Understanding Router-level Topology: Principles, Models, and Validation

David Alderson
California Institute of Technology

ISMA Workshop on Internet Topology

May 10, 2006

Acknowledgments

Primary Coauthors

- John Doyle (Caltech)
- Walter Willinger (AT&T Labs-Research)
- Lun Li (Caltech)

Contributions

- Reiko Tanaka (RIKEN, Japan)
- Matt Roughan (U. Adelaide, Australia)
- Steven Low (Caltech)
- Ramesh Govindan (USC)
- Neil Spring (U. Maryland)
- Stanislav Shalunov (Abilene)
- Heather Sherman (CENIC)

On modeling

"All models are wrong,
but some models are useful."
- G. P. E. Box

"When exactitude is elusive, it is better to be approximately right than certifiably wrong."
- B. B. Mandelbrot

the Internet as an inspiration for the development of elegant mathematical models of networks

vs

wanting to say “something meaningful” about the Internet (something about which decision makers are concerned)

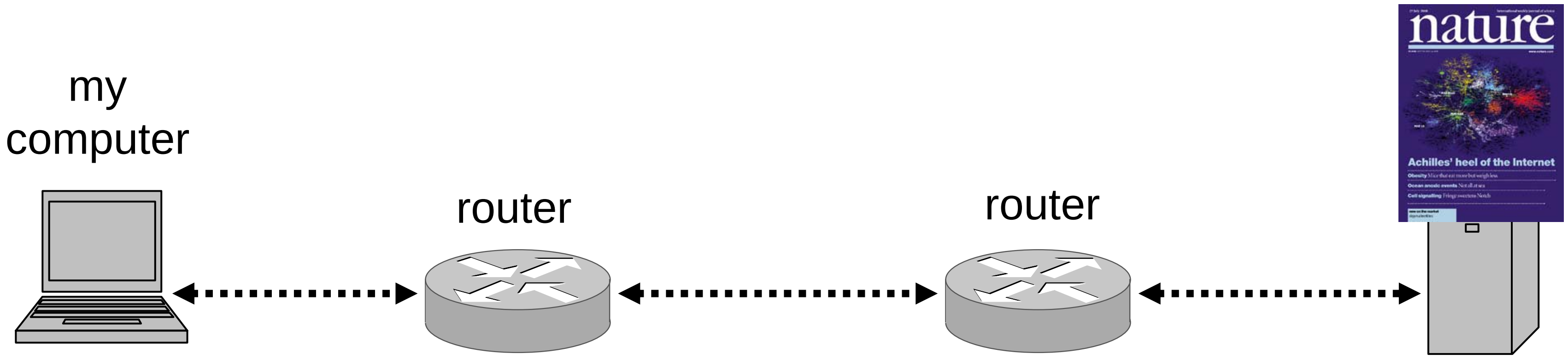
what is the MESSAGE? what “MATTERS”? and TO WHOM?

who has RESPONSIBILTY for the message?

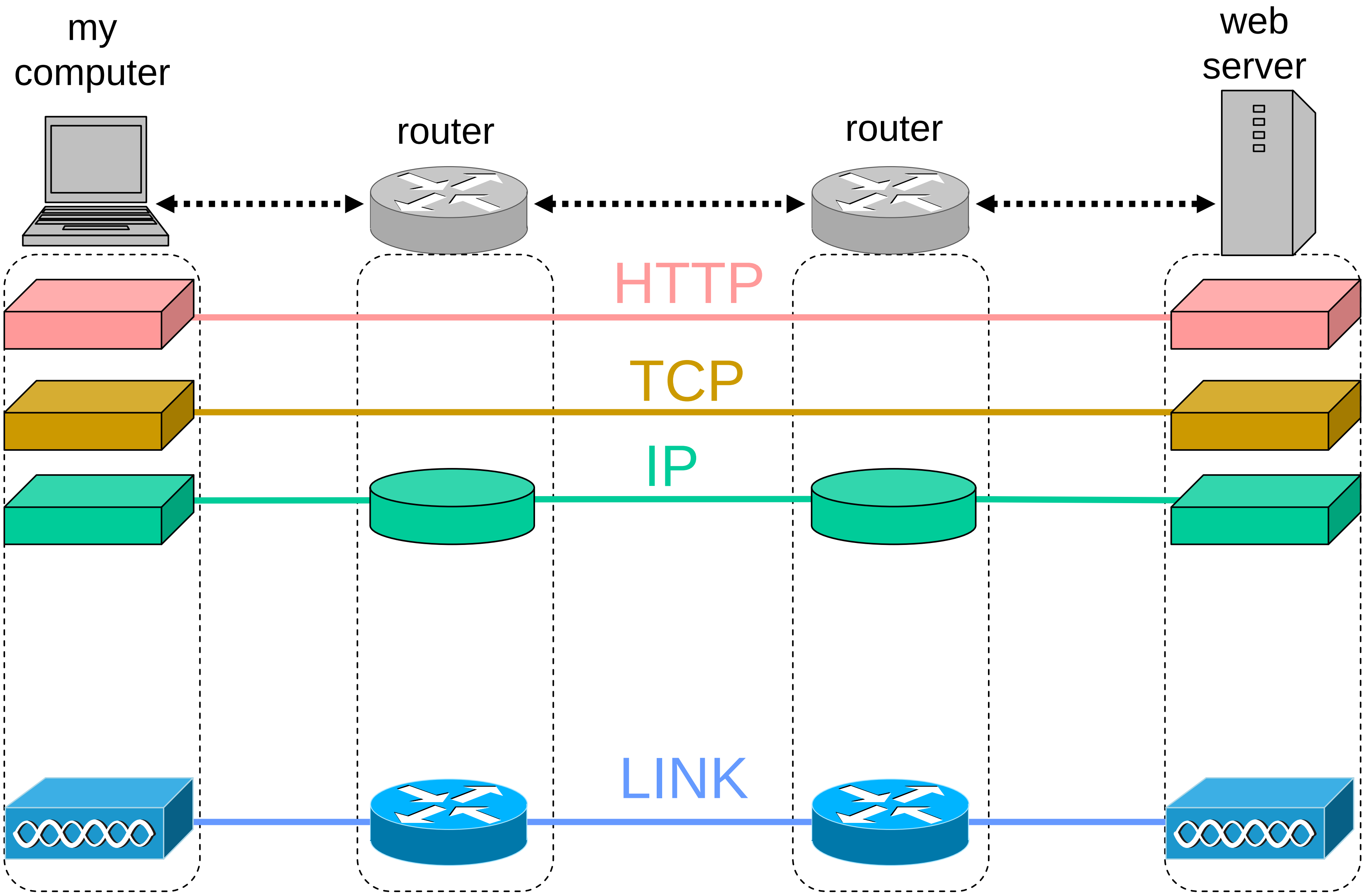
The application of graph theory and statistics to the study of Internet topology without the details of system architecture and engineering can lead to incorrect (and possibly misleading) conclusions.

Let’s consider the router-level Internet

The Router-Level Internet

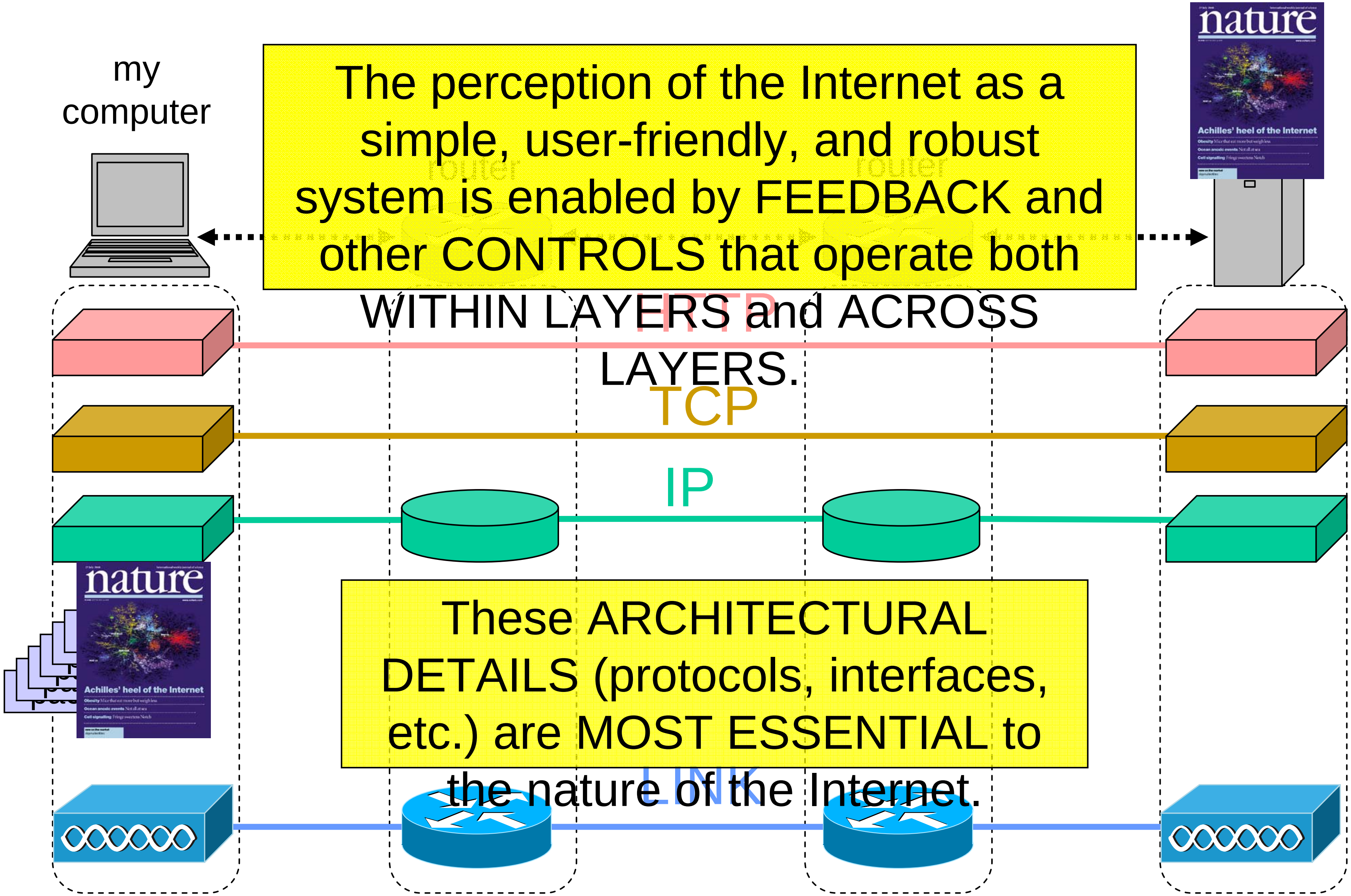


The Internet is a LAYERED Network



The Internet is a LAYERED Network

The perception of the Internet as a simple, user-friendly, and robust system is enabled by FEEDBACK and other CONTROLS that operate both WITHIN LAYERS and ACROSS LAYERS.



These ARCHITECTURAL DETAILS (protocols, interfaces, etc.) are MOST ESSENTIAL to the nature of the Internet.

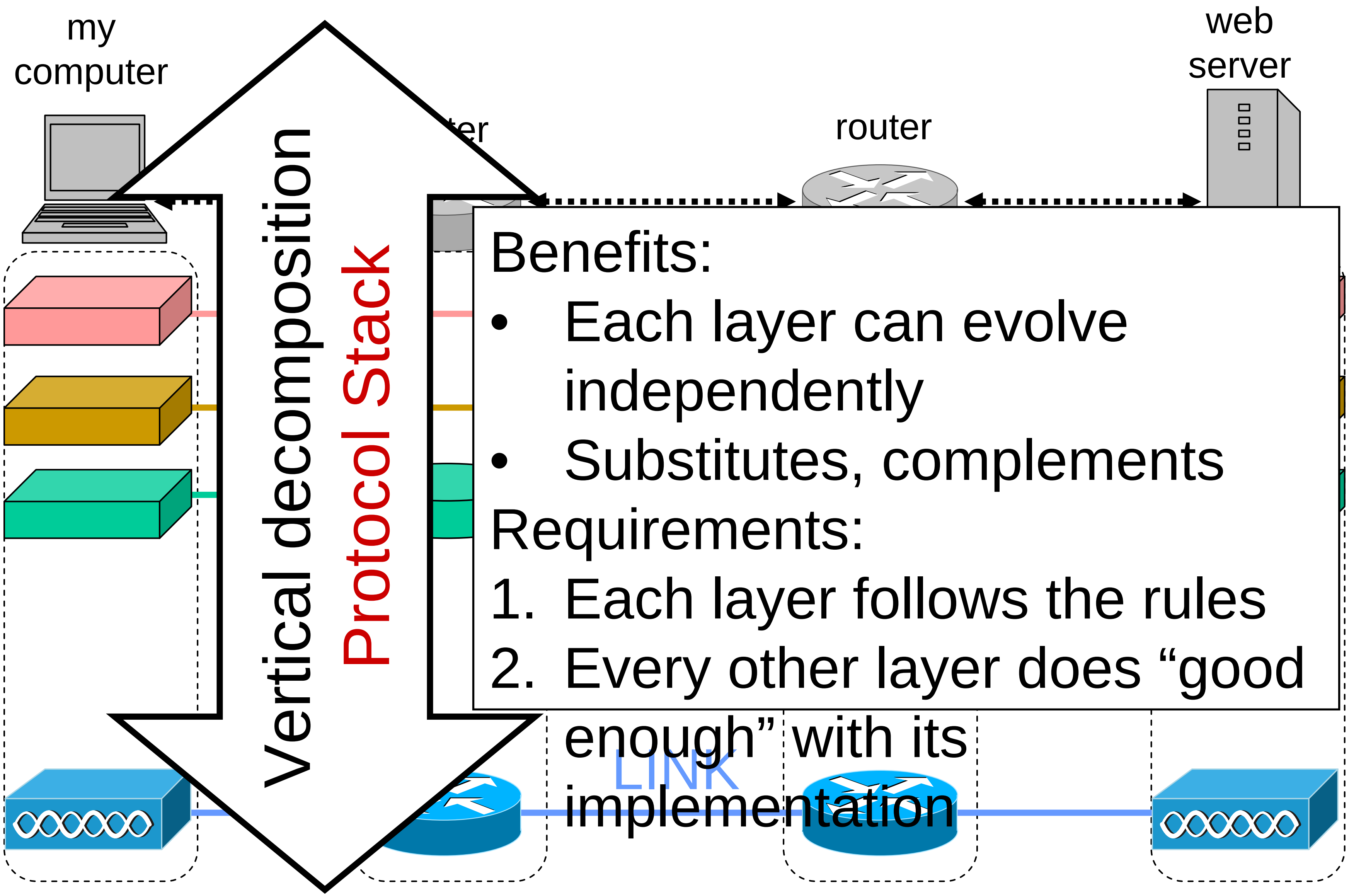
Internet structure can be viewed as a solution to a DESIGN problem

- **physical constraints on components**
  - **distance/delay, capacity**
- **functional constraints on the system as a whole**
  - **"X-ities": functionality, maintainability, adaptability, evolvability, etc.**

design approach: *modularity*

- **simplify the problem by breaking it up**
- **but still with provable properties as if it were an integrated whole**

Internet Architecture: Dual Decomposition



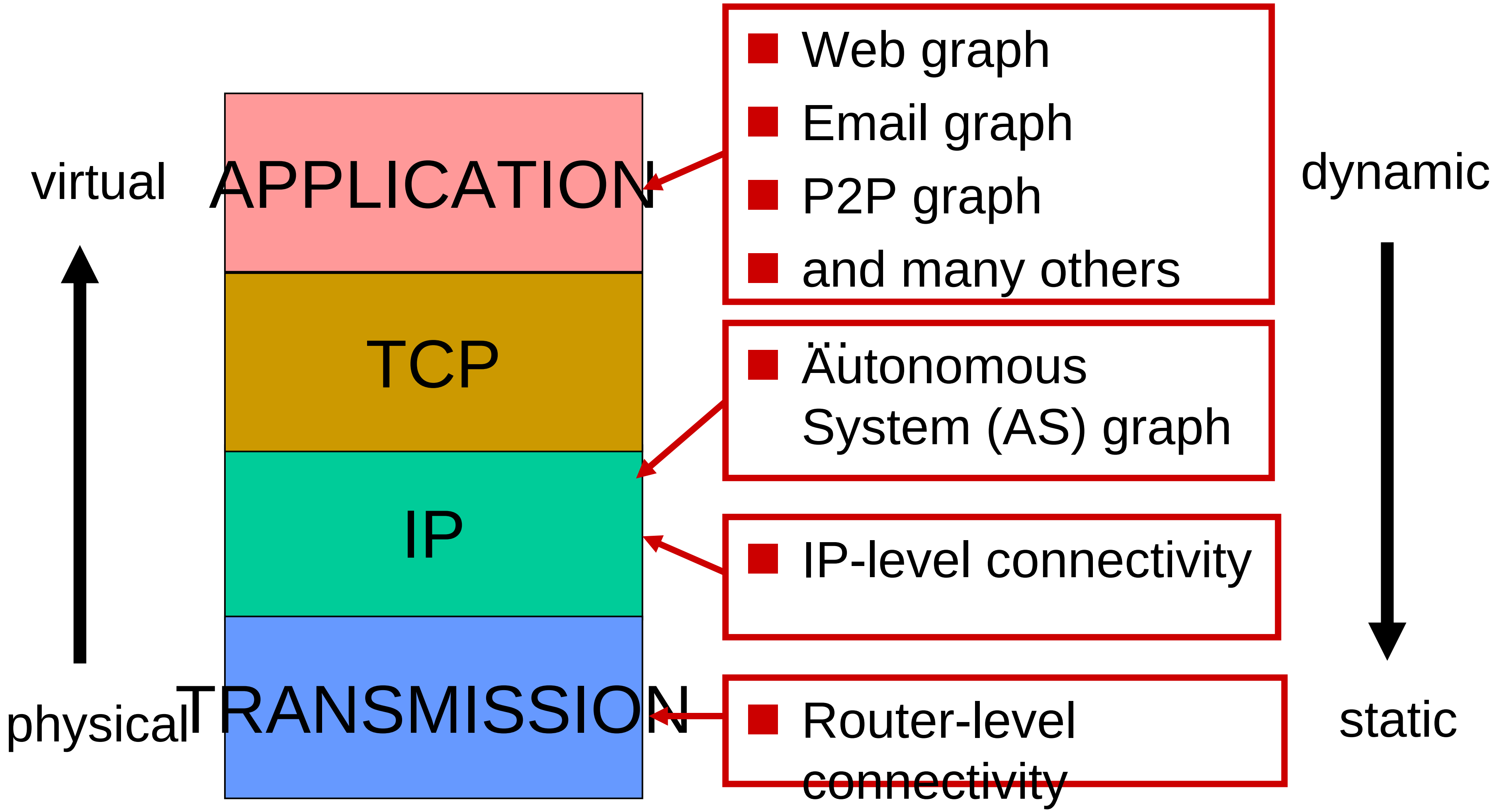
Benefits:

