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?

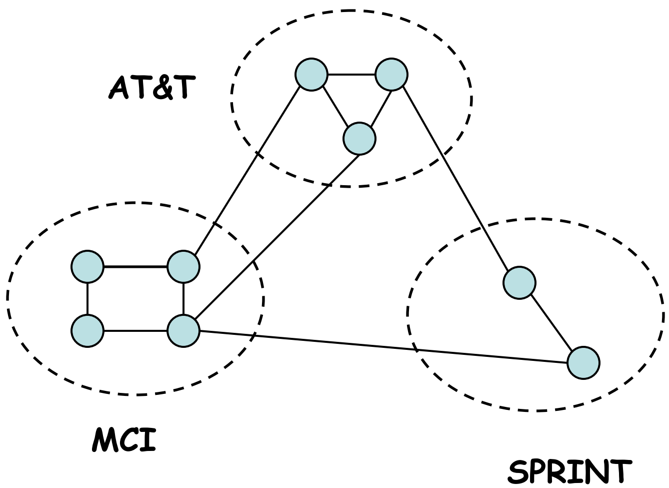
- Survivability/robustness to node/POP/AS failure
- Path lengths / diameter
- Congestion / hot spots / bottlenecks
- Redundancy

Star? Tree? Mesh? Random?

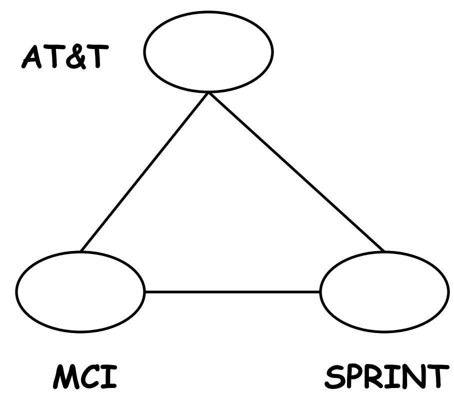
Why study topology?

- 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

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

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

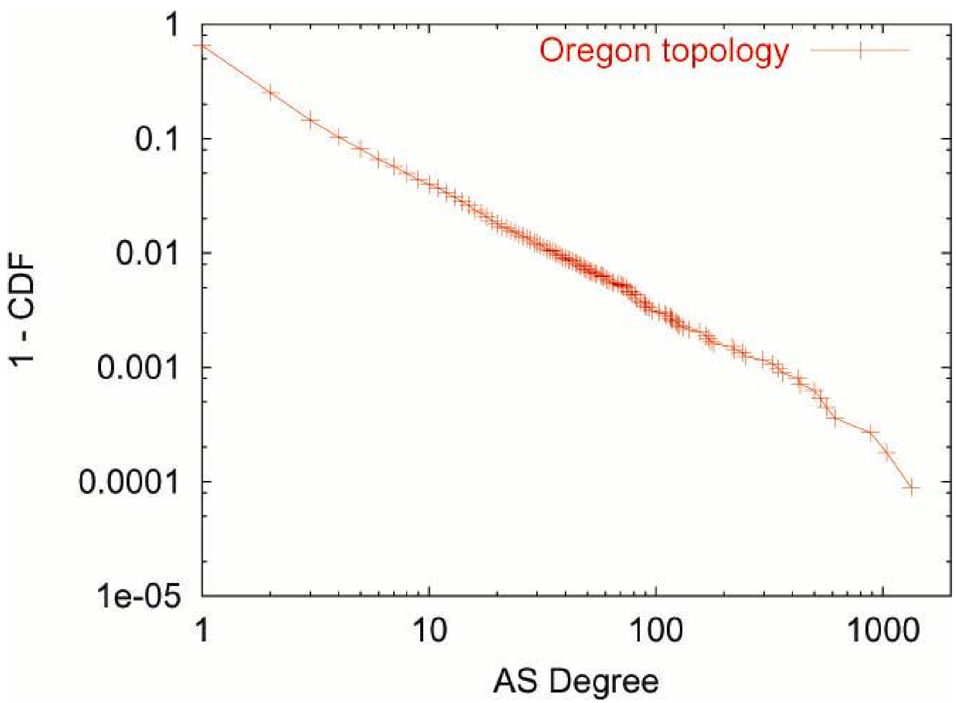
- Real networks exhibit
  - Hierarchical structure
  - Specialized nodes (transit, stub..)
  - Connectivity requirements
  - Redundancy
- Characteristics incorporated into the Georgia Tech Internetwork Topology Models (GT-ITM) simulator (E. Zegura, K.Calvert and M.J. Donahoo, 1995)

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…..

