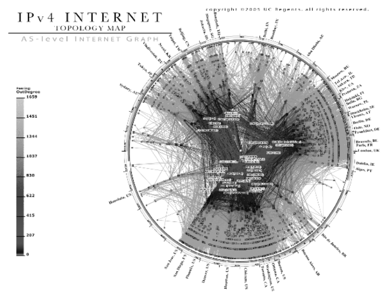
Internet Topology



Can we characterize the Internet's Topology?

How to generate realistic Internet topology for simulations?

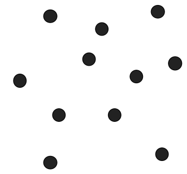
Model Internet as a Graph

Router-Level,
node = router
edge = 1-hop link

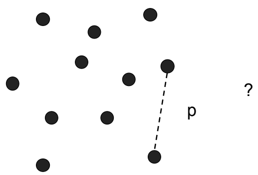
AS-Level,
node = AS domain
edge = Peering

Generating Random Graph

Randomly generate points on a plane

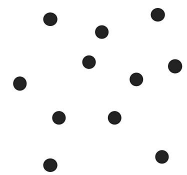


Connects two nodes with fixed probability p

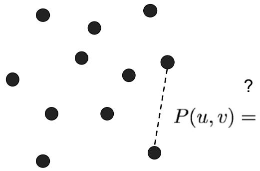


Waxman's Method

Randomly generate points on a plane



Connects two points with probability P(u,v)



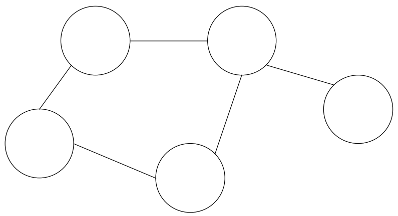
$$P(u,v) = \alpha e^{\frac{-d(u,v)}{\beta L}}$$

L: maximum distance
d(u,v): distance between u and v

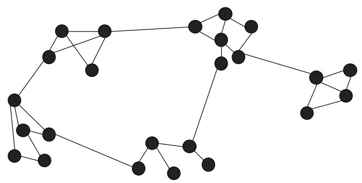
Model locality but not the structure of Internet

Transit-Stub Method

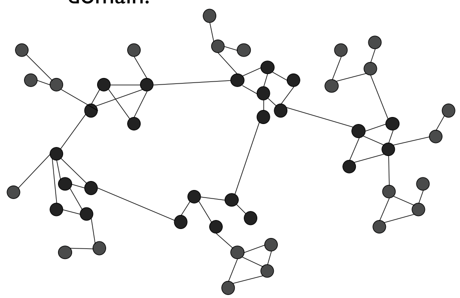
Randomly generate a graph using Waxman's method



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?

"On Power-Law Relationships of the Internet Topology"
The Faloutsos brothers, SIGCOMM '99

Use four traces of Internet topology collected between 97-98

AS-Level Topology

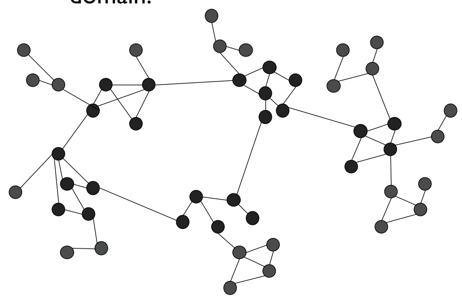
Time	Num of Nodes	Num of Edges	Max outdegree	Average outdegree
Nov 97	3015	5156	590	3.42
Apr 98	3520	6432	745	3.65
Dec 98	4398	8256	979	3.76

Router-Level Topology

1995	3888	5012	2.57

Observations: the graphs can be decomposed into two components: trees and core.

Connect stub domains to the transit domain.



40-50%

of the nodes are in trees

3

maximum depth of trees

1

depth of >80% of the trees

Time	Num of Nodes	Num of Edges	Max outdegree	Average outdegree
Nov 97	3015	5156	590	3.42
Apr 98	3520	6432	745	3.65
Dec 98	4398	8256	979	3.76

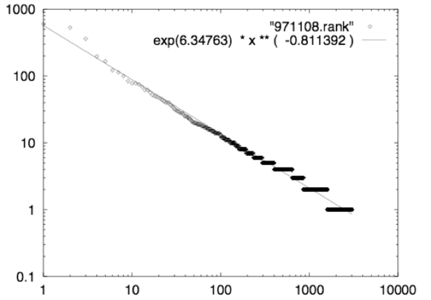
Out-degree is highly skewed!