- Each layer can evolve independently
- Substitutes, complements

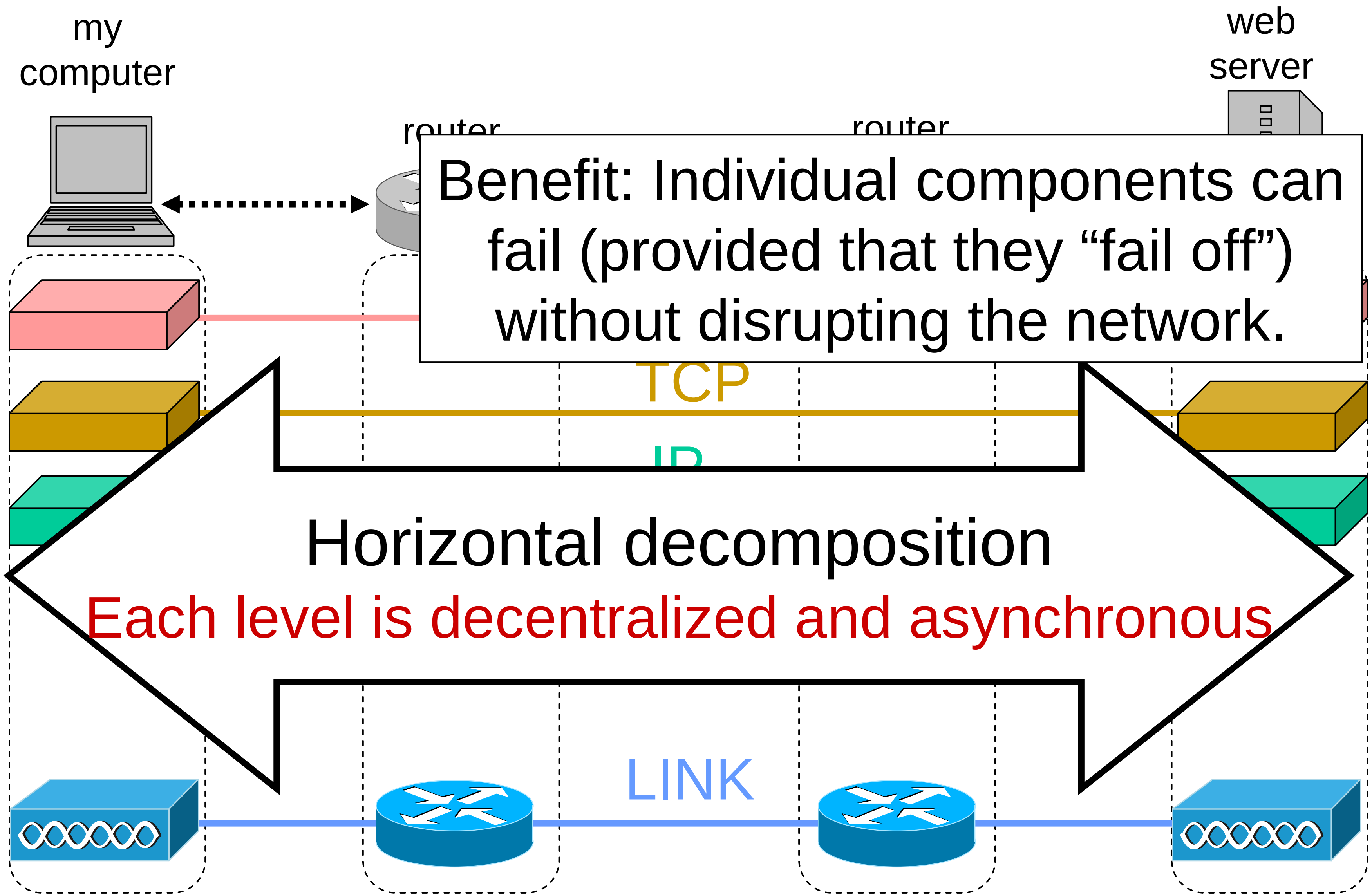
Requirements:

1. Each layer follows the rules
2. Every other layer does "good enough" with its implementation

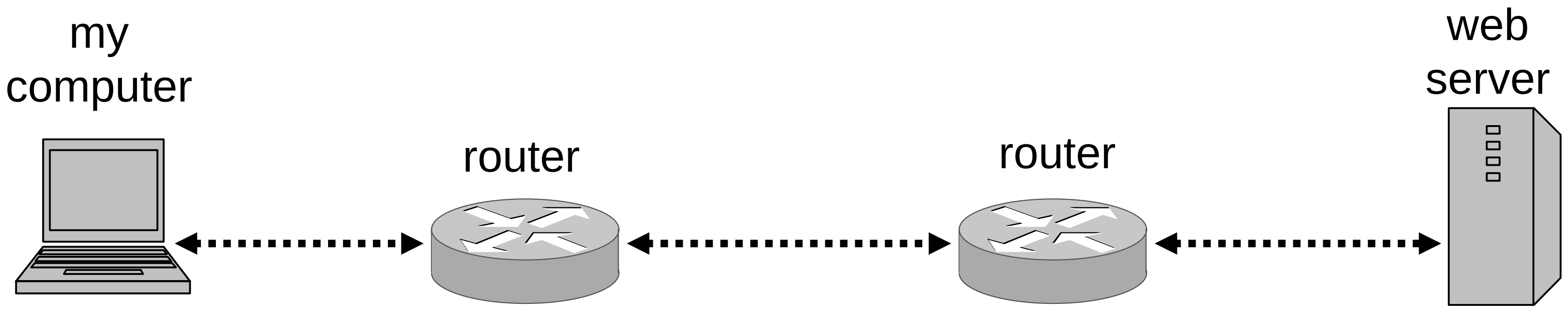
networks and their properties are different at each layer



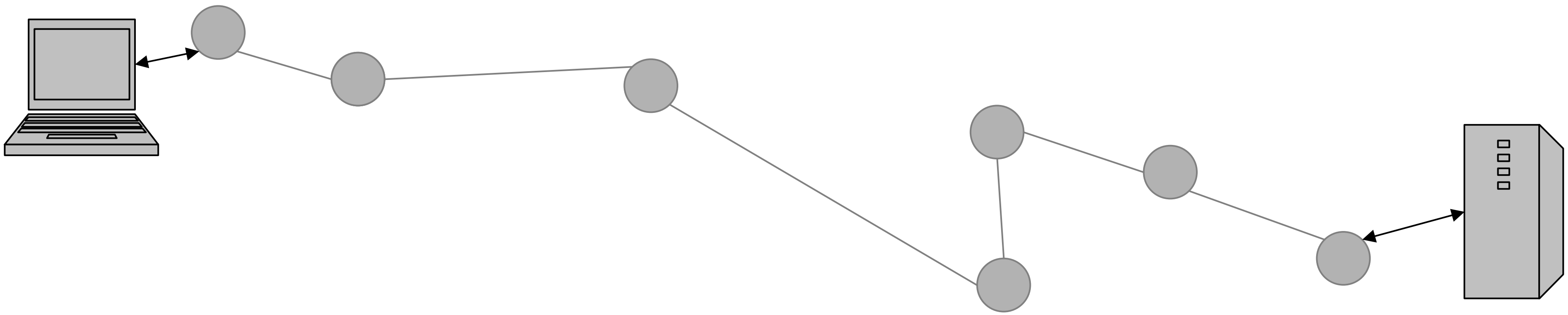
Internet Architecture: Dual Decomposition



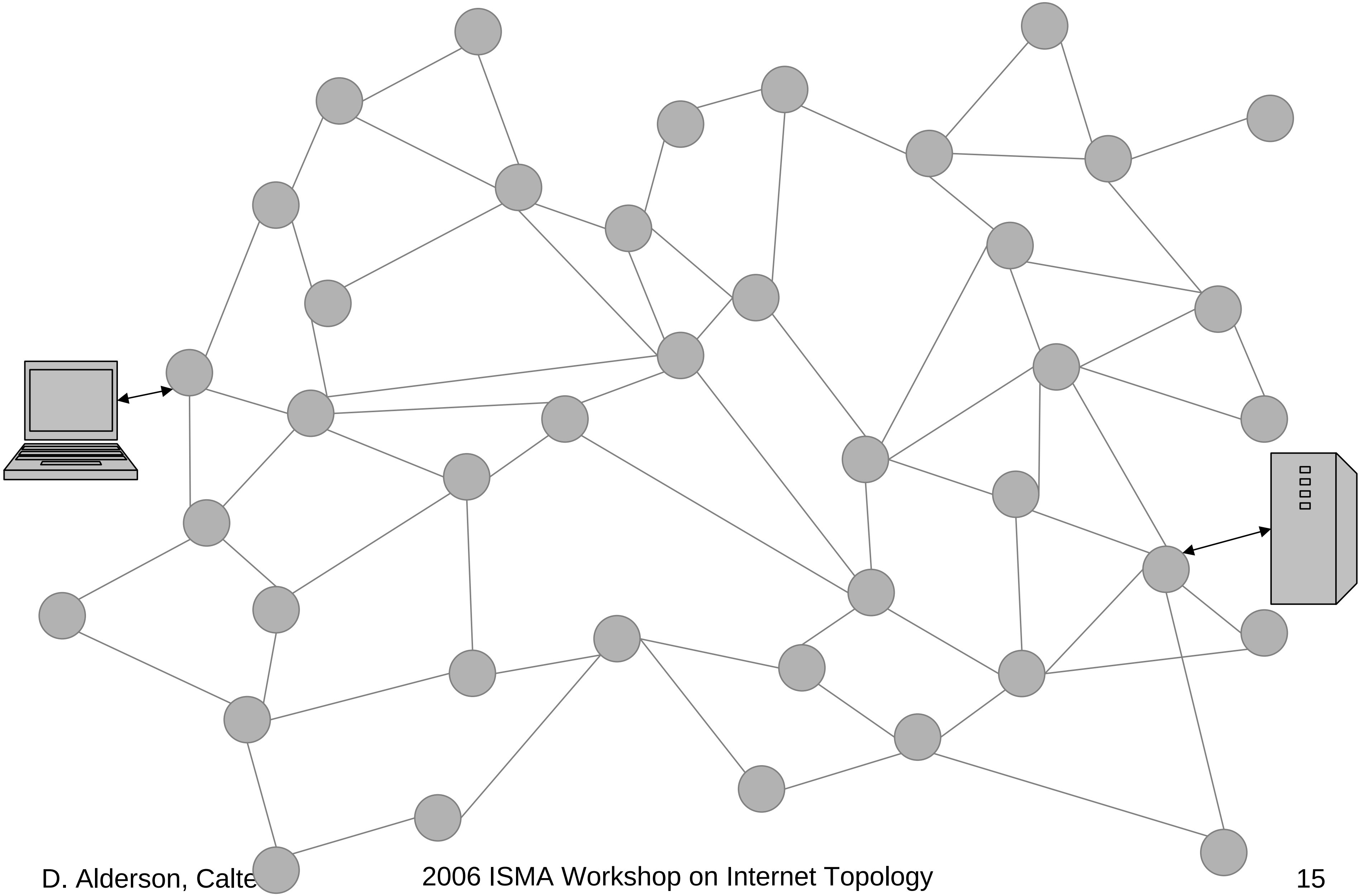
The Router-Level Internet



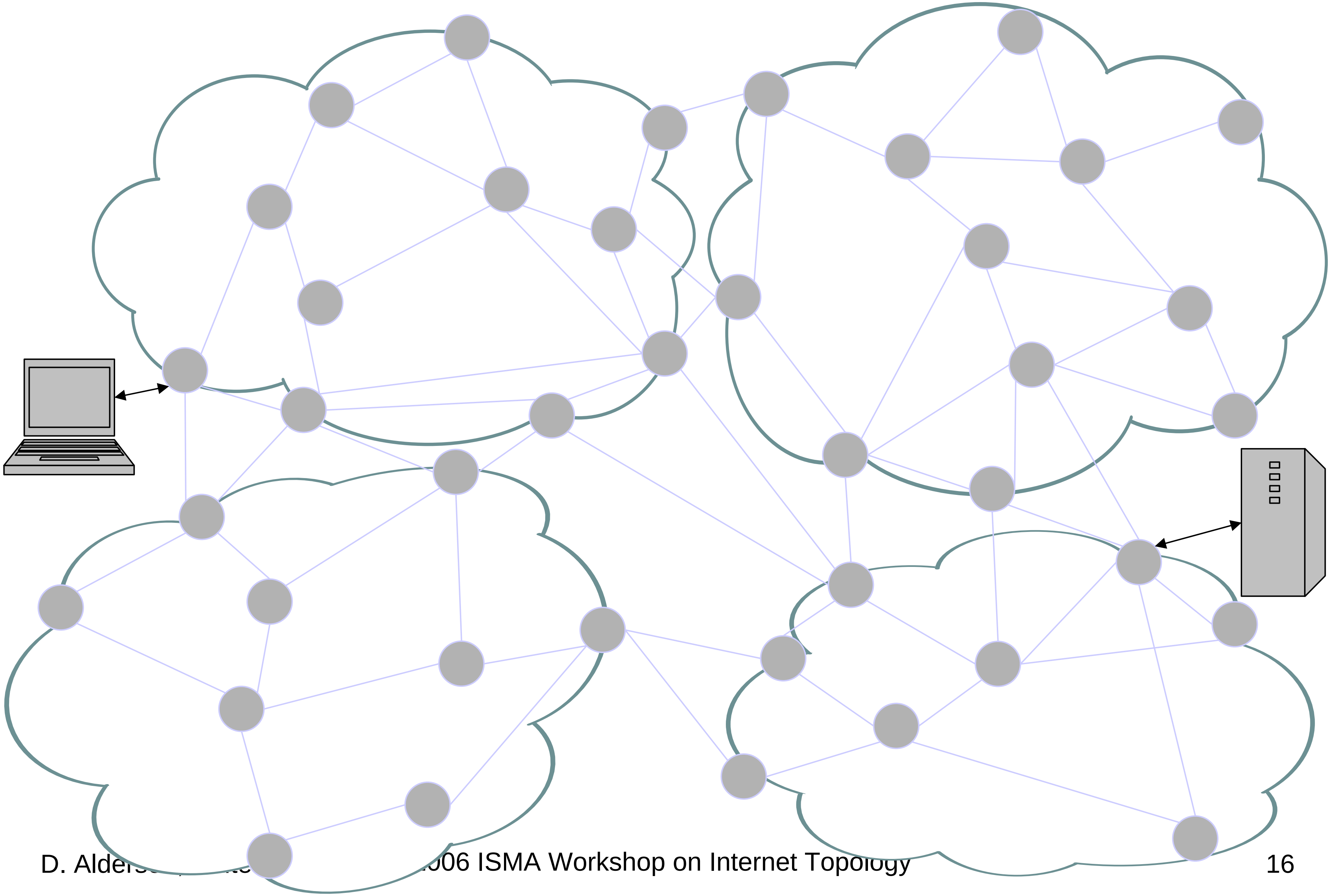
Bigger Picture: Internet Architecture



Bigger Picture: Internet Architecture

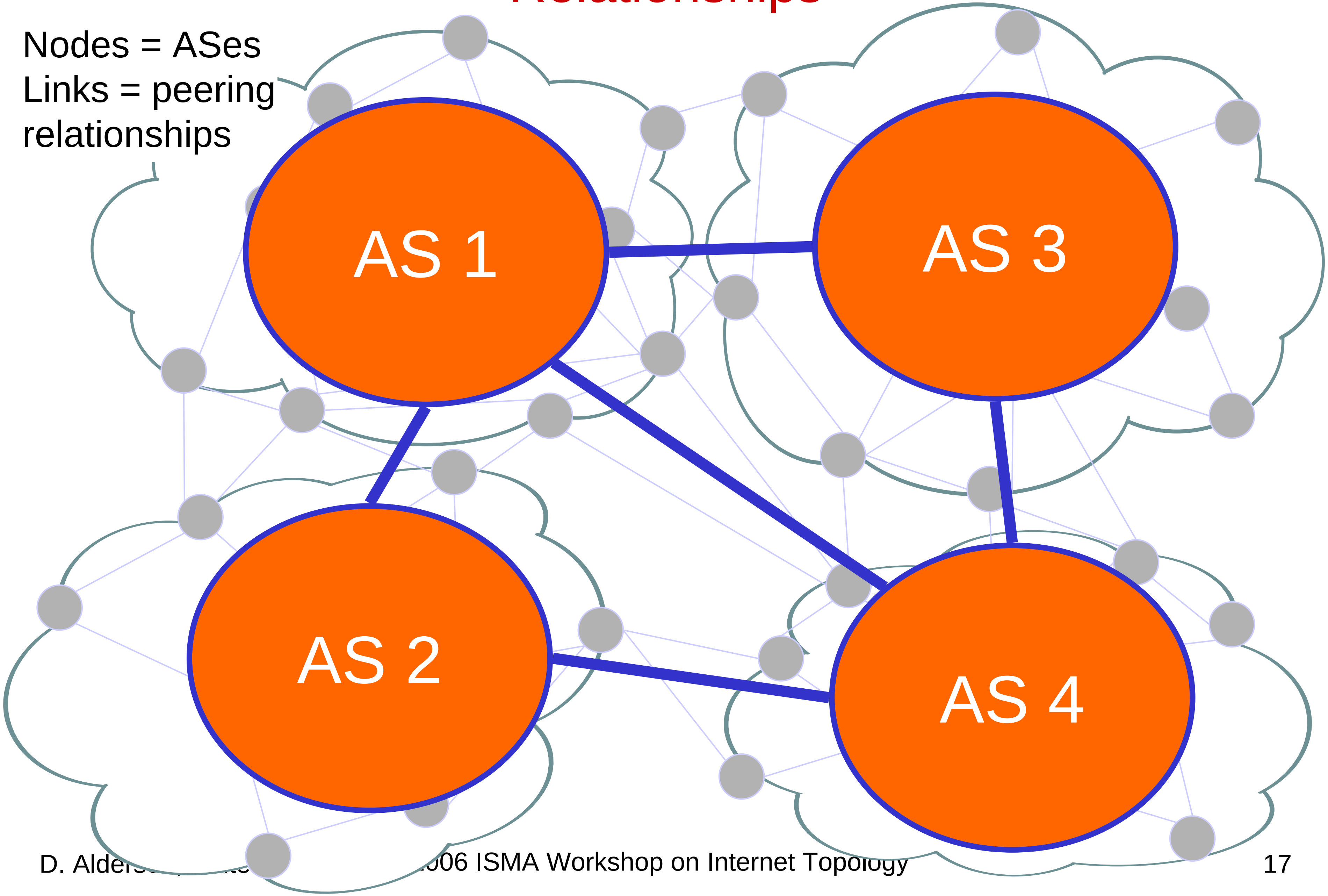


Bigger Picture: Internet Architecture



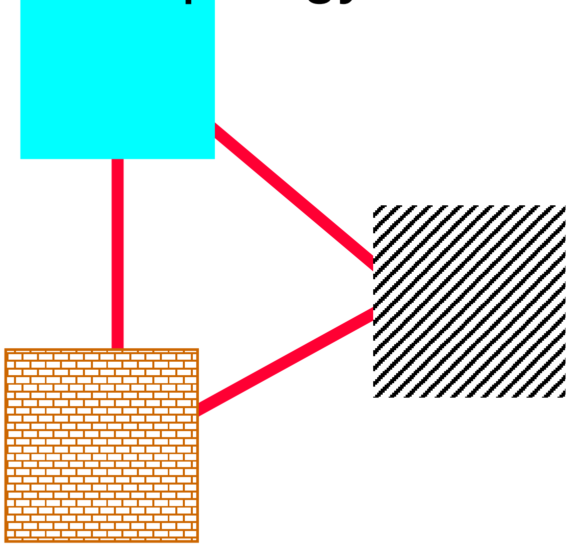
Autonomous System (AS) Graphs = Business Relationships