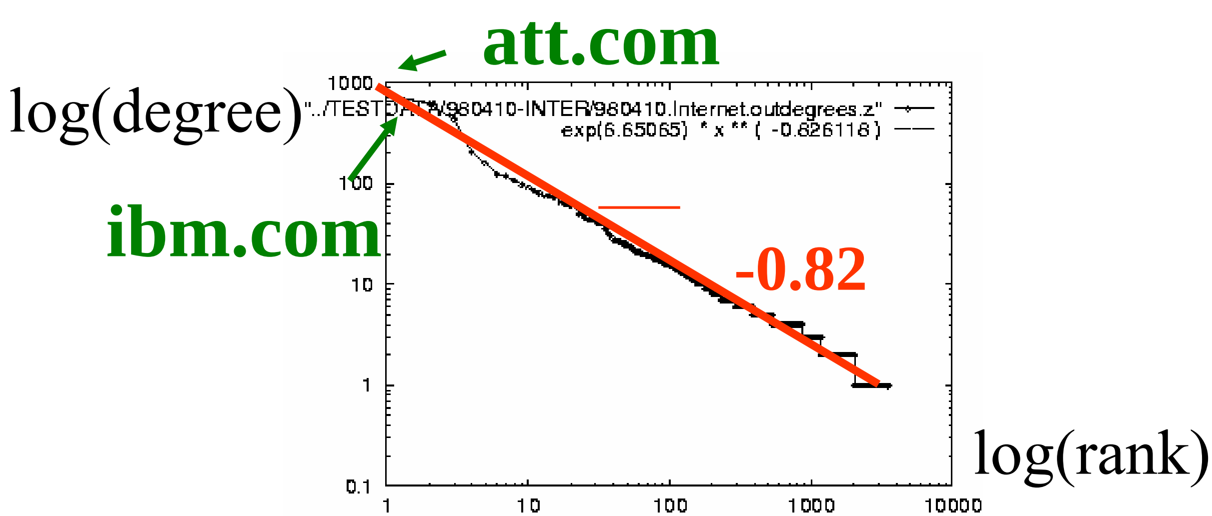
Power laws in AS level topology



AS graph is "scale-free"

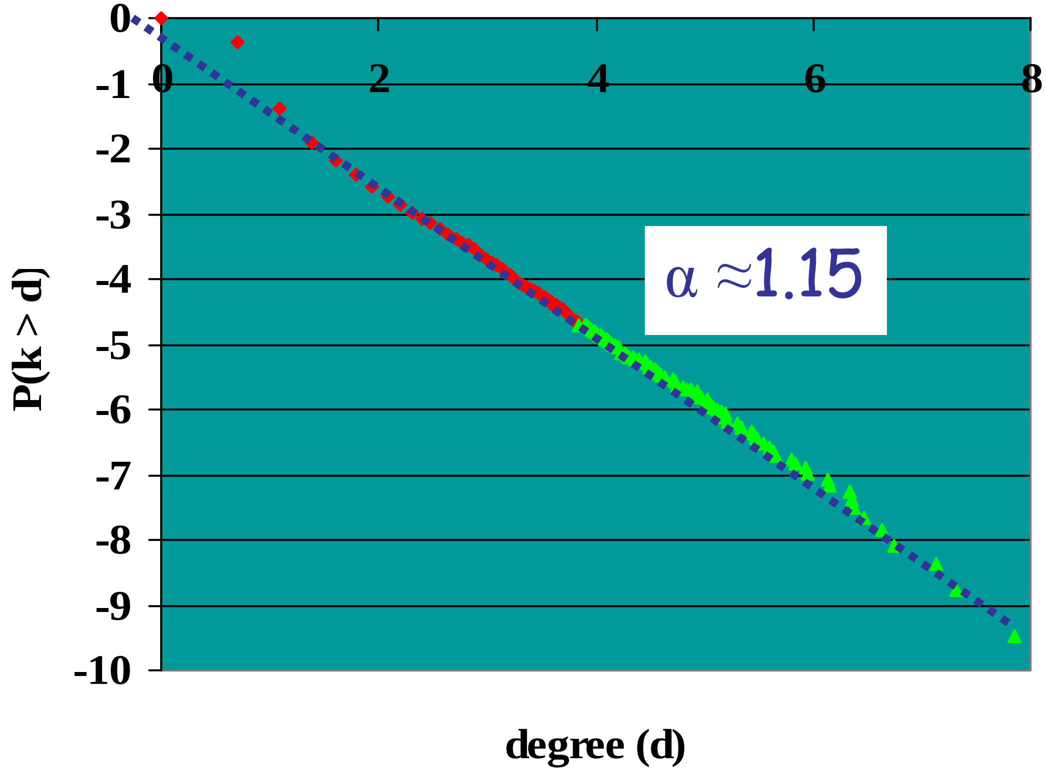
- Power law in the AS degree distribution [SIGCOMM99]

