Degrees of Kevin Bacon, in which players try to connect actors to Bacon via the movies in which they have appeared together—is scale-free. A quantitative analysis of the network of research collaborations shows that most proteins interact with only one or two others, a few are able to each other. When we investigated Baker's yeast, one of the simplest eukaryotic (nucleus-containing) cells, with thousands of proteins, we discovered a scale-free topology: although most proteins interact with only one or two others, a few are able to interact with many more. Now it has more than three networks have expanded significantly since 1890, but it gradually grew millions, with the new routers added to those that were already present. Thanks to the growth of real networks, older nodes have opportunities to acquire links. Furthermore, all nodes can make connections to other nodes. When deciding where to link a new node, people can choose from many locations. Yet most of us are connected to only a tiny fraction of the population, so that subset tends to include more connected sites because they are easier to reach. By simply linking to those sites, we exercise and reinforce a bias. This process of "preferential attachment" occurs elsewhere. In Hollywood,

---

**Examples of Scale-Free Networks**

<table>
<thead>
<tr>
<th>NETWORK</th>
<th>NODES</th>
<th>LINKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellular metabolism</td>
<td>Molecules involved in burning food for energy</td>
<td>Participation in the same biochemical reaction</td>
</tr>
<tr>
<td>Hollywood</td>
<td>Actors</td>
<td>Appearance in the same movie</td>
</tr>
<tr>
<td>Internet</td>
<td>Routers</td>
<td>Optical and other physical connections</td>
</tr>
<tr>
<td>Protein regulatory network</td>
<td>Proteins that help to regulate a cell's activities</td>
<td>Interactions among proteins</td>
</tr>
<tr>
<td>Research collaborations</td>
<td>Scientists</td>
<td>Co-authorship of papers</td>
</tr>
<tr>
<td>Sexual relationships</td>
<td>People</td>
<td>Sexual contact</td>
</tr>
<tr>
<td>World Wide Web</td>
<td>Web pages</td>
<td>URLs</td>
</tr>
</tbody>
</table>
A SCALE-FREE NETWORK grows incrementally from two to 11 nodes in this example. When deciding where to establish a link, a new node (green) prefers to attach to an existing node (red) that already has many other connections. These two basic mechanisms—growth and preferential attachment—will eventually lead to the system's being dominated by hubs, nodes having an enormous number of links.
Generating Power Law Topology (simplified)
“On the Original of Power Laws in Internet Topologies”
A Medina, I Matta, J Byers,
ACM SIGCOMM, ‘00
BRITE is no longer supported by its developers, but questions can be asked on the brite-users mailing list.

Effective engineering of the Internet is predicated upon a detailed understanding of issues such as the large-scale structure of its underlying physical topology, the manner in which it evolves over time, and the way in which its constituent components contribute to its overall function. Unfortunately, developing a deep understanding of these issues has proven to be a challenging task, since it in turn involves solving difficult problems such as mapping the actual topology, characterizing it, and developing models that capture its emergent behavior. Consequently, even though there are a number of topology models, it is an open question as to how representative the topologies they generate are of the actual Internet. Our goal is to produce a topology generation framework which improves the state of the art and is based on design principles which include representativeness, inclusiveness, and interoperability. **Representativeness** leads to synthetic topologies that accurately reflect many aspects of the actual Internet topology (e.g. hierarchical structure, degree distribution, etc.). **Inclusiveness** combines the strengths of as many generation models as possible in a single generation tool. **Interoperability** provides interfaces to widely-used simulation applications such as ns, SSF and OmNet++ as well as visualization applications. We call such a tool a **universal topology generator**.

BRITE is an approach towards universal topology generation. We designed BRITE to be:
Randomly generate a small graph
Incremental Growth:
Add one node at a time
Preferential Attachment:
Connects to a neighbor $i$ with a probability $p = \frac{d_i}{\sum_{j \in C} d_j}$
Figure 6: Log-log plot of outdegree $d_v$ vs. rank for a 5000-node Waxman topology (left), a 4040-node Transit-Stub topology (middle) and a 5000-node BRITE topology with preferential connectivity and incremental growth (right). The correlation coefficient is 0.81 for the Waxman topology, 0.87 for the Transit-Stub topology, and 0.96 for the BRITE topology.