Nodes = ASes
Links = peering relationships

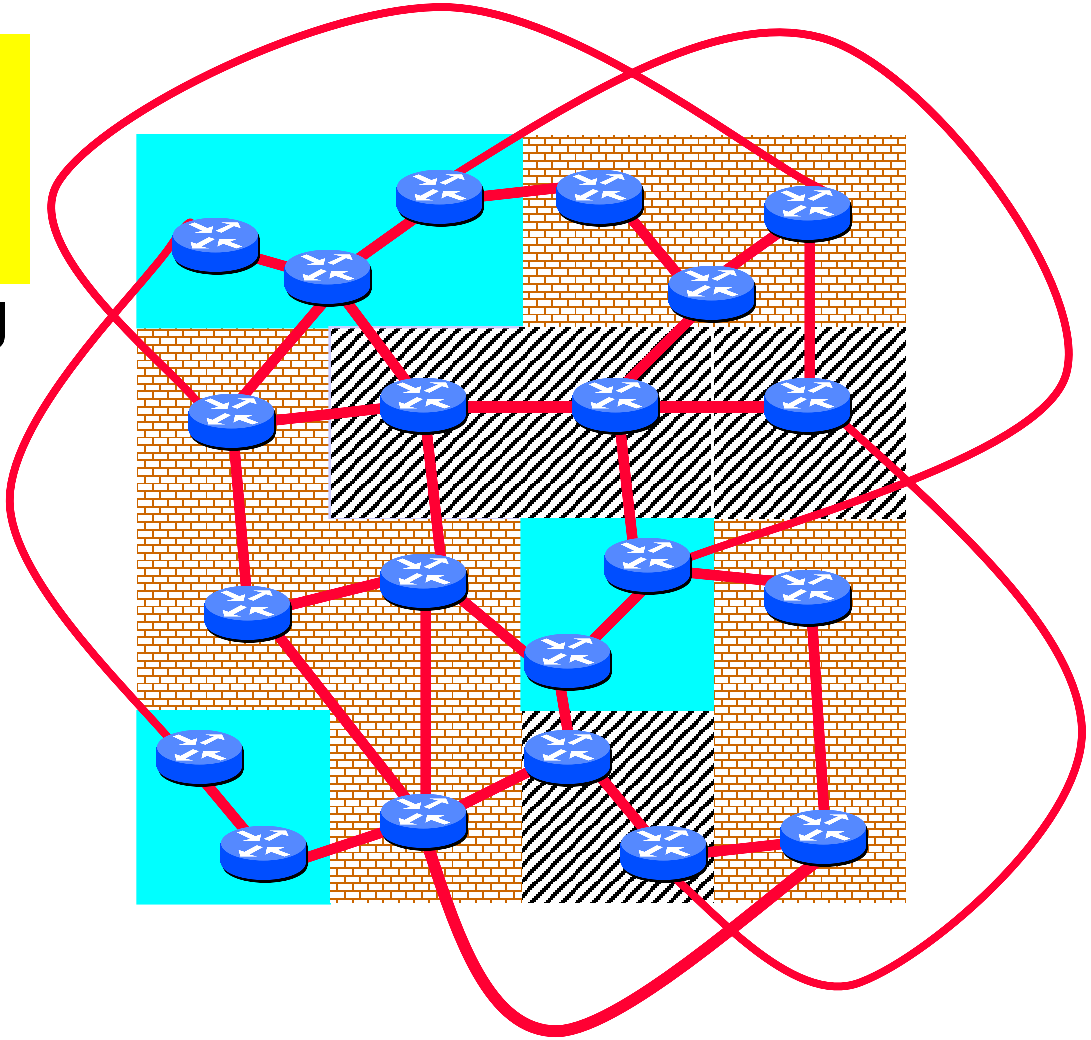


AS graphs obscure topology!

see talks by Hyunseok Chang and others on Thursday AM for more on AS topology modeling



The AS graph may look like this.



Reality may be closer to this…

Courtesy Tim Griffin

MESSAGE #1: specify WHICH aspect of Internet topology

- There is **no** "generic" Internet topology
- Router-level, IP-level, AS-level, application-level, ...
- Details of each make a big difference

PITFALL: Lack of specificity causes confusion

- Albert, Jeong, and Barabasi (2000) study robustness properties of the Internet by equating AS-level topology with router-level topology

  ⇒Knocking out nodes in the AS graph??

- Berger, Borgs, Chayes, and Saberi (2005) study the spread of viruses on the Internet by equating the Web graph with the router-level topology.

  ⇒Virus propagation on the Web graph??

Unfortunately, direct inspection of Internet topology is generally NOT possible

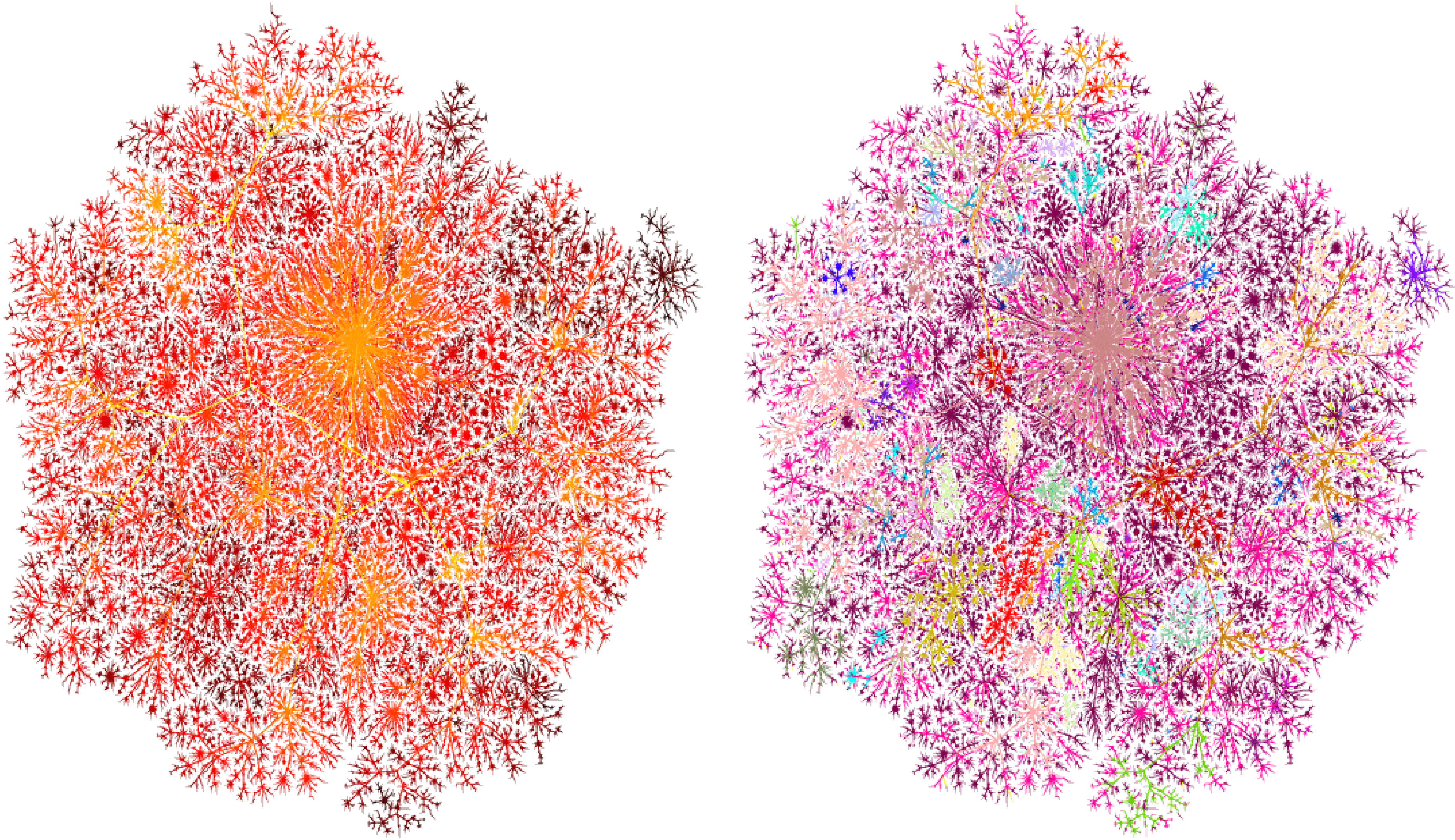
- Economic incentive for ISPs to obscure network structure
- Recent trend
  - Empirical measurement studies
  - Generative models
- Obstacles
  - Mismatch between what we want to measure and can measure
  - Imperfect measurements
  - What macro/microscopic statistics characterize a topology?
  - How to determine what matters?

Remainder of talk: focus on router-level topology

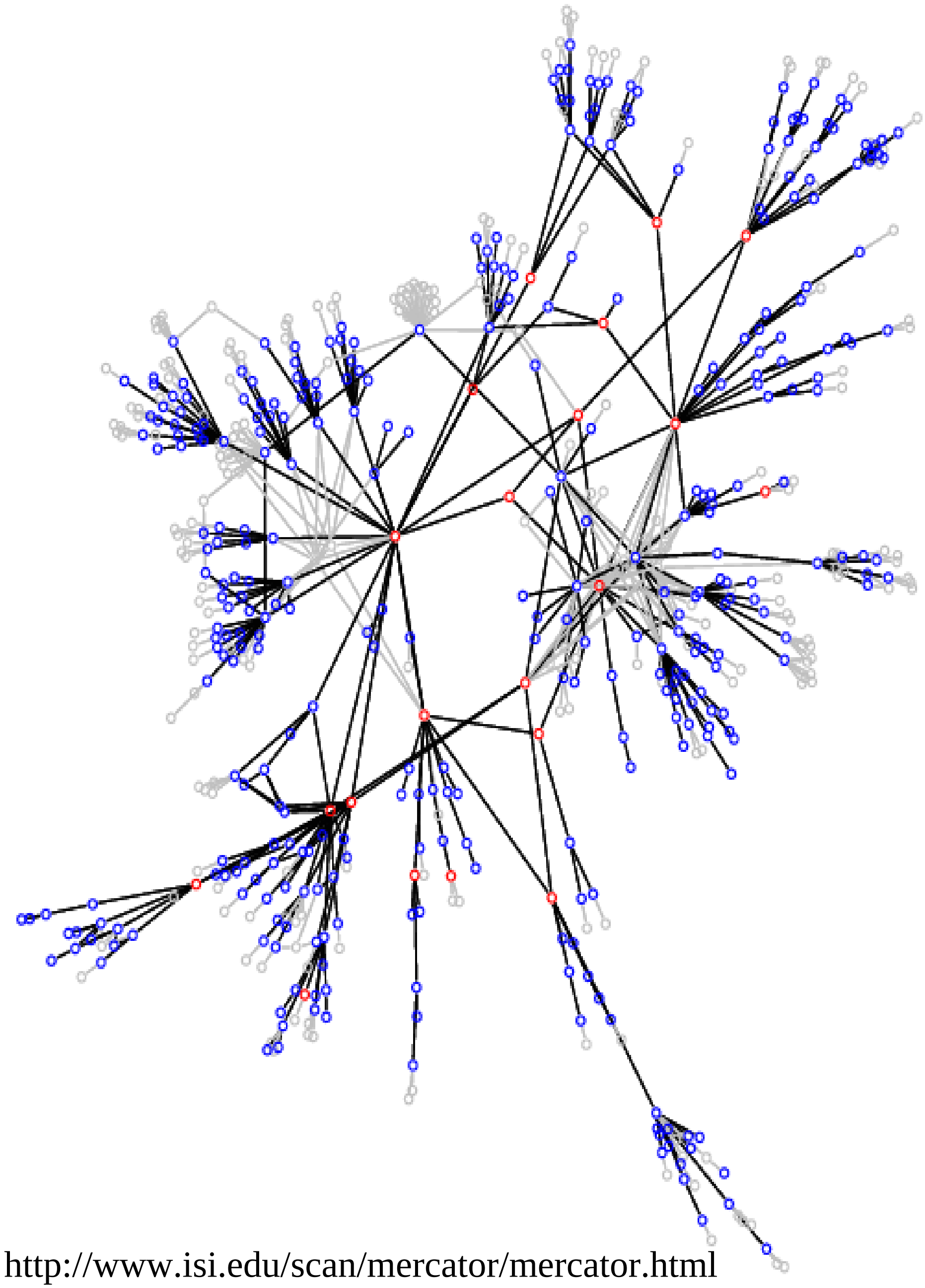
considerable progress in measuring router-level topology…

- `traceroute` tool
  - Discovers compliant (i.e., IP) routers along path between selected network host computers

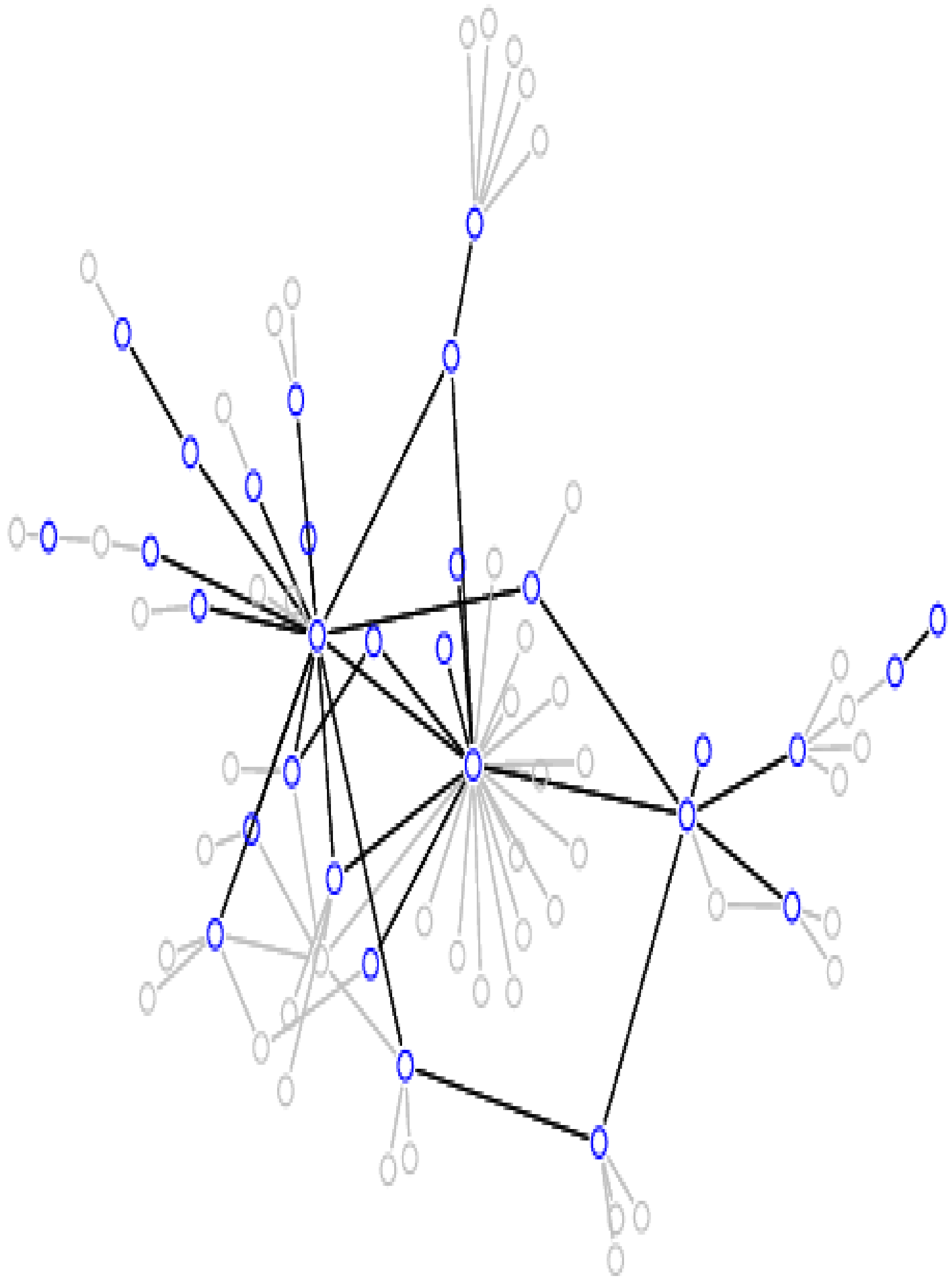
- Large-scale traceroute experiments
  - Pansiot and Grad (router-level map from around 1995)
  - Cheswick and Burch (mapping project 1997–)
  - Mercator (router-level maps from around 1999 by R. Govindan and H. Tangmunarunkit)
  - Skitter (ongoing mapping project by CAIDA folks)
  - Rocketfuel (state-of-the-art router-level maps of individual ISPs by UW folks)