internet domains



Power Laws

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



topology from BGP tables

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

Generating power law graphs

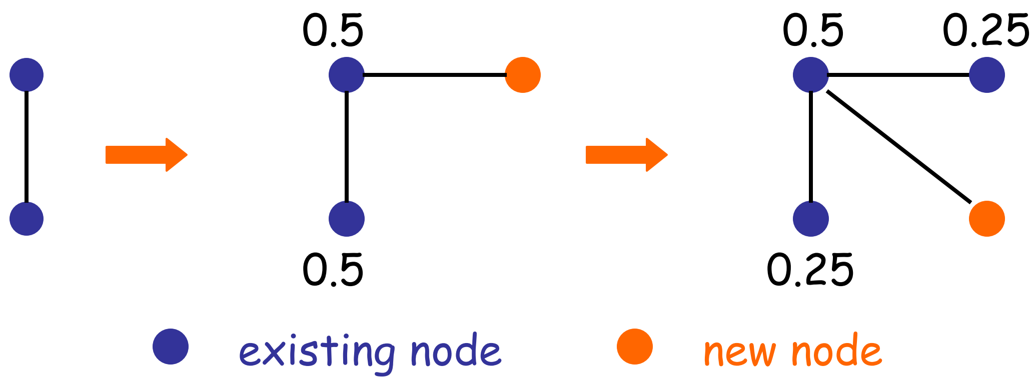
Goal: construct network of size N with degree power law, $P(d>x) \sim x^{-\alpha}$

- 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)

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

$$\Pi(k_i) = k_i / \sum k_j$$



may contain multi-edges, self-loops

“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)

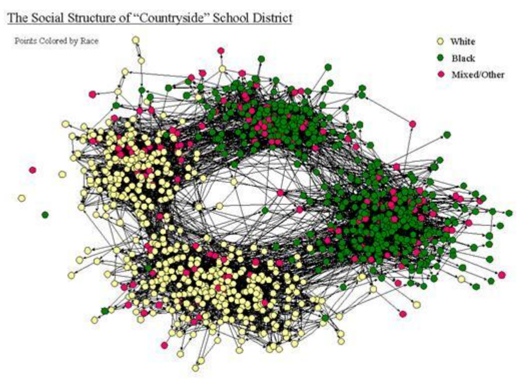
Is the router-level Internet graph scale-free?

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

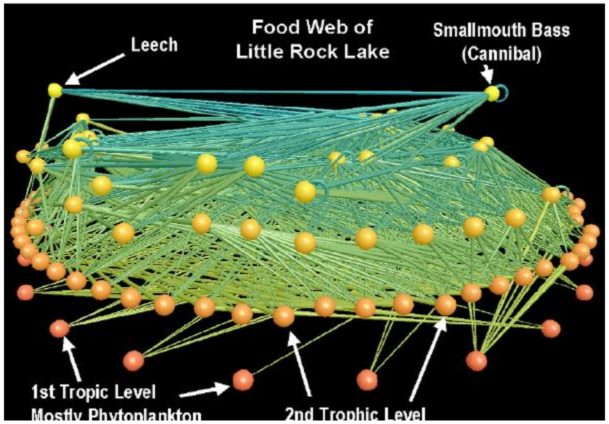
Topology constraints

- 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.

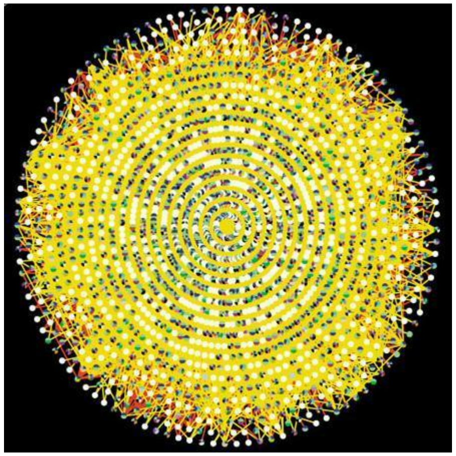
Network and graph mining



Friendship Network
[Moody '01]



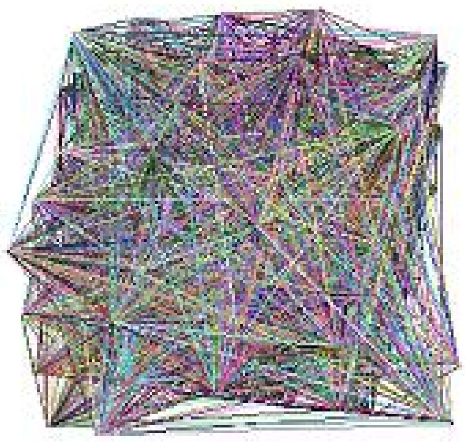
Food Web
[Martinez '91]



Protein Interactions
[genomebiology.com]

Graphs are everywhere!