Let

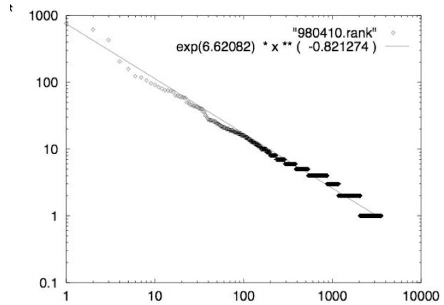
$\mathbf{d_v}$ be the **out-degree** of a node, and

$\mathbf{r_v}$ be the **rank** of a node (i.e., index in the order of decreasing outdegree)

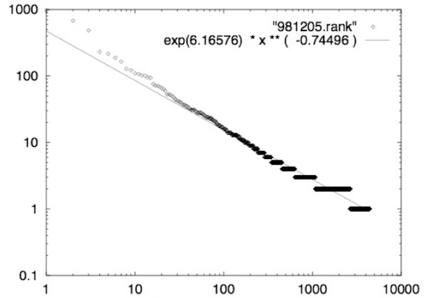
Plot d_v versus r_v on log-log scale



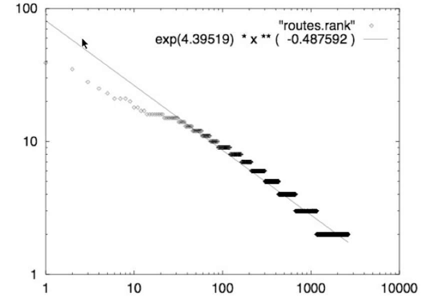
(a) Int-11-97



(b) Int-04-98



(a) Int-12-98



(b) Rout-95

$$\log d_v = \mathcal{R} \log r_v + c$$

$$d_v \propto r_v^{\mathcal{R}}$$

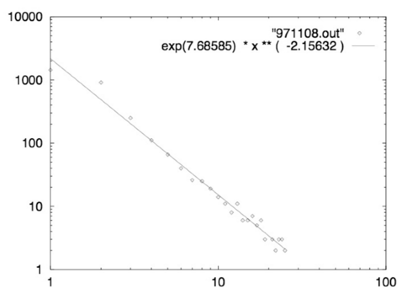
← Rank Exponent

Lemma 1: $d_v = \frac{1}{N^R} r_v^{\mathcal{R}}$

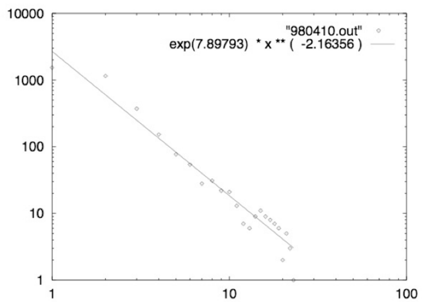
Lemma 2: $E = \frac{N}{2(\mathcal{R}+1)}\left(1 - \frac{1}{N^{\mathcal{R}+1}}\right)$

Let
f_d be the **number of nodes** with out-degree d

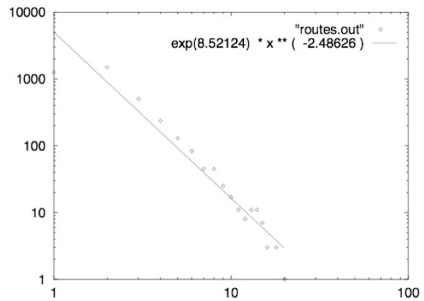
Plot f_d versus d on log-log scale



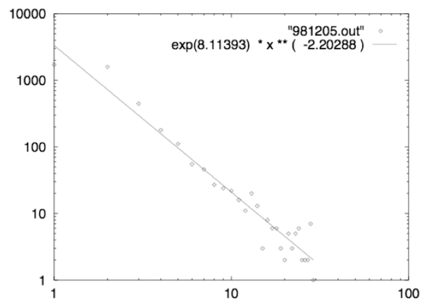
(a) Int-11-97



(b) Int-04-98



(b) Rout-95



(a) Int-12-98

$f_d \propto d^{\mathcal{O}}$

Let
P(h) be the **number of node pairs** within h hops of each other
(include self-pairs, count every pair twice)

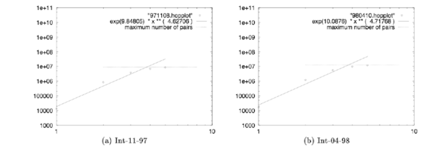
$P(0) = N$

$P(1) = N + 2E$

$P(\delta) = N^2$

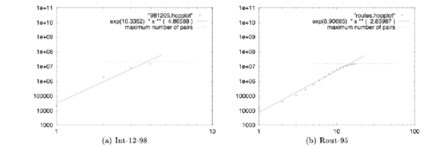
where δ is the diameter of the graph

Plot P(h) versus h on log-log scale



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

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.