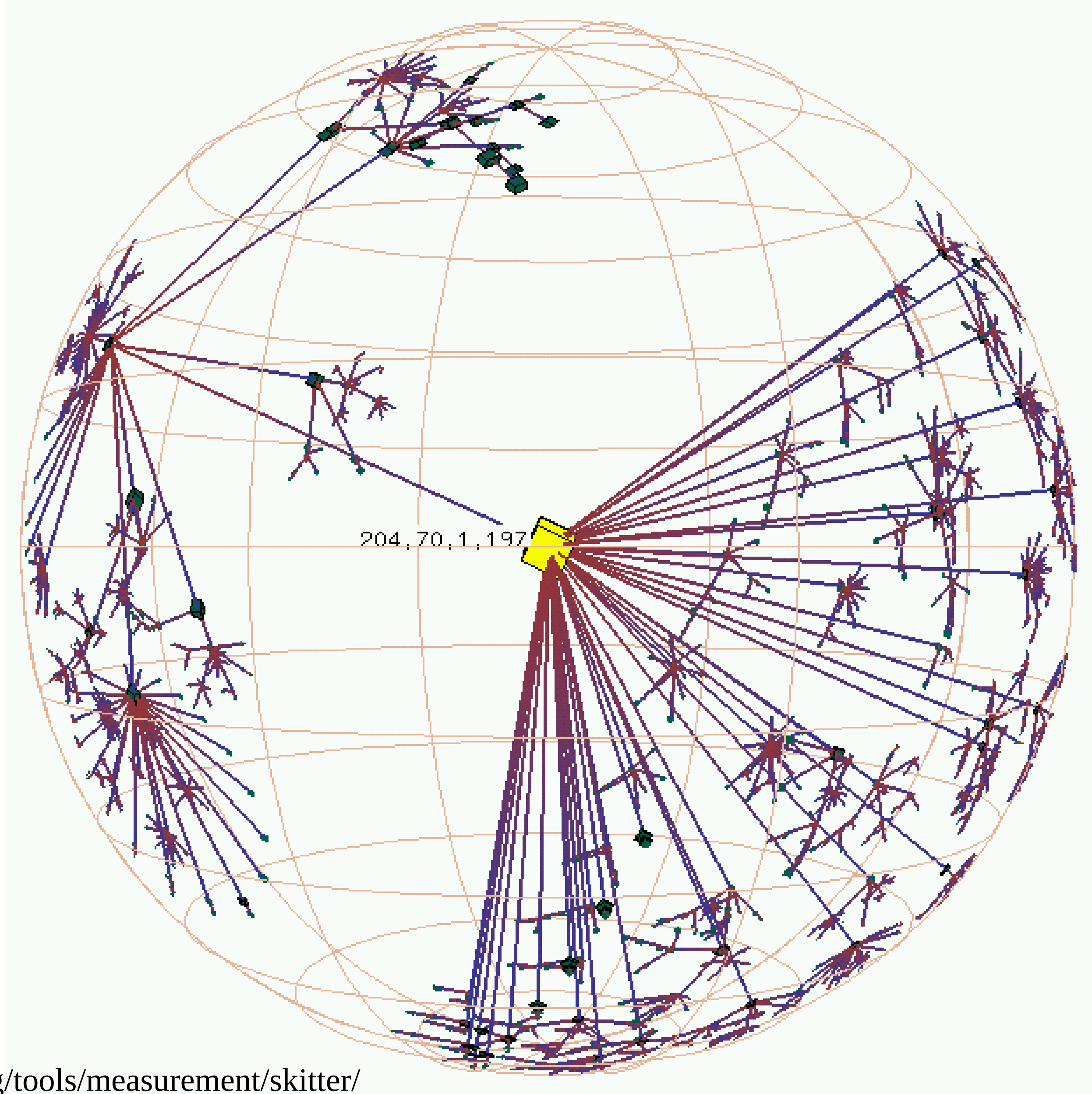


http://research.lumeta.com/ches/map/

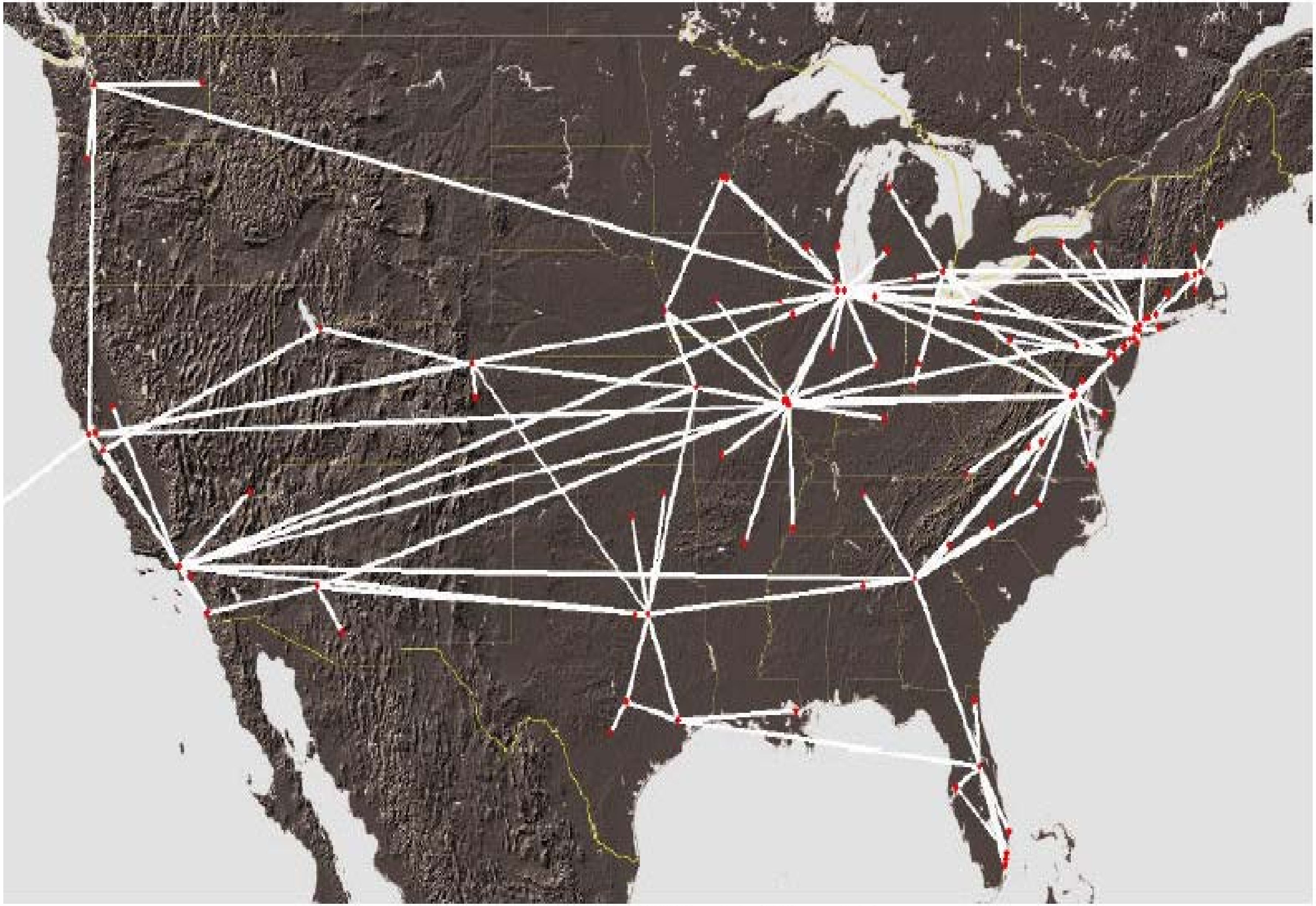


http://www.isi.edu/scan/mercator/mercator.html





http://www.caida.org/tools/measurement/skitter/

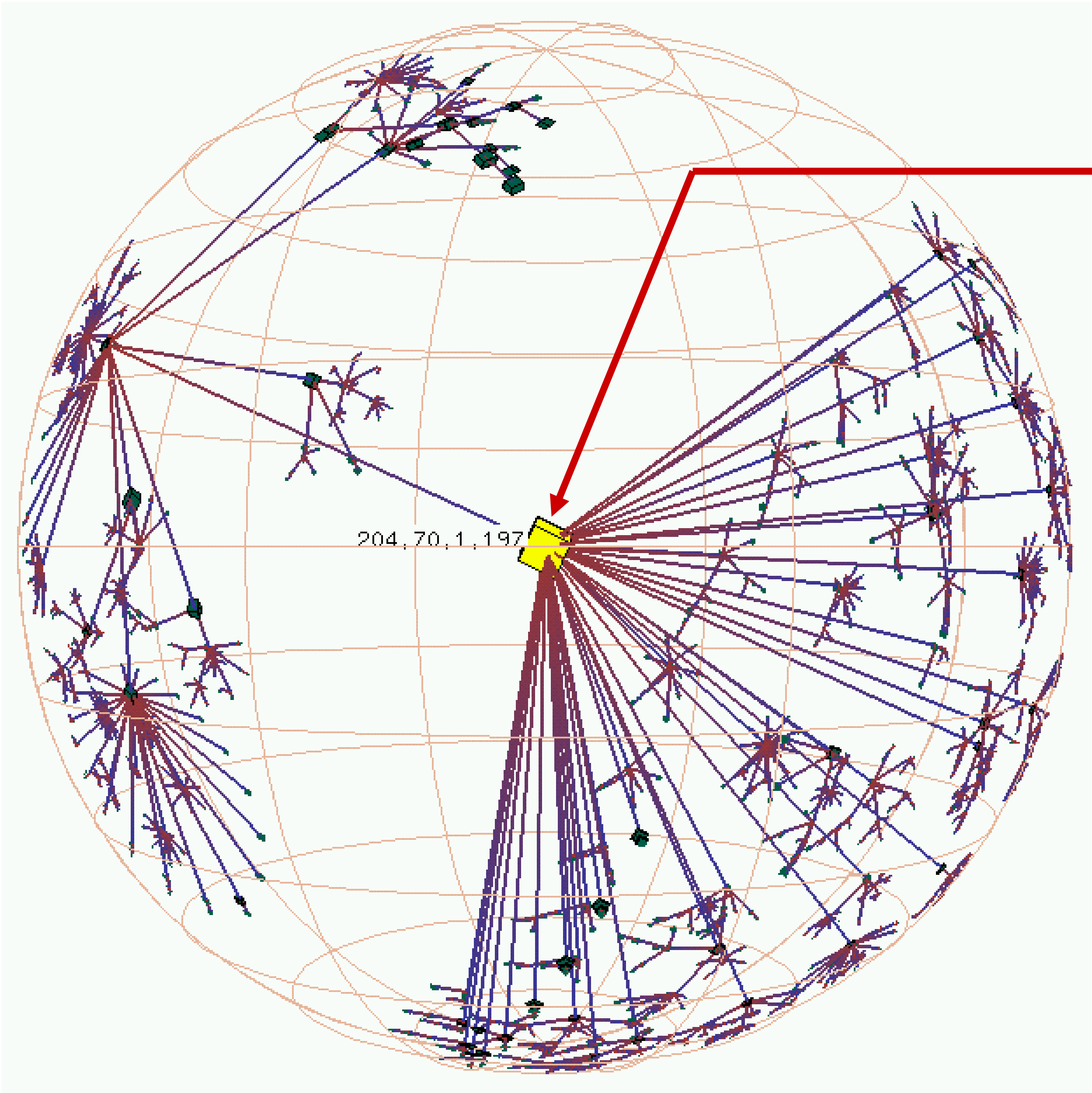


Background image courtesy JHU, applied physics labs

http://www.cs.washington.edu/research/networking/rocketfuel/bb

…but considerable drawbacks to existing approaches

- **traceroute-based measurements are ambiguous**
  - **traceroute is strictly about IP-level connectivity**
  - **traceroute cannot distinguish between high connectivity nodes that are for real and that are fake and due to underlying Layer 2 (e.g., Ethernet, ATM) or Layer 2.5 technologies (e.g., MPLS)**



- www.savvis.net
- managed IP and hosting company
- founded 1995
- offering “private IP with ATM at core”

Possible that this “node” is an entire network! (not just a router)

http://www.caida.org/tools/measurement/skitter/

…but considerable drawbacks to existing approaches

- traceroute-based measurements are ambiguous
  - traceroute is strictly about IP-level connectivity
  - traceroute cannot distinguish between high connectivity nodes that are for real and that are fake and due to underlying Layer 2 (e.g., Ethernet, ATM) or Layer 2.5 technologies (e.g., MPLS)
- traceroute-based measurements are inaccurate
  - Requires some guesswork in deciding which IP addresses/interface cards refer to the same router ("alias resolution" problem)
- traceroute-based measurements are incomplete/biased
  - IP-level connectivity is more easily/accurately inferred the closer the routers are to the traceroute source(s)
  - Node degree distribution is inferred to be of the power-law type even when the actual distribution is not

see talk by Aaron Clauset et al. on Thu AM for more on this…

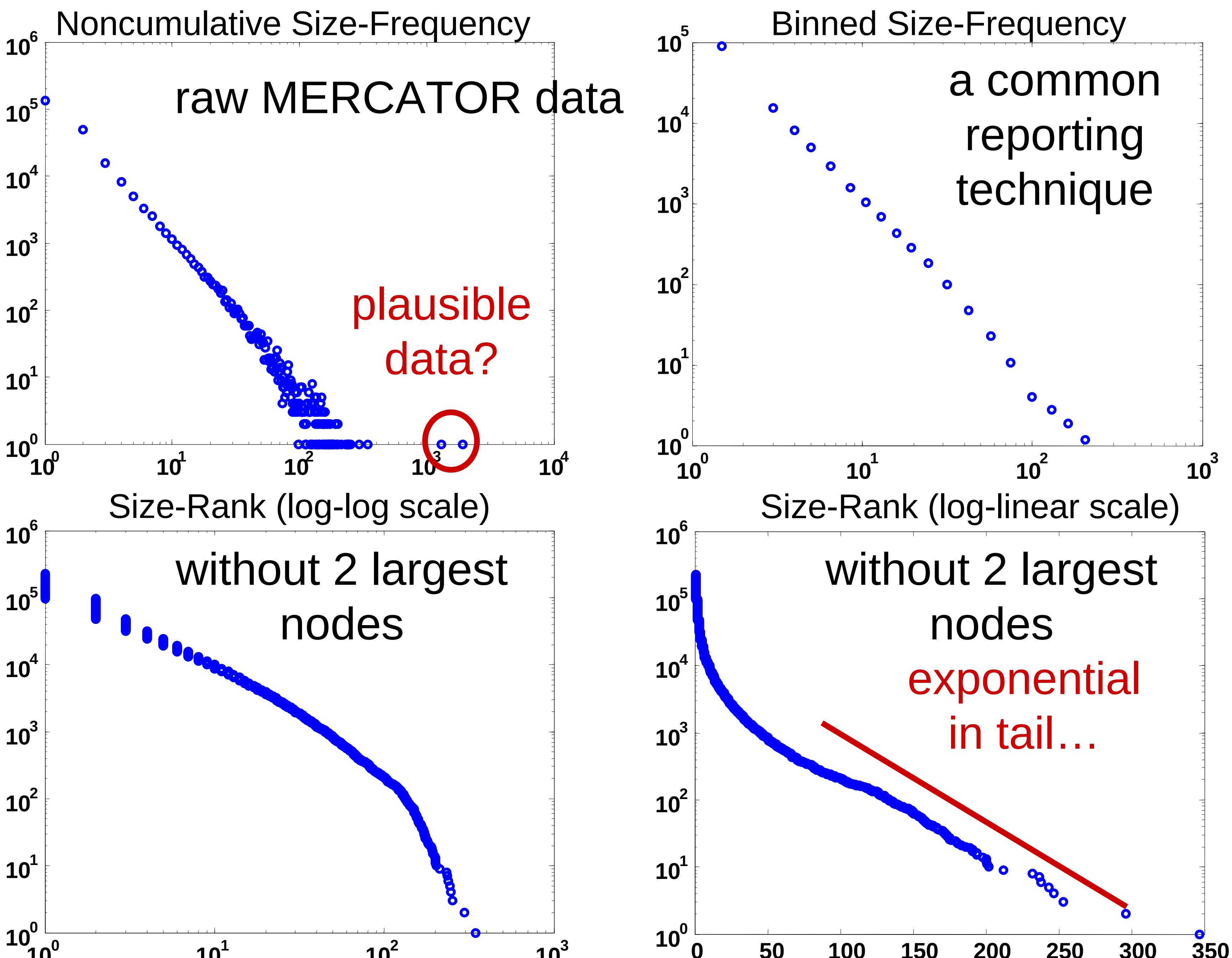
MESSAGE #2: **Idiosyncracies of network measurements require careful interpretation**

- **Each technique is typically specific to network of interest (e.g., traceroute for IP-level, BGP tables for AS-level)**
- **Even best-of-breed measurement data is ambiguous, inaccurate, and incomplete**

PITFALL: Taking (someone else's) data at face value may provide a false basis for results

- **example: use of MERCATOR data to support claims of power-law degree distribution for router-level Internet**

  ⇒Are routers with >1000 connections plausible??