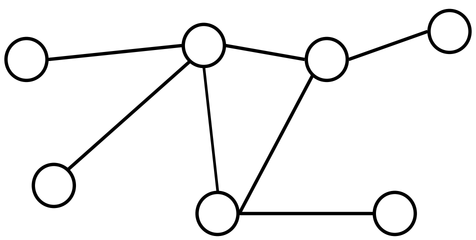
Network and graph mining

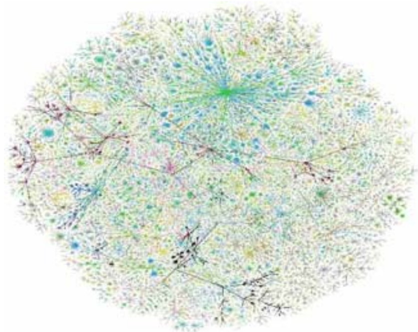


- How does the Internet look like?
- How does the web look like?
- What constitutes a 'normal' social network?
- What is the 'network value' of a customer?
- which gene/species affects the others the most?

Why

Given a graph:



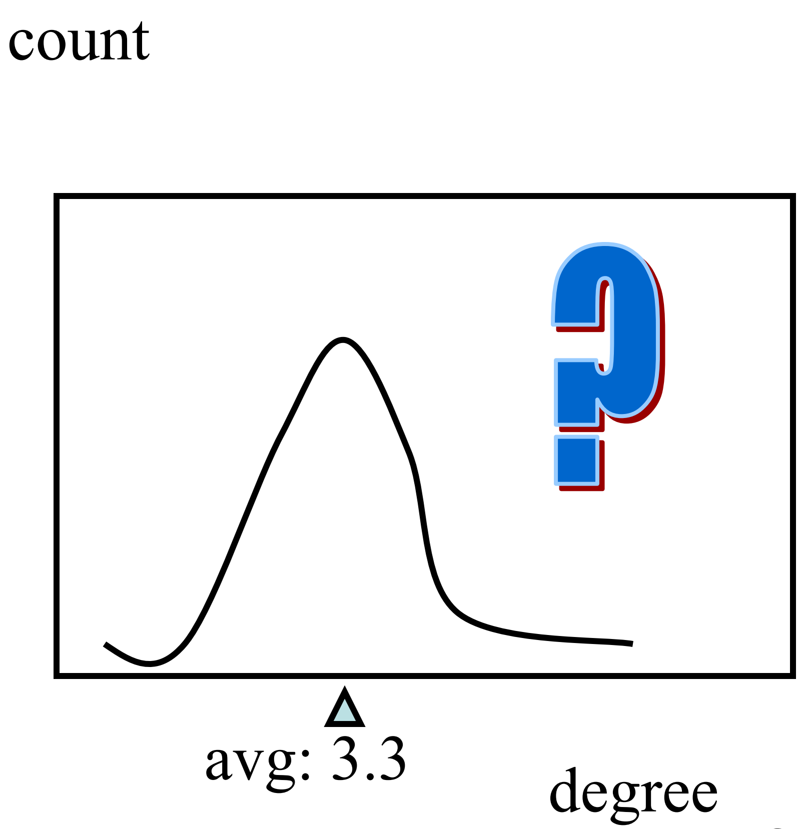


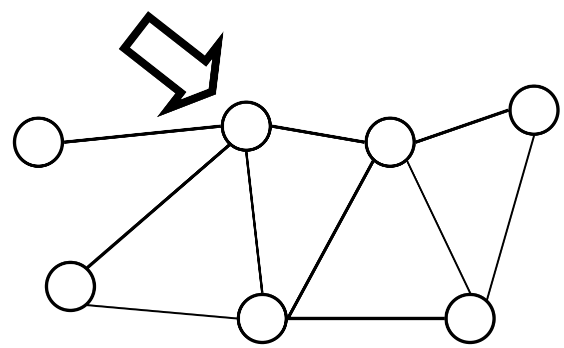
- which node to market-to / defend / immunize first?
- Are there un-natural sub-graphs? (eg., criminals' rings)?