(a) Int-12-98 (b) Rout-95

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.

$$P(h) \propto h^{\mathcal{H}}, h \ll \delta$$

$$P(h) = \begin{cases} (N+2E)h^{\mathcal{H}} & h \ll \delta \\ N^2 & h \geq \delta \end{cases}$$

$$P(\delta_{ef}) = N^2$$

IX TELECOMMUNICATIONS FORUM TELFOR 2001, Belgrade, 20-22.11.2001.

MODELING PEER-TO-PEER NETWORK TOPOLOGIES THROUGH "SMALL-WORLD" MODELS AND POWER LAWS

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

$$\delta_{ef} = \left(\frac{N^2}{N+2E} \right)^{1/\mathcal{H}}$$

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

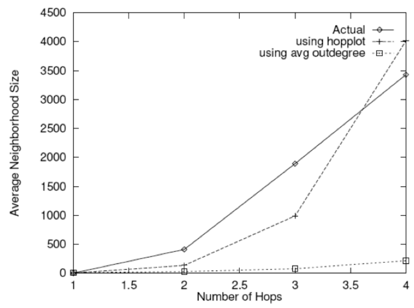
Gnutella is a highly dynamic network in which topology is constantly changing as hosts join and leave the network, establish new connections, and close the existing ones.

Average Number of Nodes within h Hops

$$NN(h) = \frac{P(h) - N}{N}$$

Average Number of Nodes within h Hops
(using average degree)

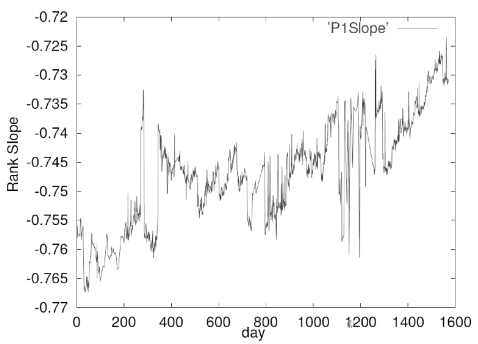
$$NN(h) = \bar{d}(\bar{d} - 1)^h$$

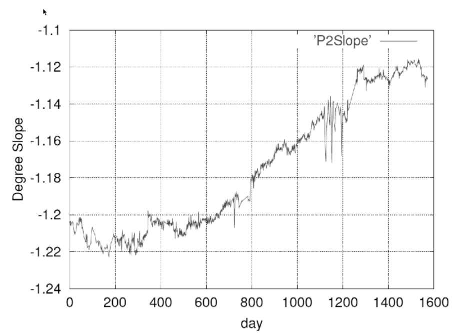


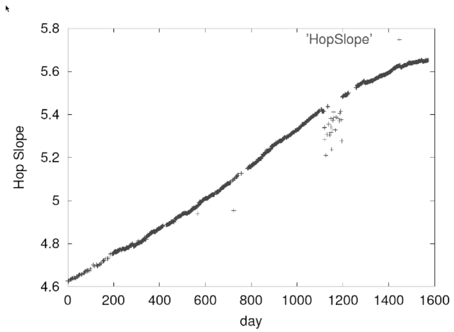
$$d_v \propto r_v^{\mathcal{R}}$$
$$f_d \propto d^{\mathcal{O}}$$
$$P(h) \propto h^{\mathcal{H}}, h \ll \delta$$

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?

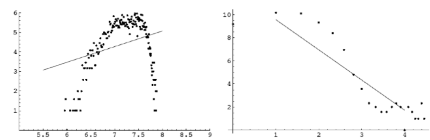


Figure 1: Log-log plot of frequency f_d vs. outdegree d for a 5000-node Waxman topology (left) and a 6660-node Transit-Stub topology (right). The correlation coefficient is 0.4 for the Waxman topology, and 0.9 for the Transit-Stub topology.

How to generate topology that follows power laws?