Noncumulative Size-Frequency
raw MERCATOR data
plausible data?
Binned Size-Frequency
a common reporting technique
Size-Rank (log-log scale)
without 2 largest nodes
Size-Rank (log-linear scale)
without 2 largest nodes
exponential in tail…
10^0 10^1 10^2 10^3 10^4 10^5 10^6
0 50 100 150 200 250 300 350

It is difficult to know what "matters" when it comes to representing router-level topology

Observation

- Long-range links are expensive
- Real networks are not random, but have obvious hierarchy.
- Internet topologies exhibit power law degree distributions (Faloutsos et al., 1999)

Modeling Approach

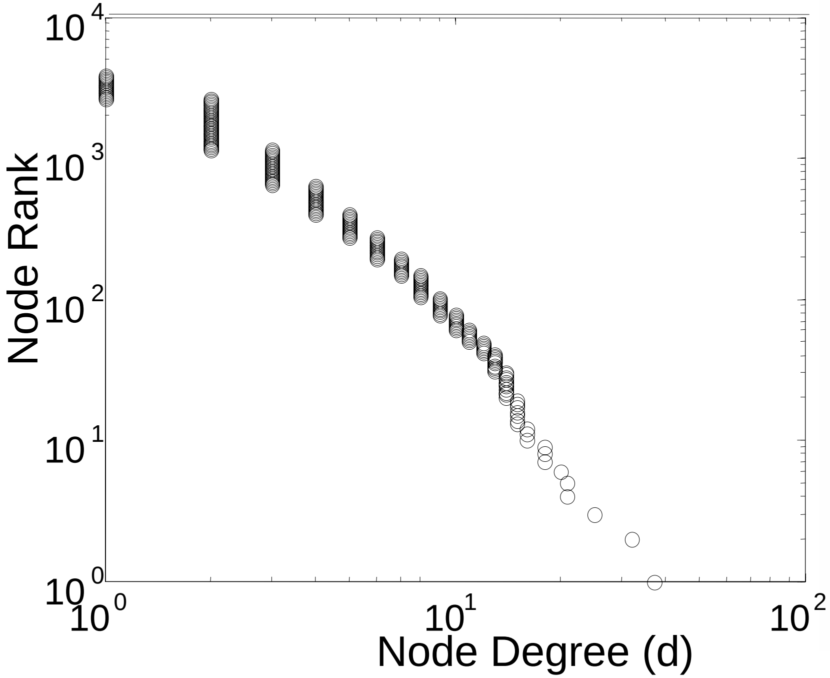
- Random graph models (Waxman, 1988)
- Structural models (GT-ITM Calvert/Zegura, 1996)
- Degree-based models replicate power-law degree sequences (e.g. scale-free networks, 1999-2004)

Power Laws and Internet Topology

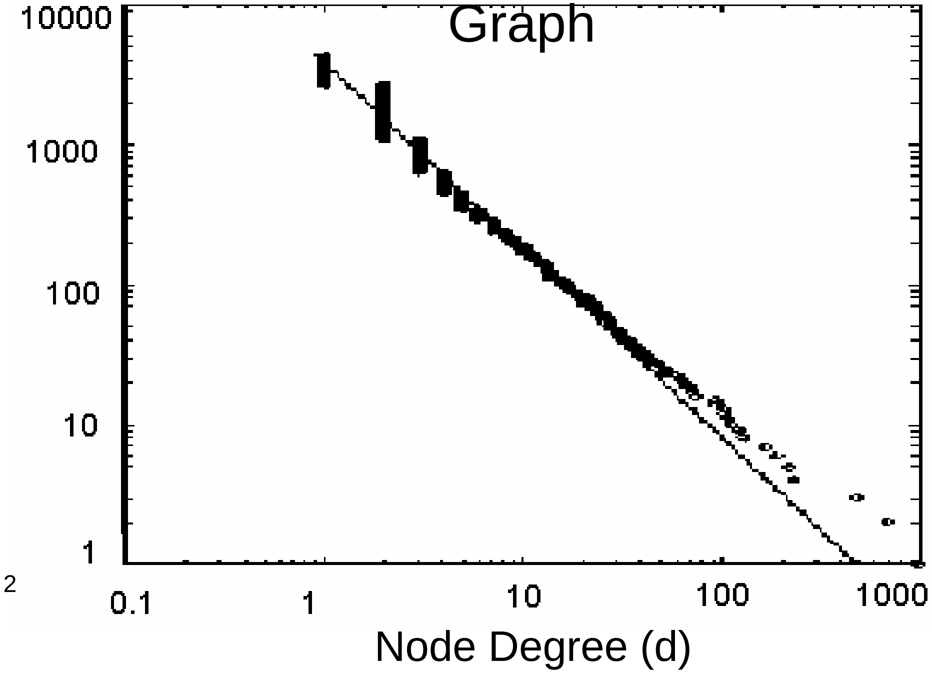
<u>Source</u>: Faloutsos et al. (1999)

Rank: R(d) = P (D>d) x #nodes

Router-Level Graph



Autonomous System (AS) Graph



Node Degree: d = # connections

- A random variable X is said to follow a *power law* with *index* $\alpha >$ 0 if $$P[X > x] \approx cx^{-\alpha}, \text{ as } x \to \infty$$
- Led to active research in *degree-based* network models

Degree-Based Network Models

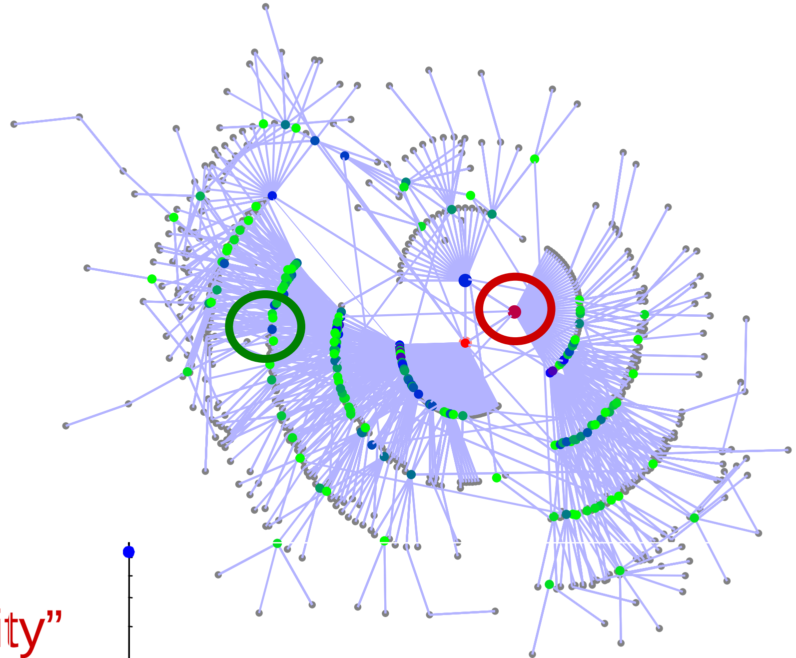
- Basic Idea: traditional random graphs [Erdös & Renyí, 59] do not produce power laws, so develop new models that explicitly attempt to match the observed (power law) distribution in node degree

- **Preferential Attachment**
  - Incremental growth + new nodes attach to high-degree nodes
  - "Rich get richer"—power laws in asymptotic limit
  - Scale-free networks [Barabási & Albert, 99]
  - Generators: Inet, GPL, AB, BA, BRITE, CMU power-law generator

- **Expected Degree Sequence**
  - Based on random graph models that skew probability distribution to produce power laws in expectation
  - Power law random graph (PLRG) [Aiello et al., 00]
  - Generalized random graph (GRG) [Chung & Lu, 03]

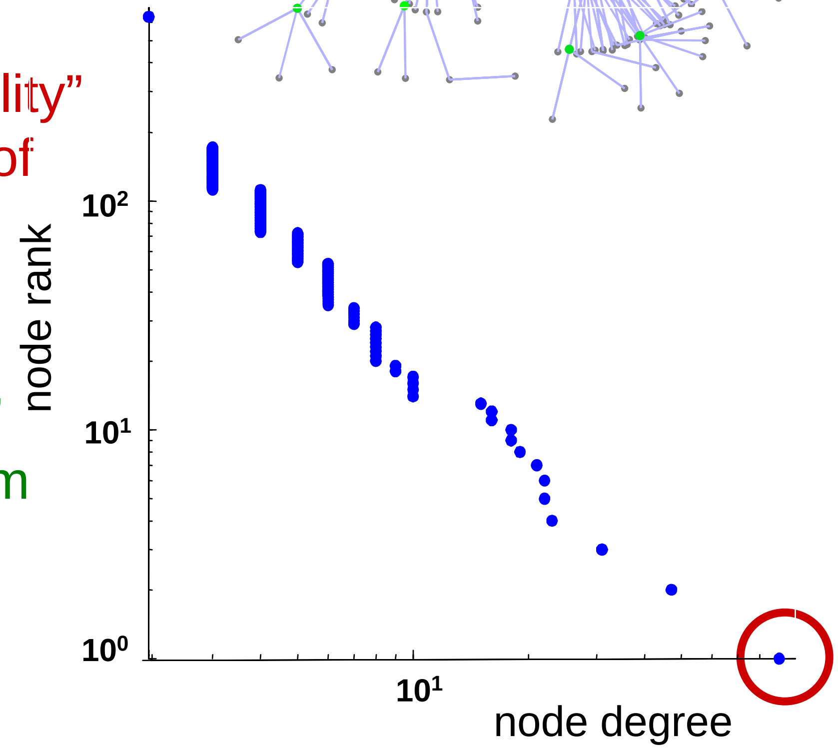
“Scale-free” networks and the “Achilles’ heel” of the Internet

Reference: R. Albert, H. Jeong, and A.-L. Barabási. Attack and error tolerance of complex networks. *Nature* 406, 378-382, 2000.



“Attack vulnerability” = Targeted loss of hub fragments network

“Error tolerance” = Loss of random node has little effect



The literature on Scale-Free Networks claims broad implications for the Internet and other networks

Power laws in network connectivity…

⇔ Are necessary and sufficient for "scale-free structure"

⇔ Imply critically connected "hubs"

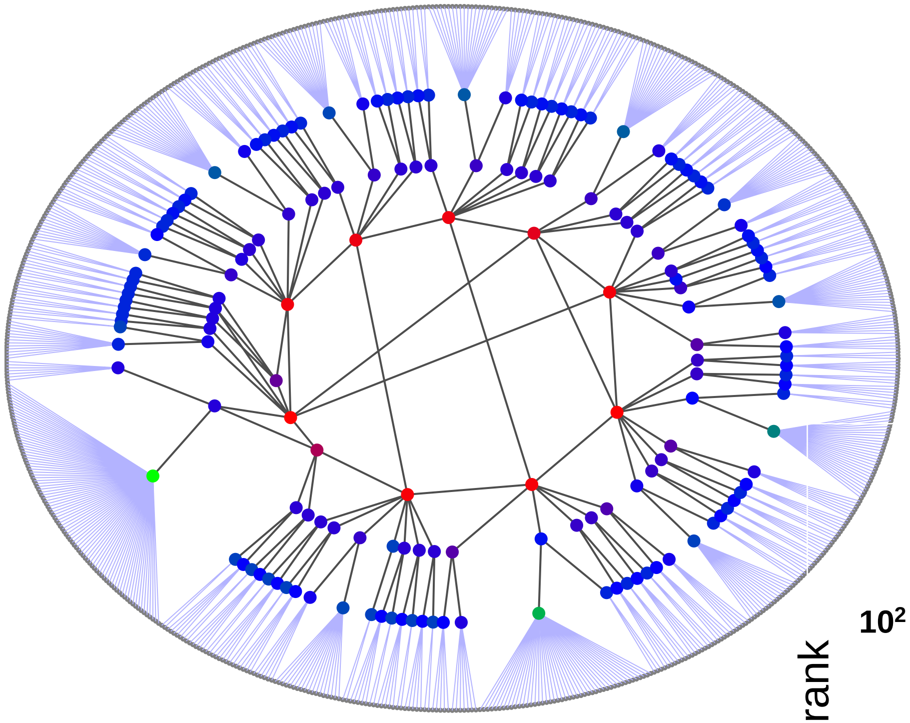
⇒ Create an Achilles' heel vulnerability

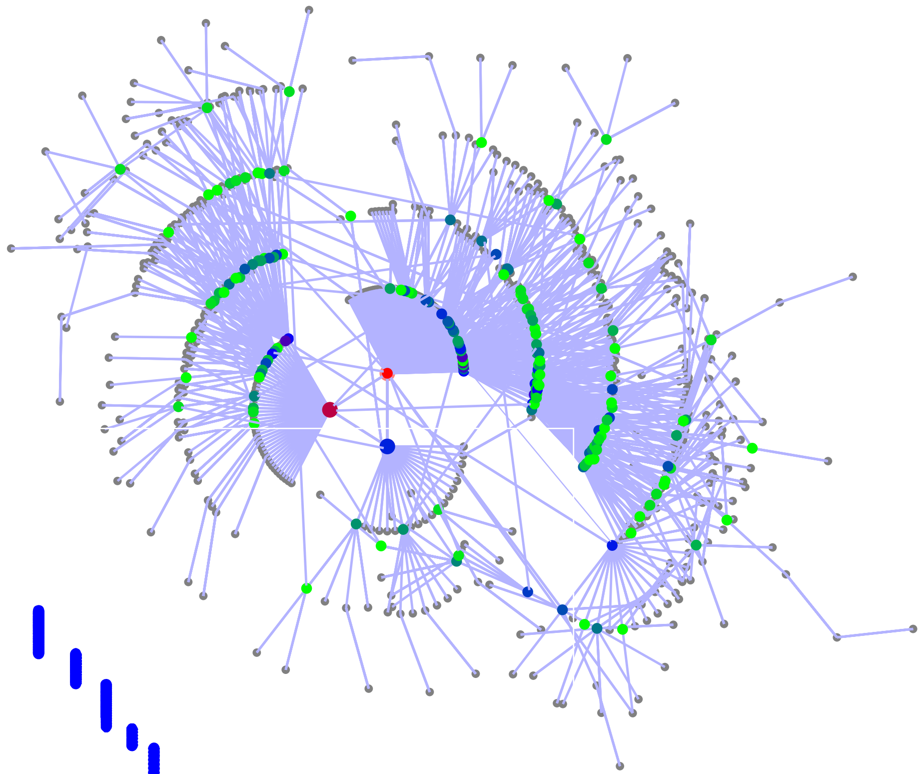
⇒ Yield a zero epidemic threshold for contagion

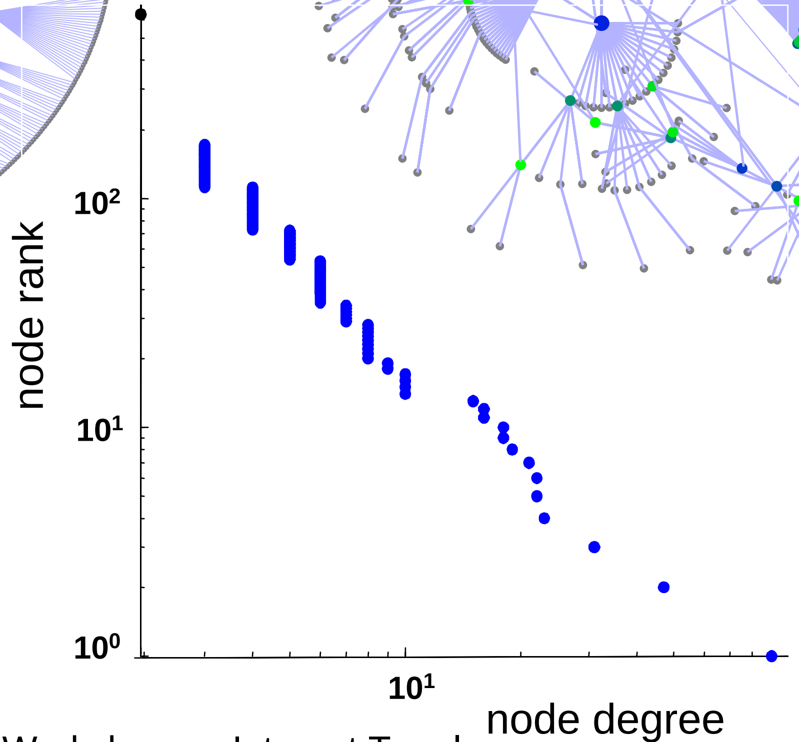
⇒ Are evidence of fundamental self-organization in networks

⇒ This self-organization is believed by some to be a *universal feature* of technological, biological, social and business networks

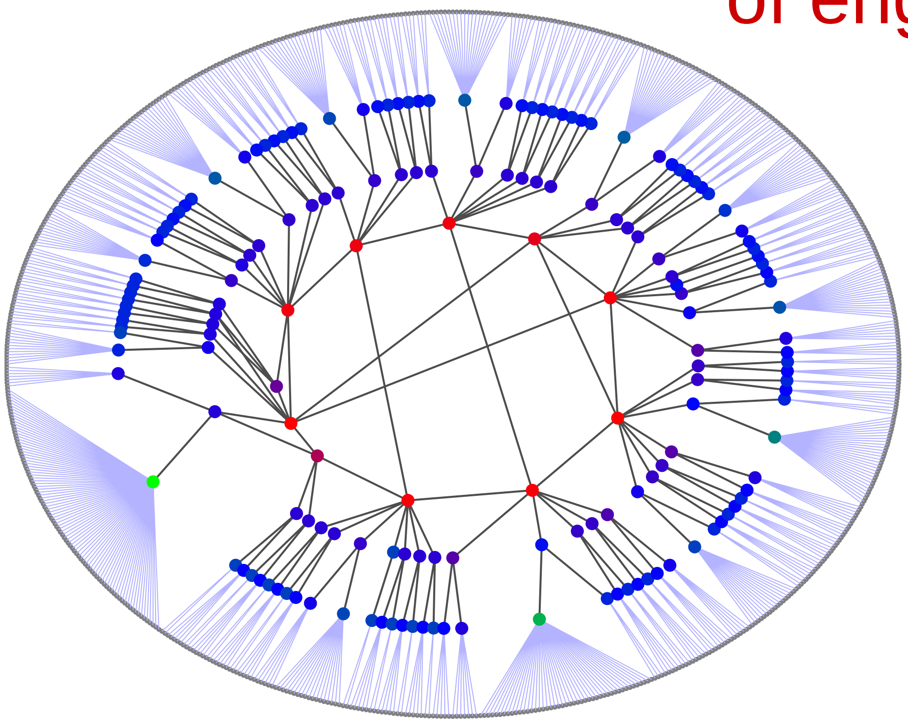
⇒ Efforts to protect complex networks should focus on the most highly-connected components