[from Lumeta: ISPs 6/1999]

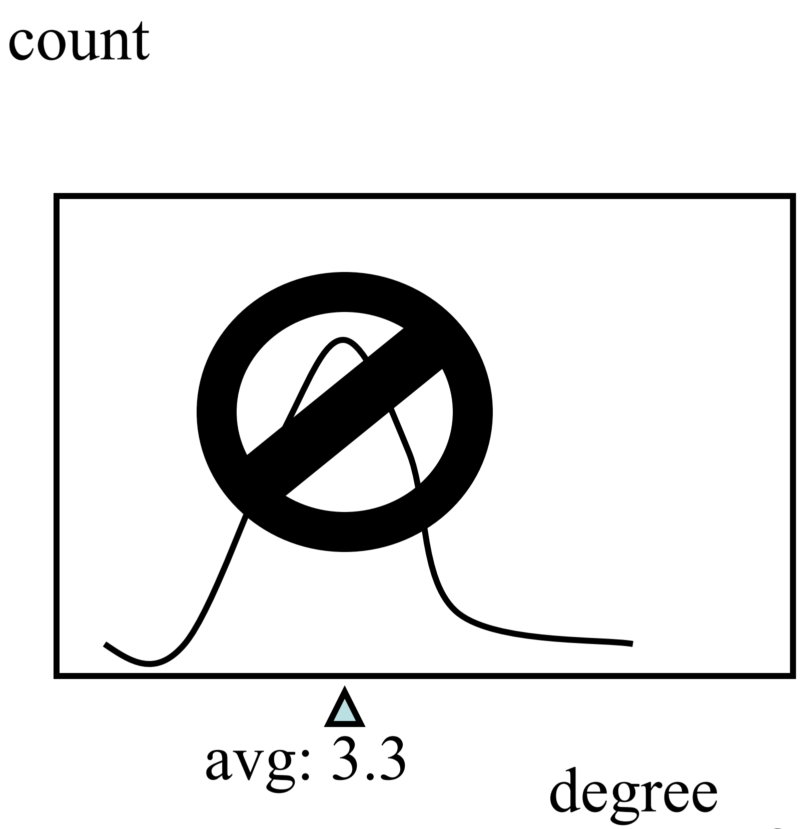
Patterns?

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



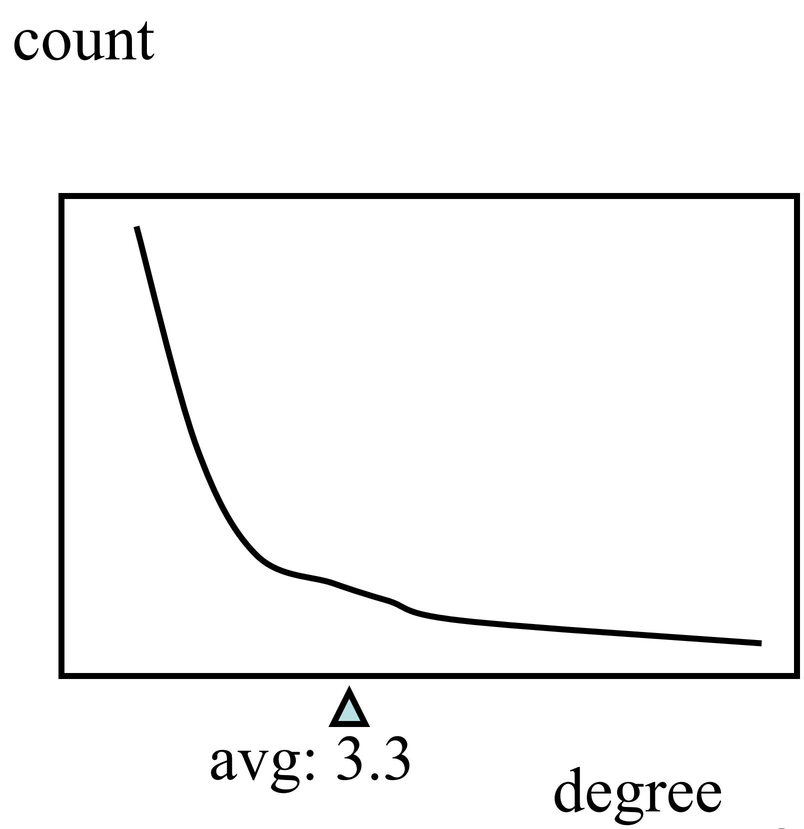


Patterns?