Where does power law comes 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 8×10^8 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

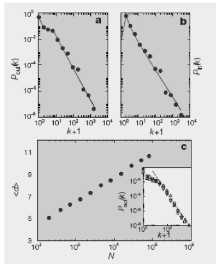
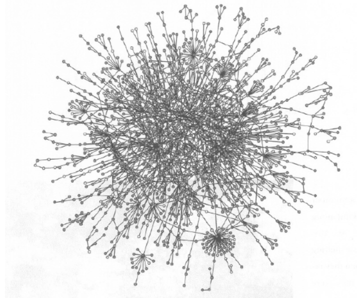


Figure 1 Distribution of links on the World-Wide Web. **a,** Outgoing links (URLs found on an HTML document); **b,** incoming links (URLs pointing to a certain HTML document). Data were obtained from the complete map of the nd.edu domain, which contains 325,729 documents and 1,469,680 links. Dotted lines represent analytical

Scientific American, "Scale-Free Networks", May 2003



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 analy-

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

Examples of Scale-Free Networks

NETWORK	NODES	LINKS
Cellular metabolism	Molecules involved in burning food for energy	Participation in the same biochemical reaction
Hollywood	Actors	Appearance in the same movie
Internet	Routers	Optical and other physical connections
Protein regulatory network	Proteins that help to regulate a cell's activities	Interactions among proteins
Research collaborations	Scientists	Co-authorship of papers
Sexual relationships	People	Sexual contact
World Wide Web	Web pages	URLs

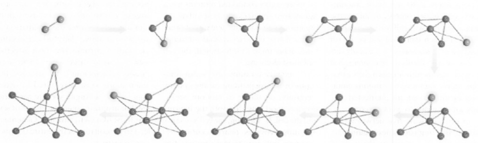
Now it has more than three networks have expanded si lywood had only a handfu 1890, but as new people joi the network grew to incluc half a million, with the roo ing to veteran actors. The only a few routers about t ago, but it gradually grew lions, with the new routers a to those that were already p work. Thanks to the grow real networks, older node: opportunities to acquire lin

Furthermore, all nodes When deciding where to li page, people can choose fron locations. Yet most of us are only a tiny fraction of the f that subset tends to include nected sites because they are By simply linking to those exercise and reinforce a bias This process of "preferentia occurs elsewhere. In Hollyw

54 SCIENTIFIC AMERICAN

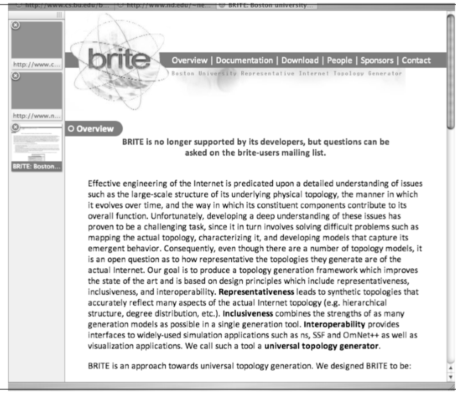
BIRTH OF A SCALE-FREE NETWORK

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.

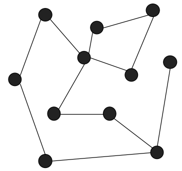


Generating Power Law Topology (simplified)

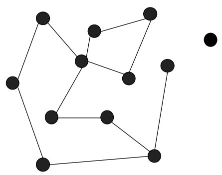
“On the Original of Power Laws in Internet Topologies”
A Medina, I Matta, J Byers,
ACM SIGCOMM, ‘00



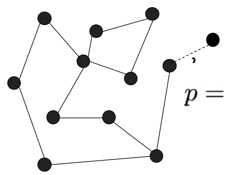
Randomly generate a small graph



Incremental Growth:
Add one node at a time



Preferential Attachment:
Connects to a neighbor i with a probability



$$p = \frac{d_i}{\sum_{j \in C} d_j}$$

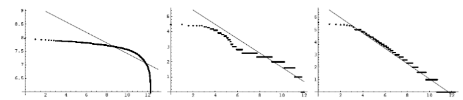
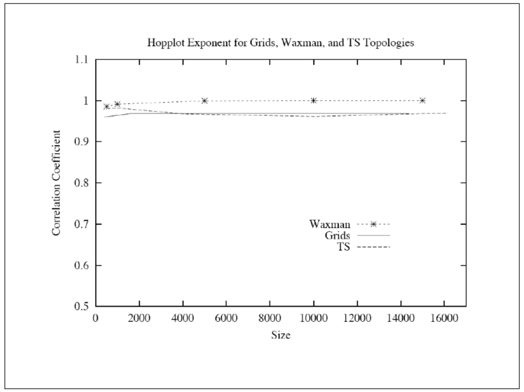


Figure 6: Log-log plot of outdegree d_v vs. rank for a 5000-node Waxman topology (left), a 4040-node Transit-Stub topology (middle) and a 5000-node BRITE topology with preferential connectivity and incremental growth (right). The correlation coefficient is 0.81 for the Waxman topology, 0.87 for the Transit-Stub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



Hopplot Exponent for Grids, Waxman, and TS Topologies
Correlation Coefficient
Size
1.1
1
0.9
0.8
0.7
0.6
0.5
0
2000
4000
6000
8000
10000
12000
14000
16000
Waxman
Grids
TS