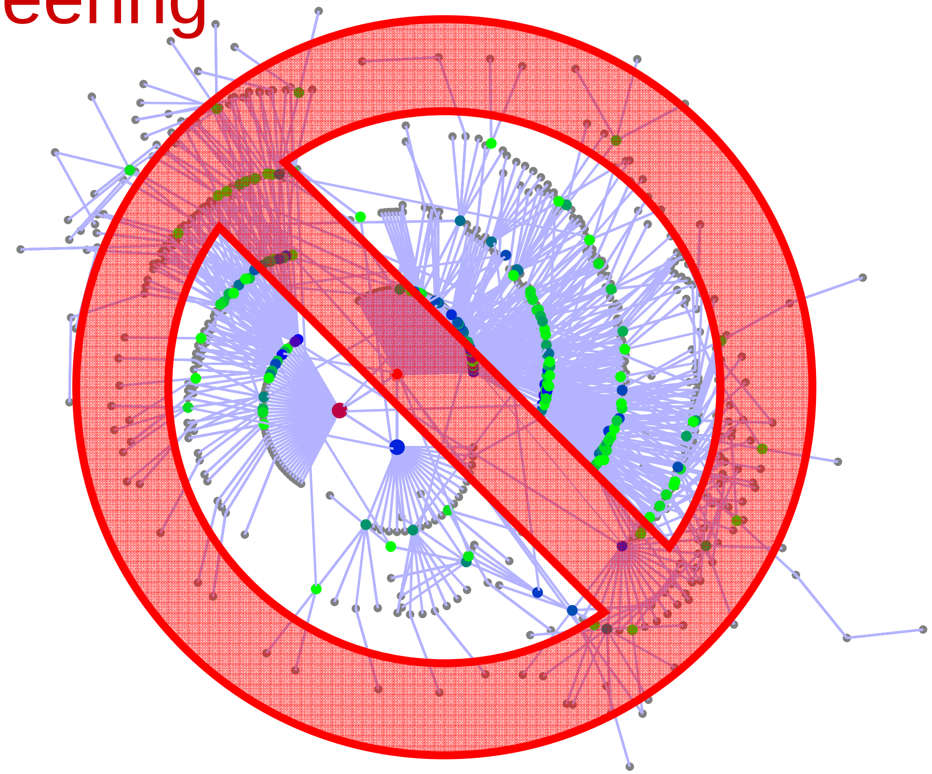





node rank
10^2
10^1
10^0
10^1
node degree

MESSAGE #3: networks with the same statistical features can be OPPOSITES in terms of engineering





- Low degree core
- Result of design
- High performance and robustness

- High degree central "hubs"
- From random construction
- Poor performance and robustness

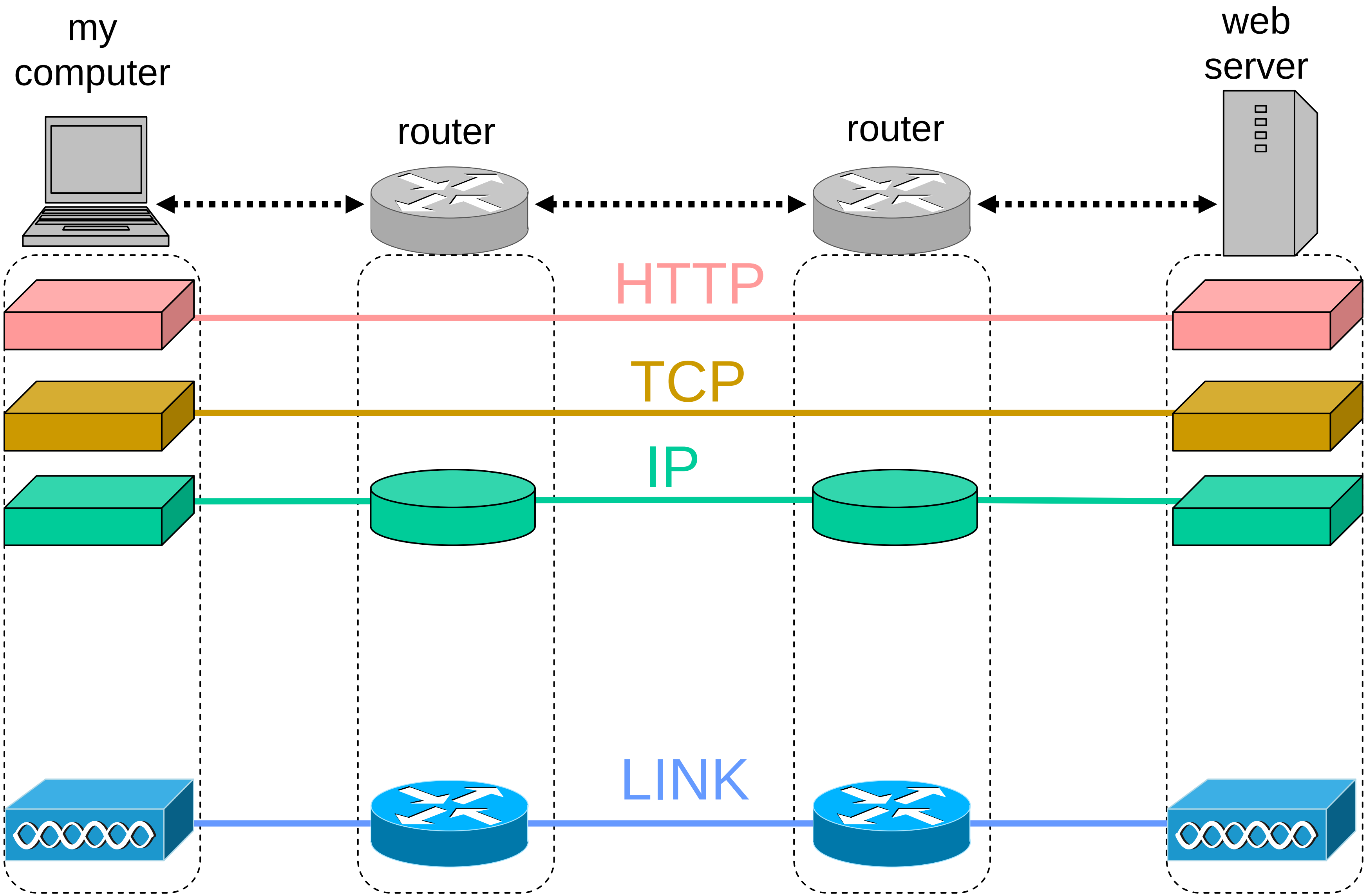
Trends in Topology Modeling

Observation

- Long-range links are expensive
- Real networks are not random, but have obvious hierarchy.
- Internet topologies exhibit power law degree distributions (Faloutsos et al., 1999)
- Degree-based models are fundamentally inconsistent with engineering reality

Modeling Approach

- Random graph models (Waxman, 1988)
- Structural models (GT-ITM Calvert/Zegura, 1996)
- Degree-based models replicate power-law degree sequences (scale-free networks, 1999-2004)
- Optimization-driven models yield topologies consistent with design tradeoffs of network engineers (SIGCOMM'04)


my
computer
router
router
web
server
HTTP
TCP
IP
LINK

Our Perspective

- Who builds real router-level topologies?

  the "decision makers" are individual ISPs

- How do technology and cost influence deployment?

  they provide CONSTRAINTS on what the ISP can do

- How does one evaluate a "good" design?

  network PERFORMANCE can be measured in terms of tra

- What drives their structure?

  some form of an (implicit) OPTIMIZATION problem, although actual "design" may be decentralized and heuristic

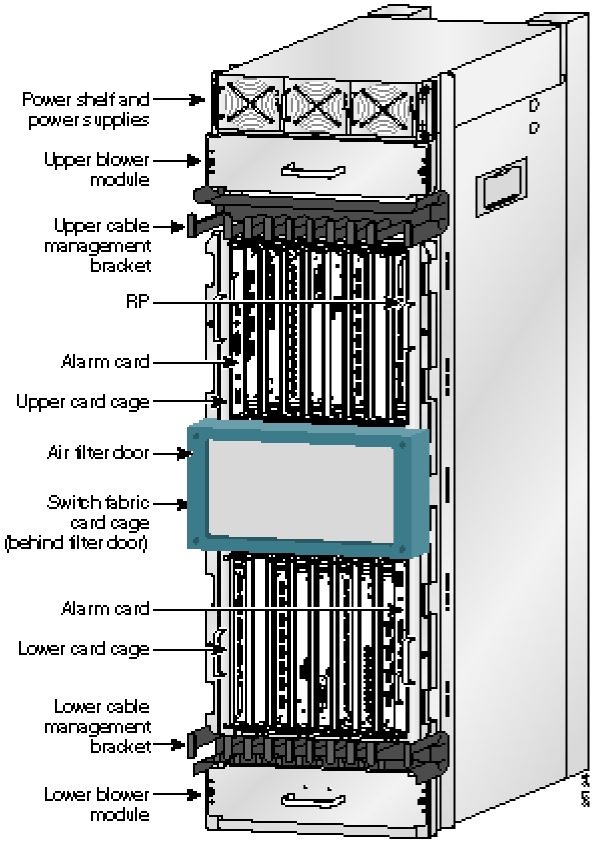
- What about power laws?

  a mere consequence of the inputs to the optimization prob

Cisco 12000 Series Routers

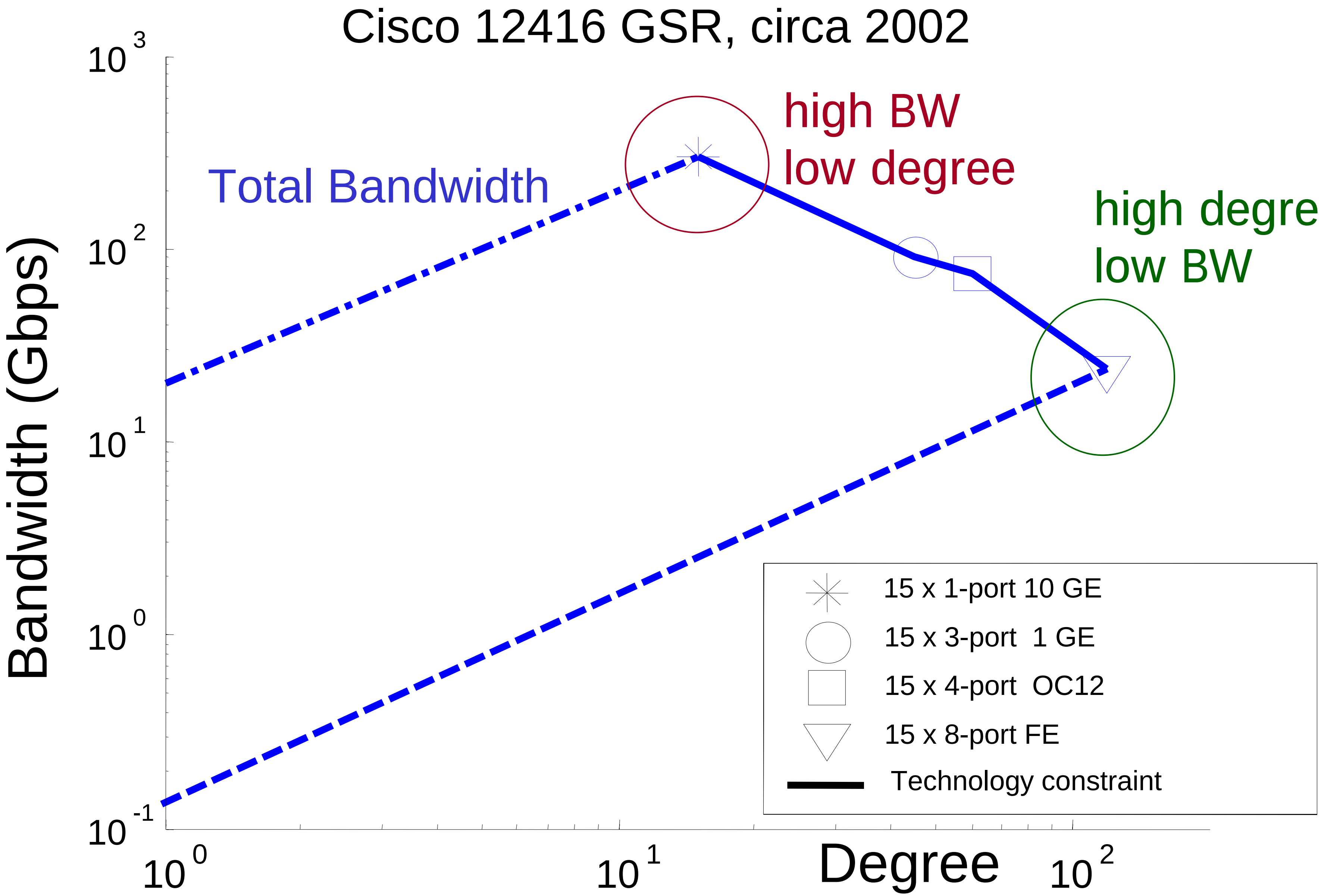
- Modular in design, creating flexibility in configuration.
- Router capacity is constrained by the number and speed of line cards inserted in each slot.

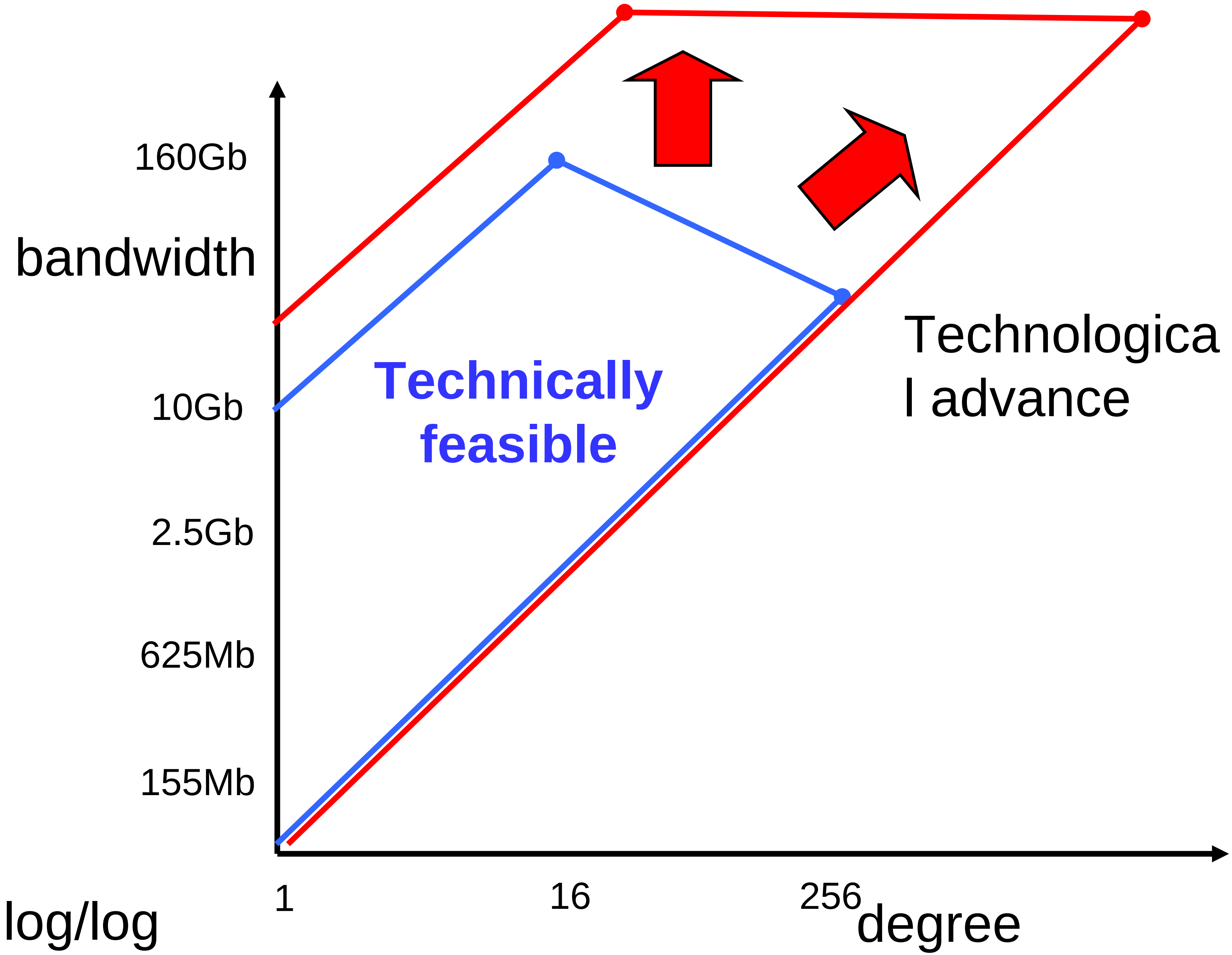
Chassis	Rack size	Slots	Switching Capacity
12416	Full	16	320 Gbps
12410	1/2	10	200 Gbps
12406	1/4	6	120 Gbps
12404	1/8	4	80 Gbps



