The Shape of the Internet

Slides assembled by
Jeff Chase
Duke University
(thanks to Vishal Misra and C. Faloutsos)
The Shape of the Network

Characterizing “shape”:
- AS-level topology: who connects to whom
- Router-level topology: what connects with what
- POP-level topology: where connects with where

Why does it matter?
- Survivability/robustness to node/POP/AS failure
- Path lengths / diameter
- Congestion / hot spots / bottlenecks
- Redundancy

Star? Tree? Mesh? Random?
Why study topology?

- **Correctness** of network protocols typically independent of topology
- **Performance** of networks critically dependent on topology
  - e.g., convergence of route information
- Internet **impossible** to replicate
- **Modeling of topology** needed to generate test topologies
Internet topologies

Router level

Autonomous System (AS) level

Vishal Misra
More on topologies..

• Router level topologies reflect physical connectivity between nodes
  - Inferred from tools like traceroute or well known public measurement projects like Mercator and Skitter

• AS graph reflects a peering relationship between two providers/clients
  - Inferred from inter-domain routers that run BGP and public projects like Oregon Route Views

• Inferring both is difficult, and often inaccurate

Vishal Misra
Early work

- Early models of topology used variants of Erdos-Renyi random graphs
  - Nodes randomly distributed on 2-dimensional plane
  - Nodes connected to each other w/ probability inversely proportional to distance
- Soon researchers observed that random graphs did not represent real world networks
Real world topologies

• Real networks exhibit
  - Hierarchical structure
  - Specialized nodes (transit, stub..)
  - Connectivity requirements
  - Redundancy
• Characteristics incorporated into the Georgia Tech Internetwork Topology Models (GT-ITM) simulator (E. Zegura, K. Calvert and M.J. Donahoo, 1995)
So...are we done?

- No!
- In 1999, Faloutsos, Faloutsos and Faloutsos published a paper, demonstrating **power law relationships** in Internet graphs.
- Specifically, the **node degree distribution** exhibited power laws.

That Changed Everything.....

Vishal Misra
Power laws in AS level topology

Vishal Misra
AS graph is “scale-free”

• Power law in the AS degree distribution [SIGCOMM99]

internet domains

log(degree)  att.com  log(rank)

ibm.com

-0.82

C. Faloutsos
Power Laws

- Faloutsos\(^3\) (Sigcomm’99)
  - frequency vs. degree
  - empirical ccdf
  \[ P(d>x) \sim x^{-\alpha} \]

\[ \alpha \approx 1.15 \]

topology from BGP tables

Vishal Misra
GT-ITM abandoned..

- GT-ITM did not give power law degree graphs
- New topology generators and explanation for power law degrees were sought
- Focus of generators to match degree distribution of observed graph

Vishal Misra
Generating power law graphs

**Goal:** construct network of size $N$ with degree power law, $P(d>x) \sim x^{-\alpha}$

- power law random graph (PLRG) (Aiello et al)
- Inet (Chen et al)
- incremental growth (BA) (Barabasi et al)
- general linear preference (GLP) (Bu et al)

Vishal Misra
Barabasi model: fixed exponent

- Incremental growth
  - Initially, $m_0$ nodes
  - Step: add new node $i$ with $m$ edges
- Linear preferential attachment
  - Connect to node $i$ with probability
    \[ \prod(k_i) = \frac{k_i}{\sum k_j} \]

May contain multi-edges, self-loops

Vishal Misra
“Scale-free” graphs

- Preferential attachment leads to “scale free” structure in connectivity
- Implications of “scale free” structure
  - Few centrally located and highly connected hubs
  - Network robust to random attack/node removal (probability of targeting hub very low)
  - Network susceptible to catastrophic failure by targeted attacks (“Achilles heel of the Internet” Albert, Jeong, Barabasi, Nature 2000)
Is the router-level Internet graph scale-free?

- No...(There is no Memphis!)
- Emphasis on degree distribution - structure ignored
- Real Internet very structured
- Evolution of graph is highly constrained
Topology constraints

• Technology
  - Router out degree is constrained by processing speed
  - Routers can either have a large number of low bandwidth connections, or..
  - A small number of high bandwidth connections

• Geography
  - Router connectivity highly driven by geographical proximity

• Economy
  - Capacity of links constrained by the technology that nodes can afford, redundancy/performance they desire etc.

Vishal Misra
Network and graph mining

- Friendship Network [Moody ’01]
- Food Web [Martinez ’91]
- Protein Interactions [genomebiology.com]

Graphs are everywhere!

C. Faloutsos
Network and graph mining

- How does the Internet look like?
- How does the web look like?
- What constitutes a ‘normal’ social network?
- What is the ‘network value’ of a customer?
- Which gene/species affects the others the most?

C. Faloutsos
Why

Given a graph:

- which node to market-to / defend / immunize first?
- Are there un-natural subgraphs? (e.g., criminals’ rings)?

[from Lumeta: ISPs 6/1999]

C. Faloutsos
Patterns?

- avg degree is, say 3.3
- pick a node at random - guess its degree, exactly (→ “mode”)

avg: 3.3

degree

C. Faloutsos
Patterns?

- avg degree is, say 3.3
- pick a node at random - guess its degree, exactly (-> "mode")
- A: 1!!

C. Faloutsos
Patterns?

- Avg degree is, say 3.3
- Pick a node at random - what is the degree you expect it to have?
- A: 1!!
- $A'$: very skewed distr.
- Corollary: the mean is meaningless!
- (and std $\rightarrow$ infinity (!))

C. Faloutsos
Power laws - discussion

• do they hold, over time?
  • Yes! for multiple years [Siganos+]
  • do they hold on other graphs/domains?
  • Yes!
  - web sites and links [Tomkins+], [Barabasi+]
  - peer-to-peer graphs (gnutella-style)
  - who-trusts-whom (epinions.com)
Time Evolution: rank $R$

- The rank exponent has not changed! [Siganos+]

C. Faloutsos
The Peer-to-Peer Topology

- Number of immediate peers (= degree), follows a power-law

[C. Faloutsos]

Graph: Gnutella snapshot from Dec. 28, 2000 (\(|\rho|=0.94\))
epinions.com

- who-trusts-whom
  [Richardson + Domingos, KDD 2001]

Count

(out) degree

C. Faloutsos
Why care about these patterns?

• better graph generators [BRITE, INET]
  - for simulations
  - extrapolations
• ‘abnormal’ graph and subgraph detection

C. Faloutsos
Even more power laws:

library science (Lotka’s law of publication count); and
citation counts: (*citeseer.nj.nec.com* 6/2001)
Even more power laws:

- web hit counts [w/ A. Montgomery]

C. Faloutsos
Power laws, cont'd

- In- and out-degree distribution of web sites [Barabasi], [IBM-CLEVER]

\[
\log \text{ indegree} = -\log(\text{freq})
\]

from [Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, Andrew Tomkins]

C. Faloutsos
Power laws, cont’d

- In- and out-degree distribution of web sites
  [Barabasi], [IBM-CLEVER]

from [Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, Andrew Tomkins]

C. Faloutsos
Mapping the Internet

• At this point in the session, we discussed the SIGCOMM 2002 RocketFuel paper, based on slides in pdf form from Neil Spring.