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

Patterns?

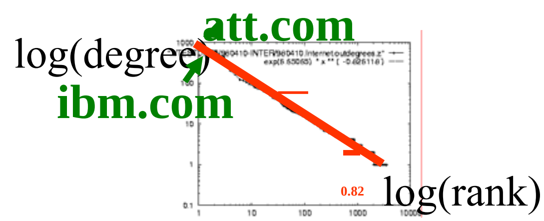


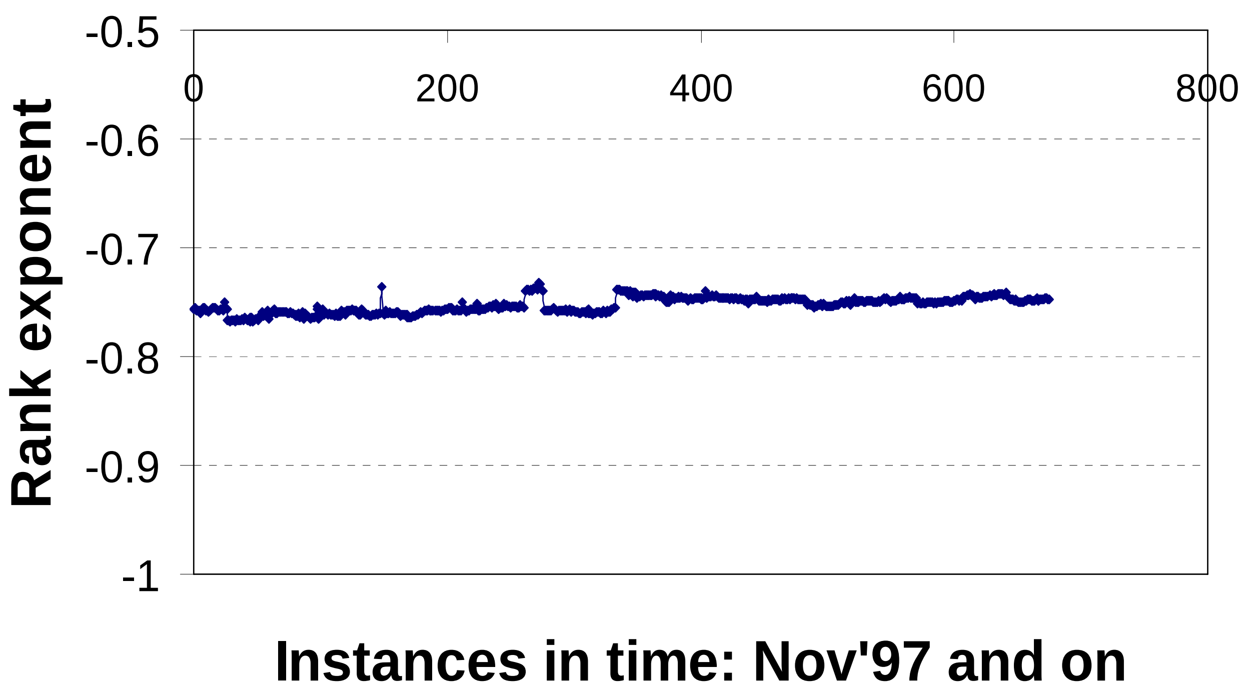
- avg degree is, say 3.3
- pick a node at random - what is the degree you expect it to have?
- *A*: 1!!
- *A'*: very skewed distr.
- Corollary: **the mean is meaningless**!
- (and std -> infinity (!))