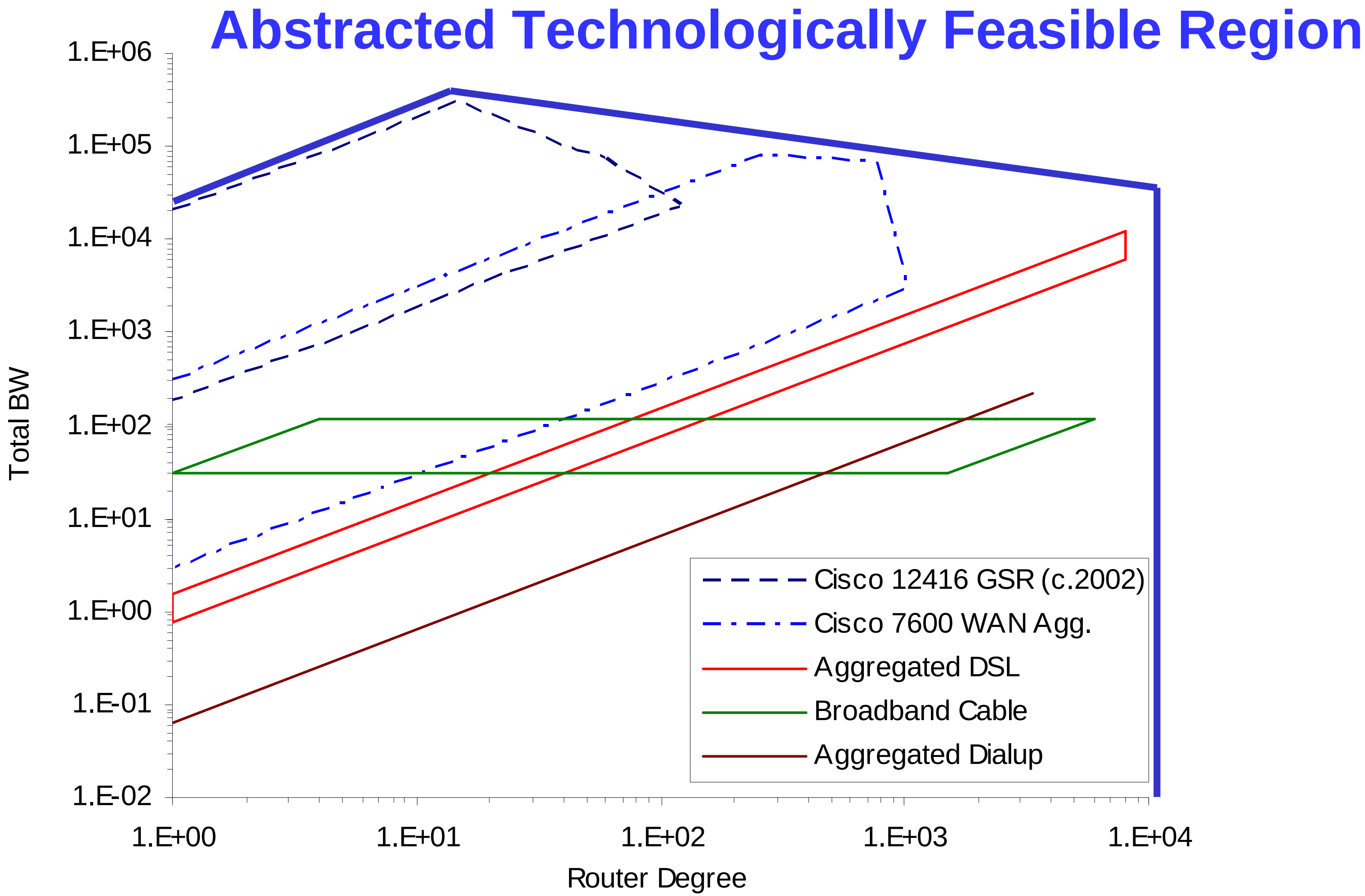
Source: www.cisco.com

Router Technology Constraint



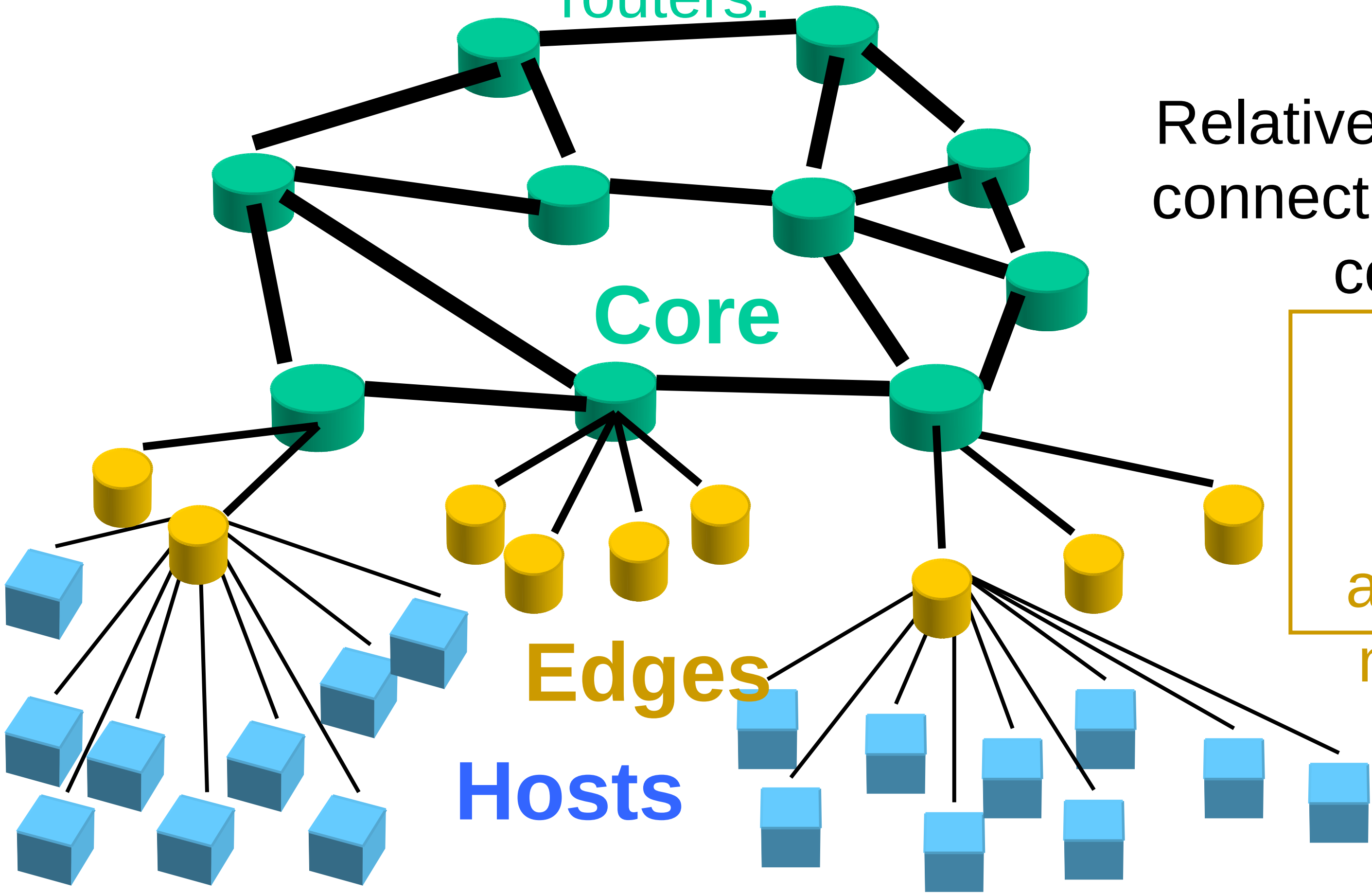

160Gb
bandwidth
10Gb
2.5Gb
625Mb
155Mb
Technically feasible
Technological advance
1
16
256
degree
log/log

router models specialize by “role”



Heuristically Optimal Topology

Sparse, mesh-like core of fast, low-degree routers.

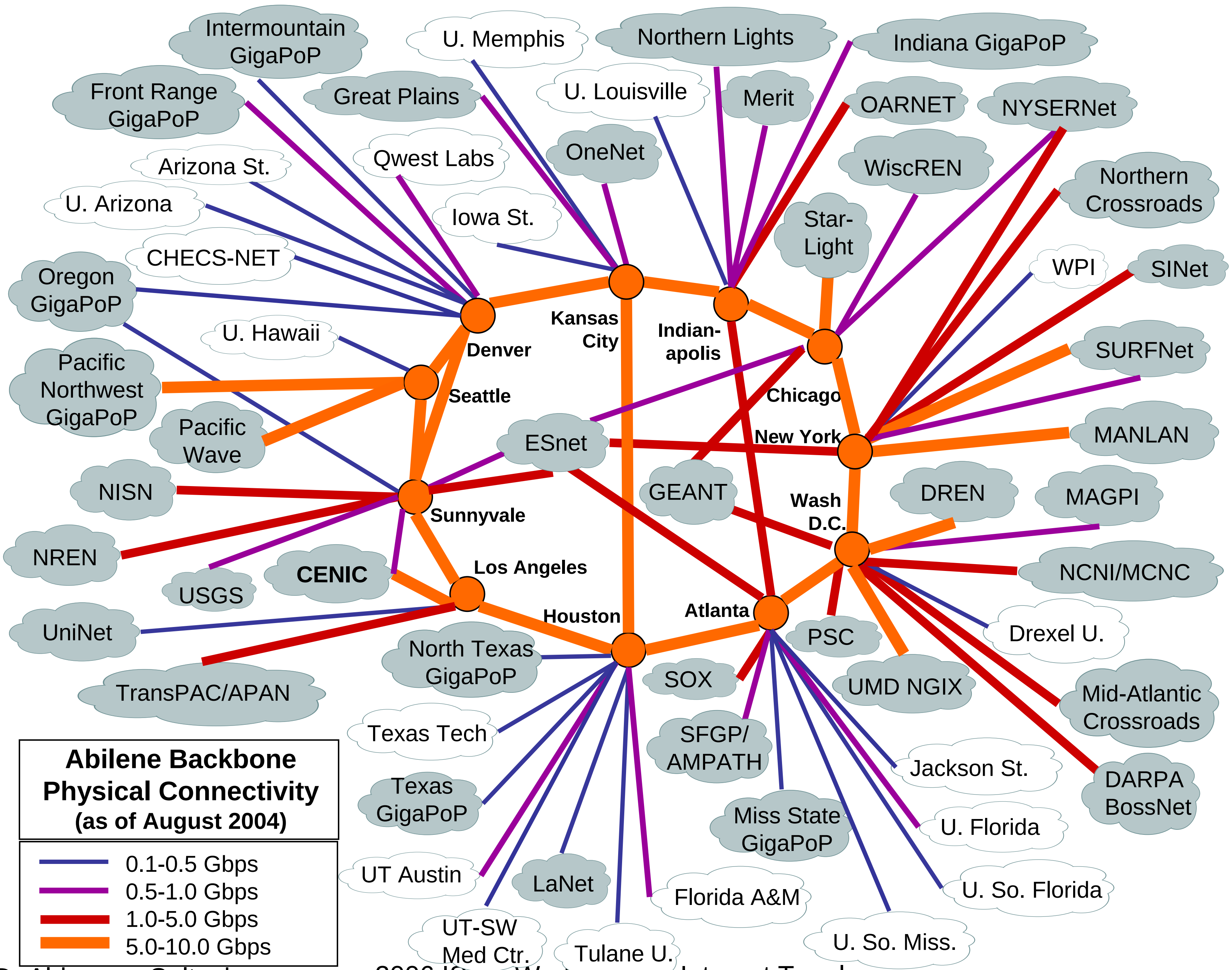


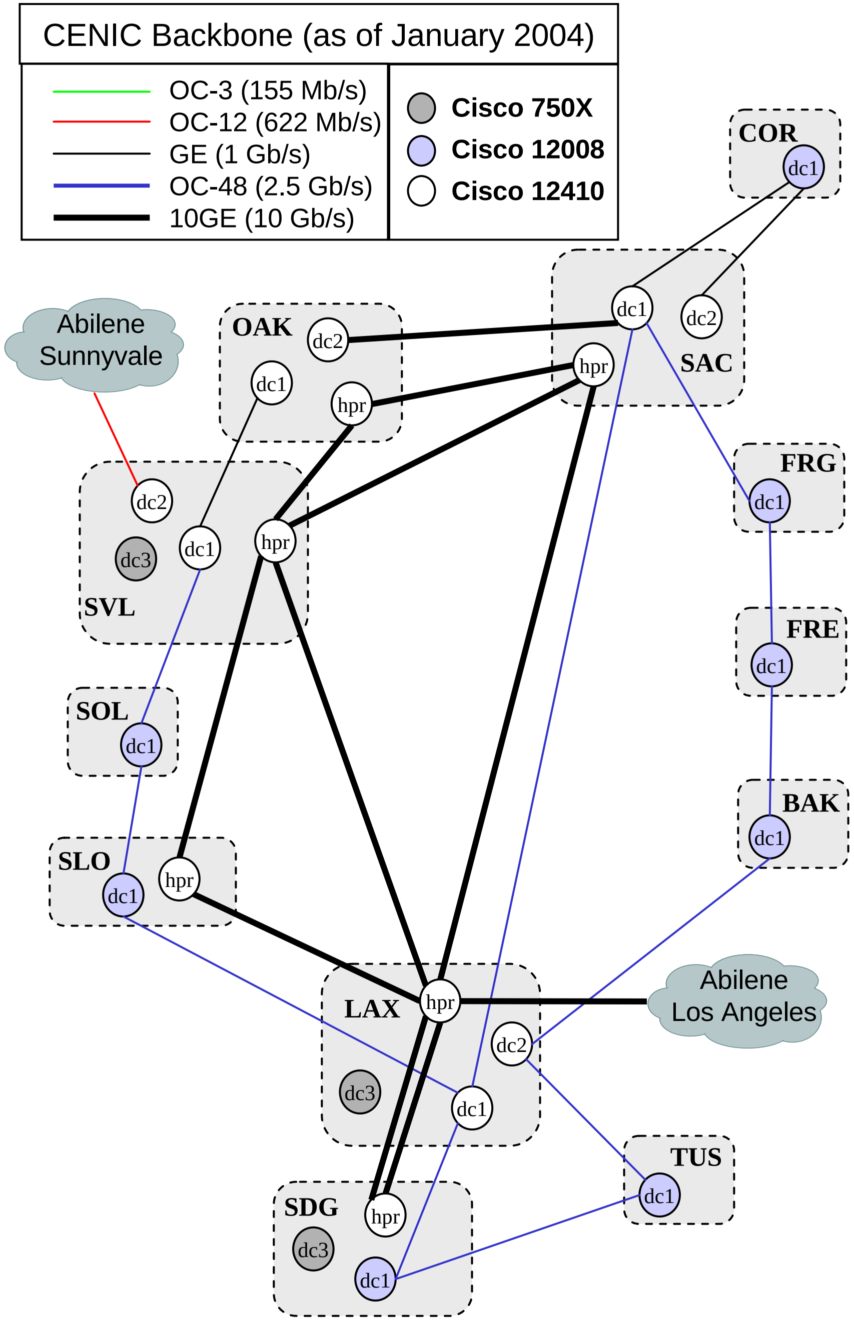
Relatively uniform connectivity within core.

High cost of links drives traffic aggregation at network edge

Possibly high variability in connectivity at edge.

High degree nodes are at the edges.


Intermountain GigaPoP
U. Memphis
Northern Lights
Indiana GigaPoP
Front Range GigaPoP
Great Plains
U. Louisville
Merit
OARNET
NYSERNet
Arizona St.
Qwest Labs
OneNet
WiscREN
Northern Crossroads
U. Arizona
Iowa St.
Star-Light
CHECS-NET
WPI
SINet
Oregon GigaPoP
Kansas City
Indian-apolis
Denver
U. Hawaii
SURFNet
Chicago
Seattle
Pacific Northwest GigaPoP
Pacific Wave
ESnet
New York
MANLAN
NISN
GEANT
DREN
MAGPI
Wash D.C.
Sunnyvale
NREN
CENIC
Los Angeles
NCNI/MCNC
USGS
Atlanta
Houston
UniNet
PSC
Drexel U.
North Texas GigaPoP
SOX
UMD NGIX
TransPAC/APAN
Mid-Atlantic Crossroads
Texas Tech
SFGP/ AMPATH
Jackson St.
DARPA BossNet
Texas GigaPoP
Miss State GigaPoP
U. Florida
UT Austin
LaNet
Florida A&M
U. So. Florida
UT-SW Med Ctr.
Tulane U.
U. So. Miss.
Abilene Backbone Physical Connectivity (as of August 2004)
0.1-0.5 Gbps
0.5-1.0 Gbps
1.0-5.0 Gbps
5.0-10.0 Gbps

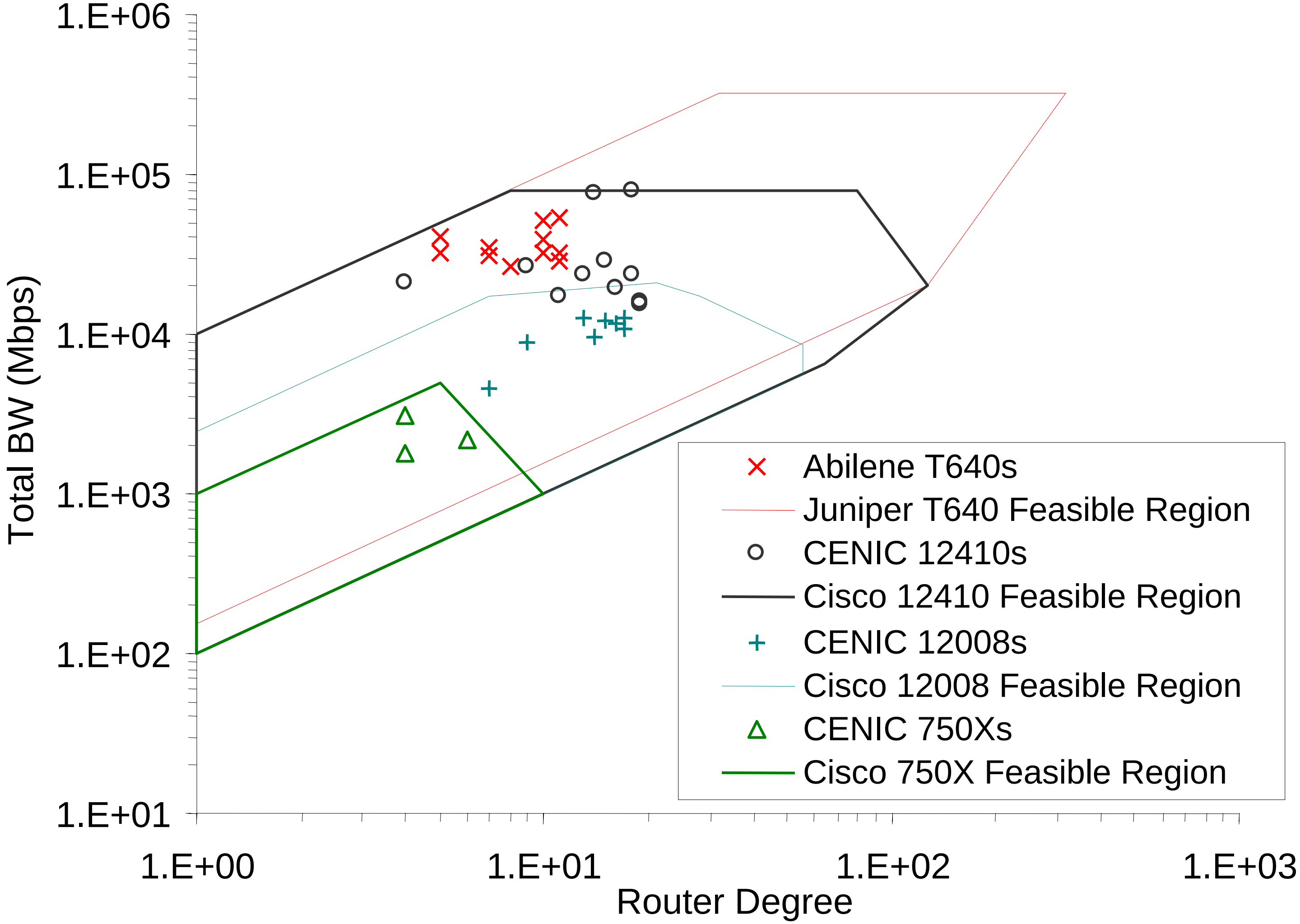


The Corporation for Education Network Initiatives in California (CENIC) acts as ISP for the state's colleges and universities
http://www.cenic.org