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*



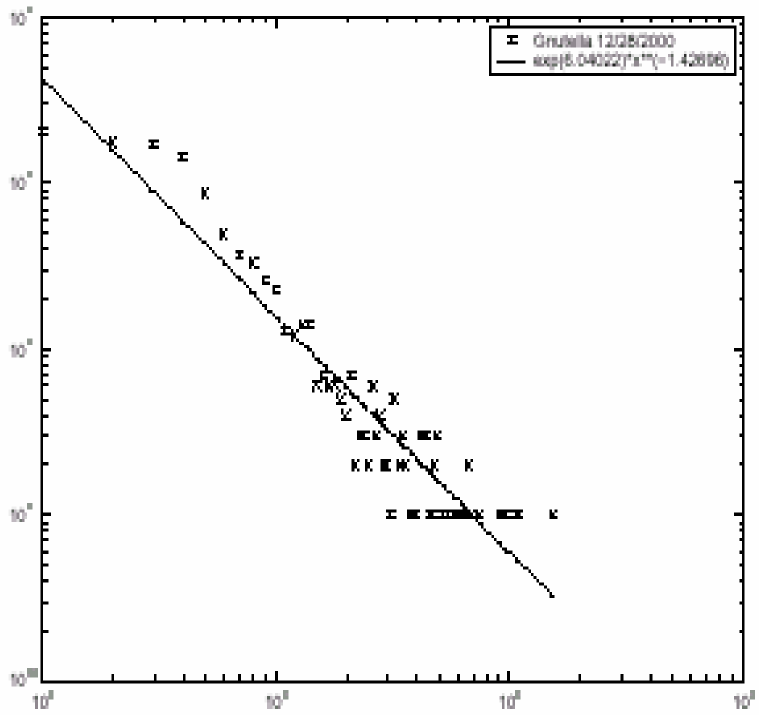


Domain level

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

The Peer-to-Peer Topology

count



(a) Gnutella snapshot from Dec. 28, 2000 (|r|=0.94)

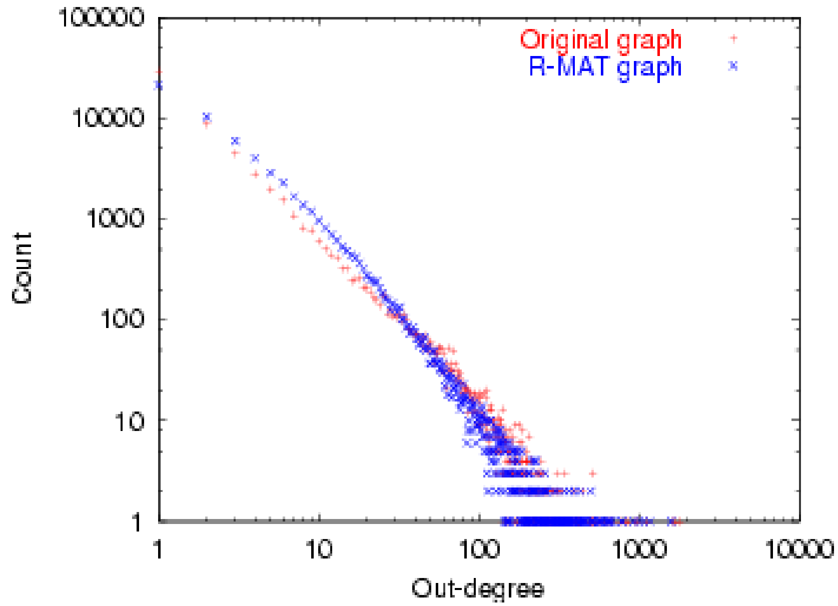
[Jovanovic+]

degree

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

epinions.com

count

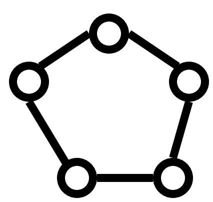


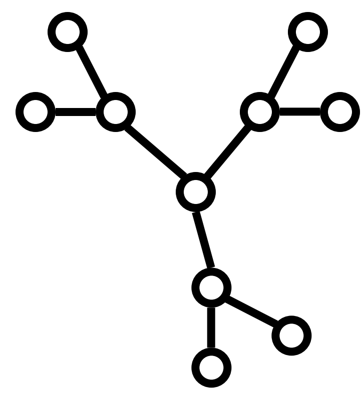
(out) degree

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

Why care about these patterns?

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

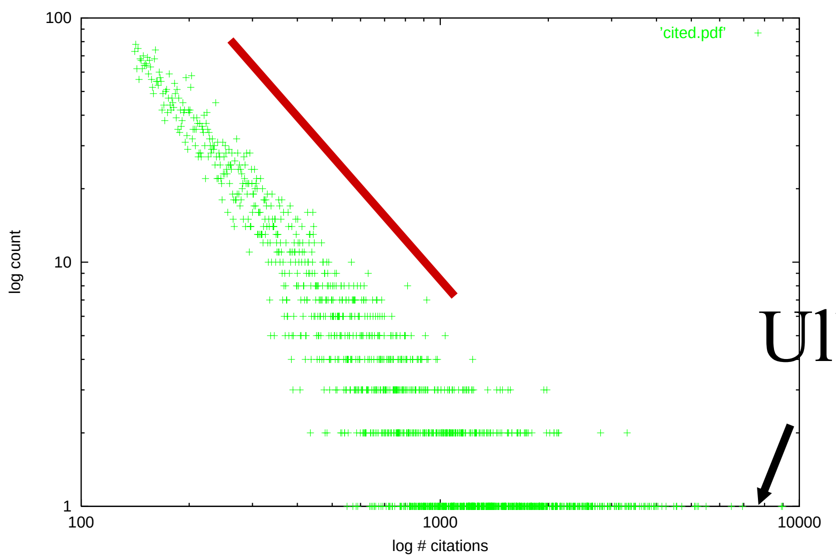




Even more power laws:

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

log(count)



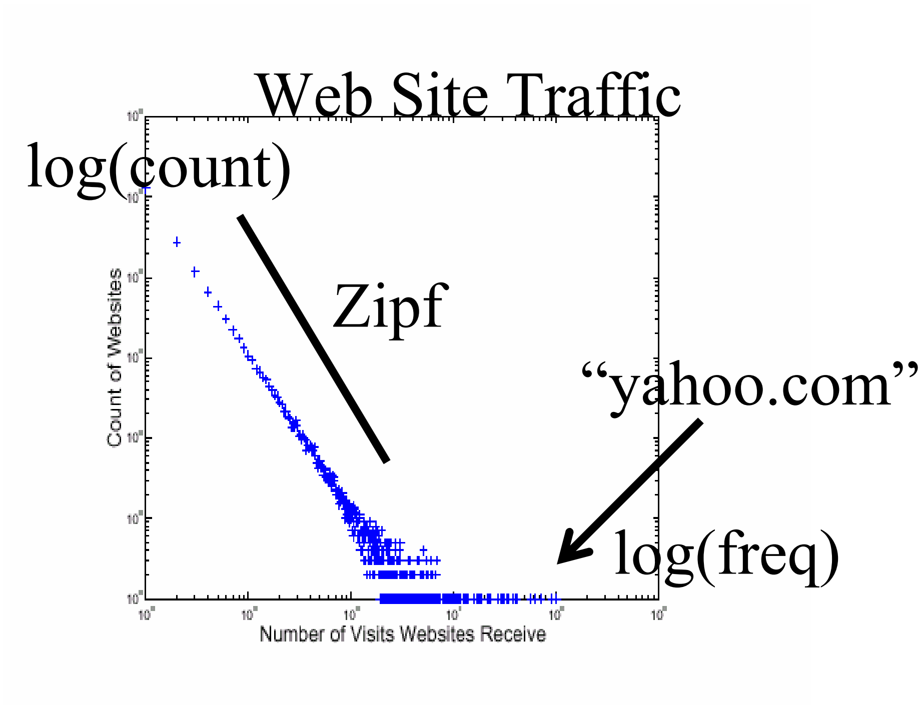
Ullman

log(#citations)

C. Faloutsos

Even more power laws:

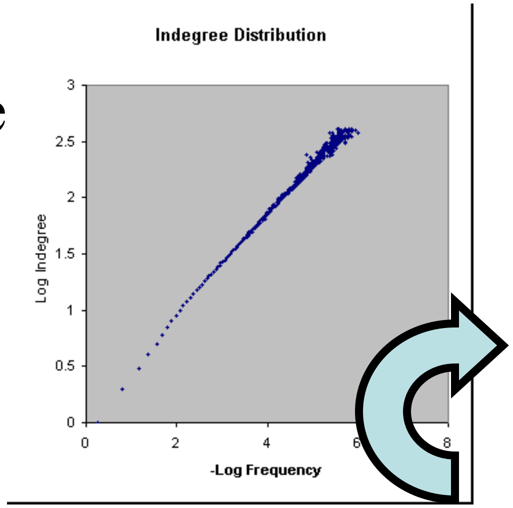
- web hit counts [w/ A. Montgomery]



Power laws, cont'd

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

log indegree



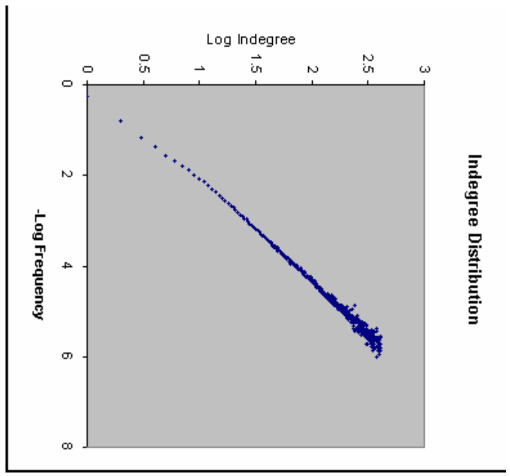
- log(freq)

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

Power laws, cont'd

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

log(freq)



log indegree

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

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.

www.cs.umd.edu/~nspring/talks/sigcomm-rocketfuel.pdf