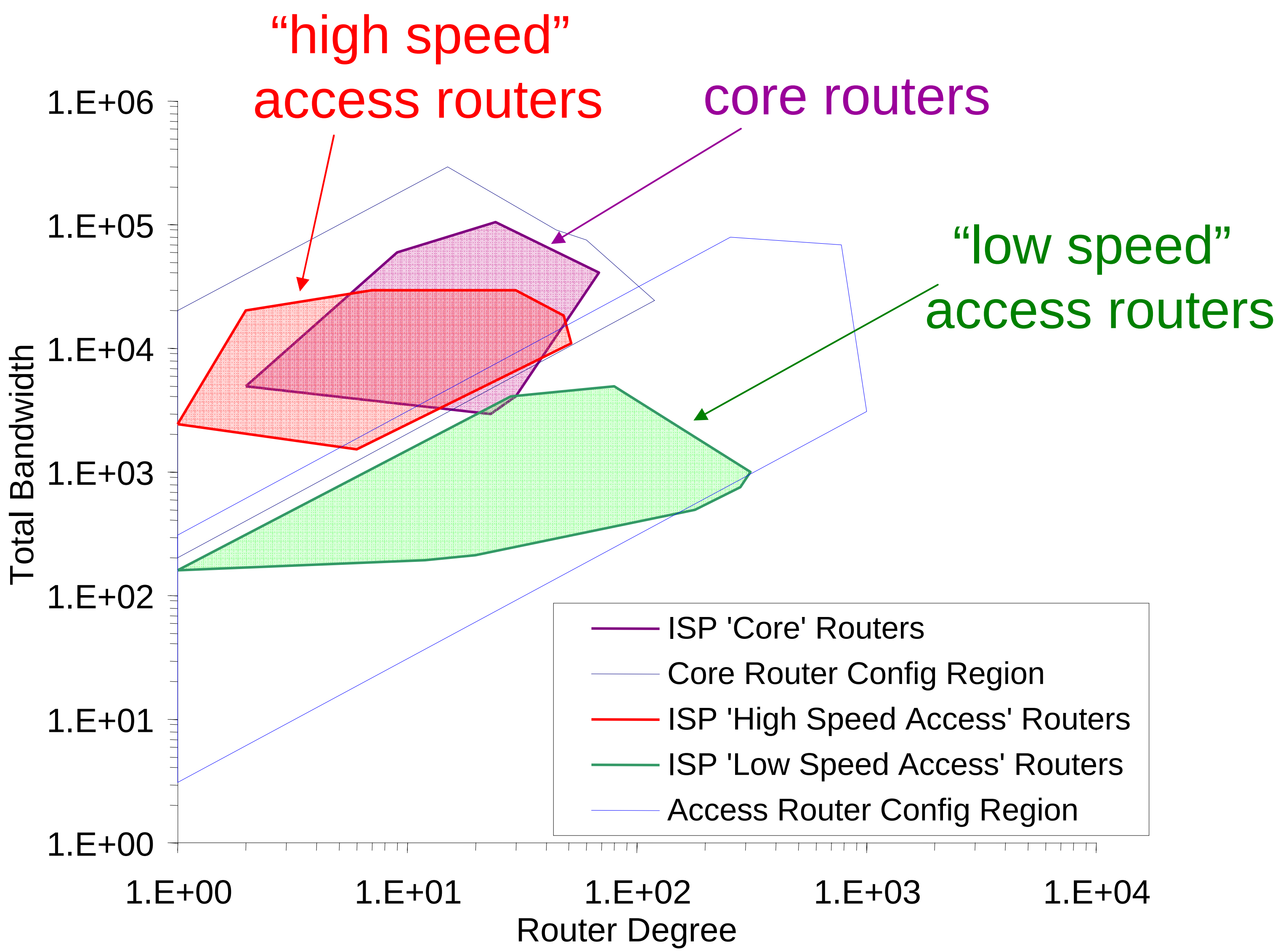
Like Abilene, its backbone is a sparsely-connected mesh, with relatively low connectivity and minimal redundancy.

- no high-degree hubs?
- no Achilles' heel?

Router Deployment: Abilene and CENIC

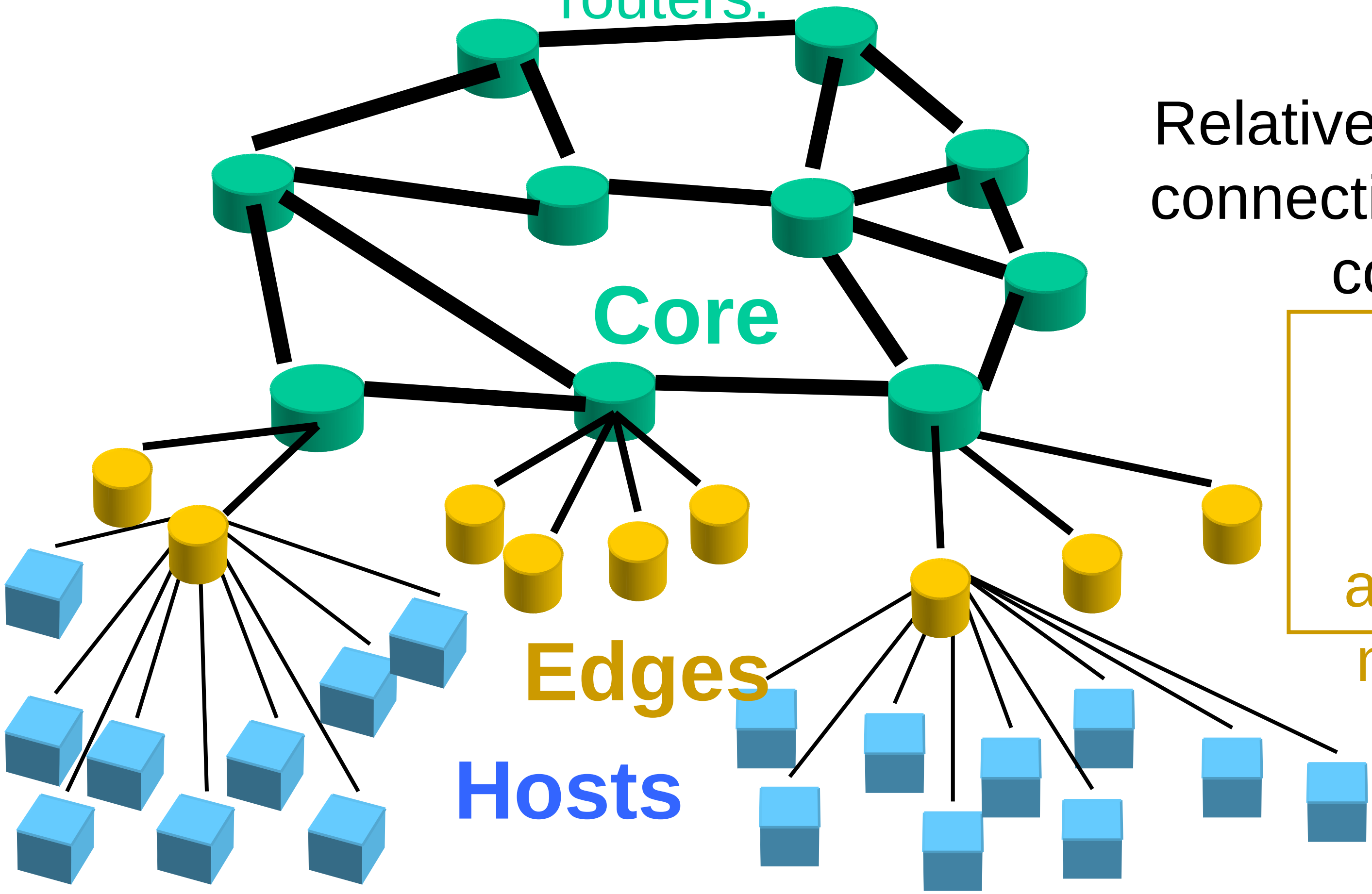


AT&T Router Deployment (c.2003)



Heuristically Optimal Topology

Sparse, mesh-like core of fast, low-degree routers.



Relatively uniform connectivity within core.

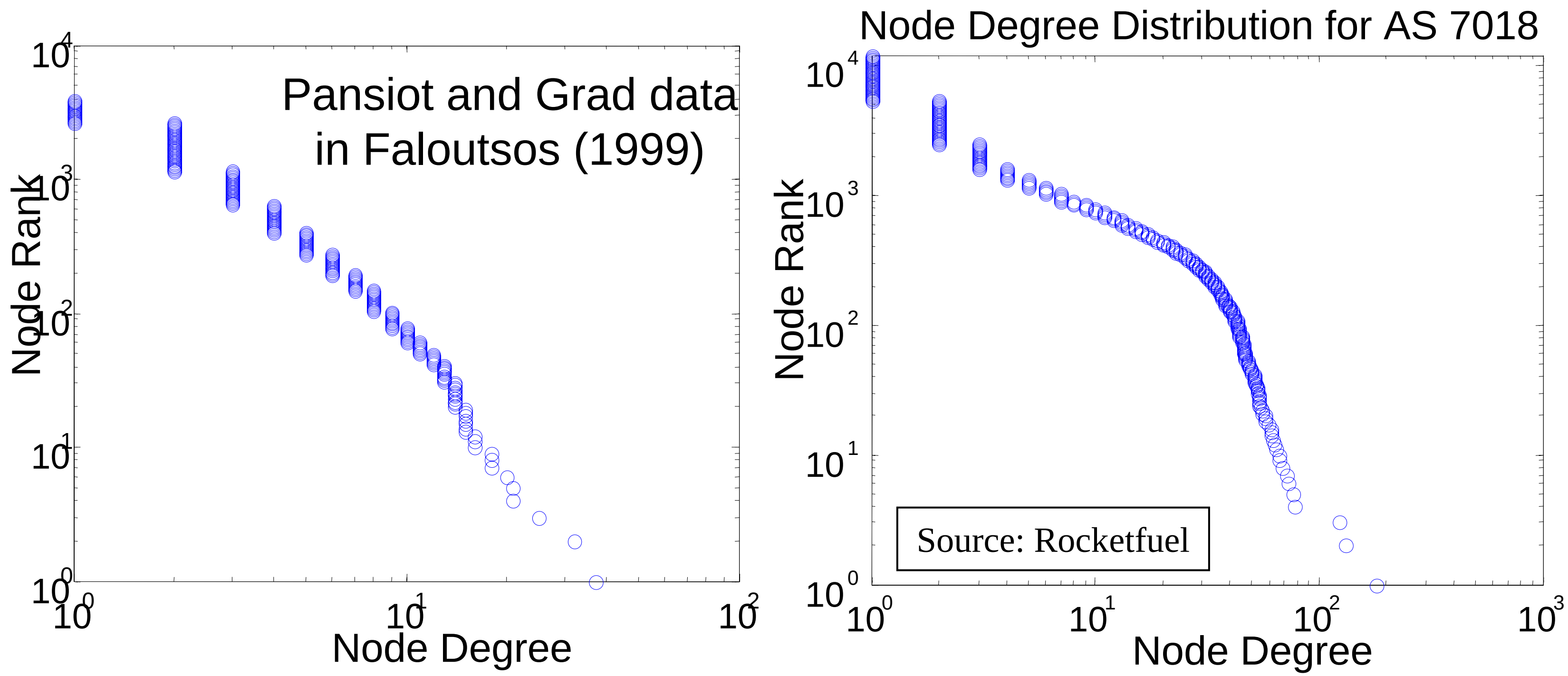
High cost of links drives traffic aggregation at network edge

Possibly high variability in connectivity at edge.

High degree nodes are at the edges.

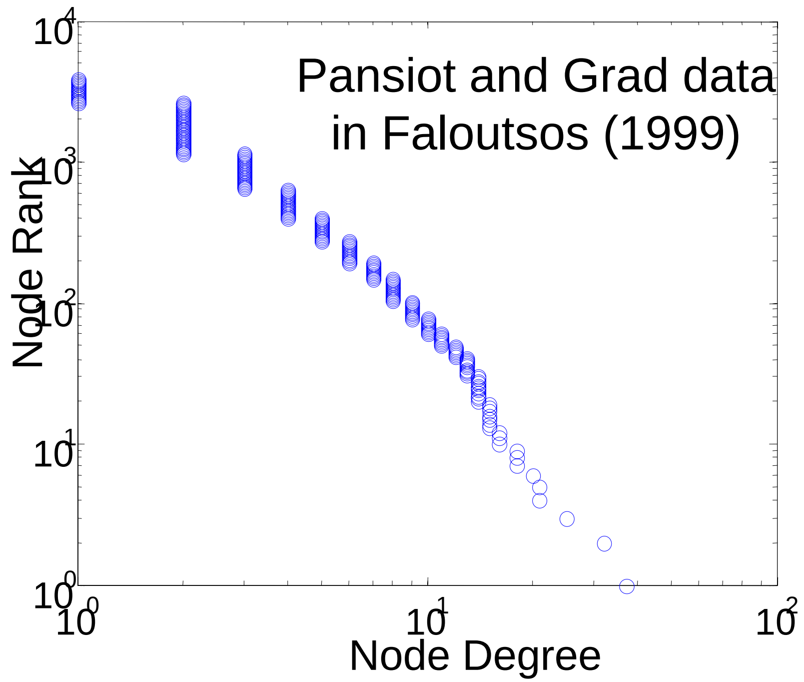
A Closer Look Router-Level Measurement Data

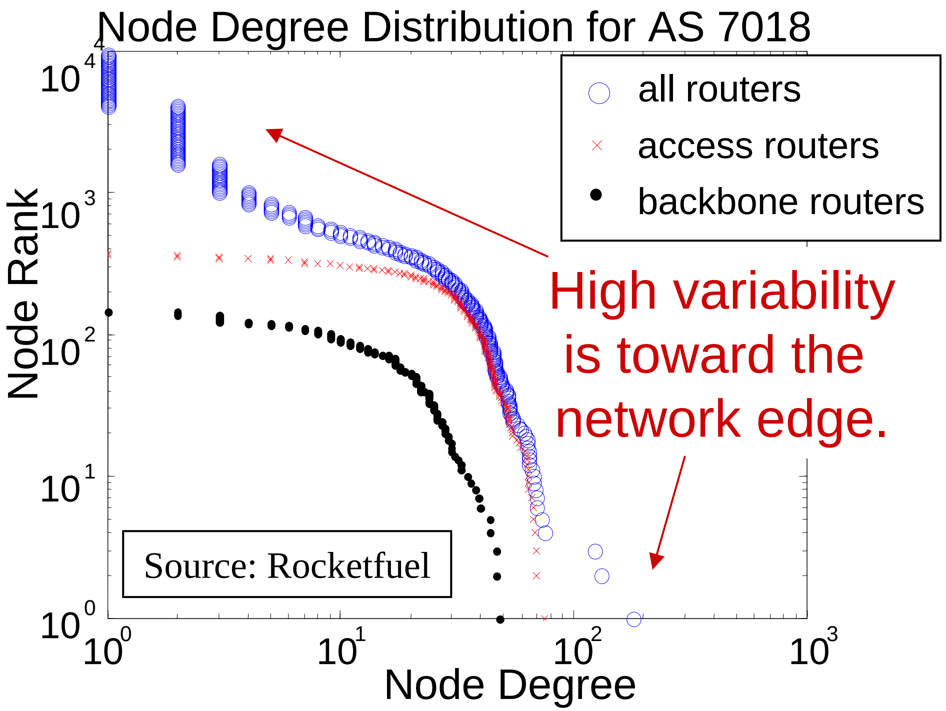
- Rocketfuel Project: Higher fidelity maps of individual ISPs
- Shows that core routers do not follow a power law distribution



A Closer Look Router-Level Measurement Data

- Rocketfuel Project: Higher fidelity maps of individual ISPs
- Shows that core routers do not follow a power law distribution





Again, traceroute measurement data requires careful scrutiny

Nonetheless, power laws in aggregate connectivity are plausible.

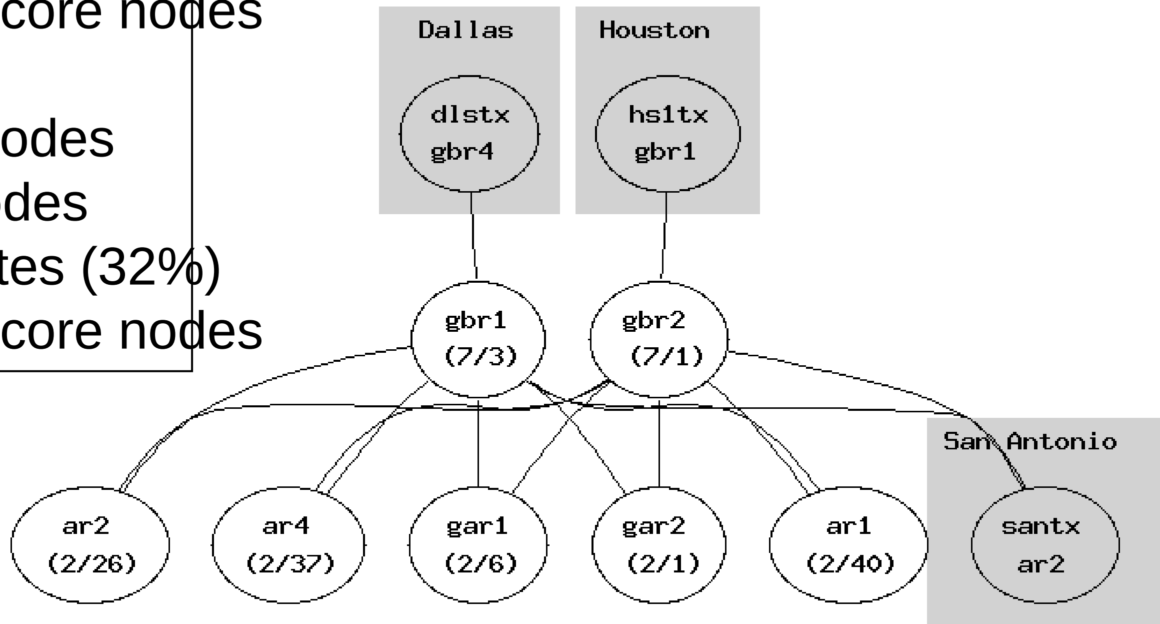
Rocketfuel: Interpretation, Validation, Augmentation

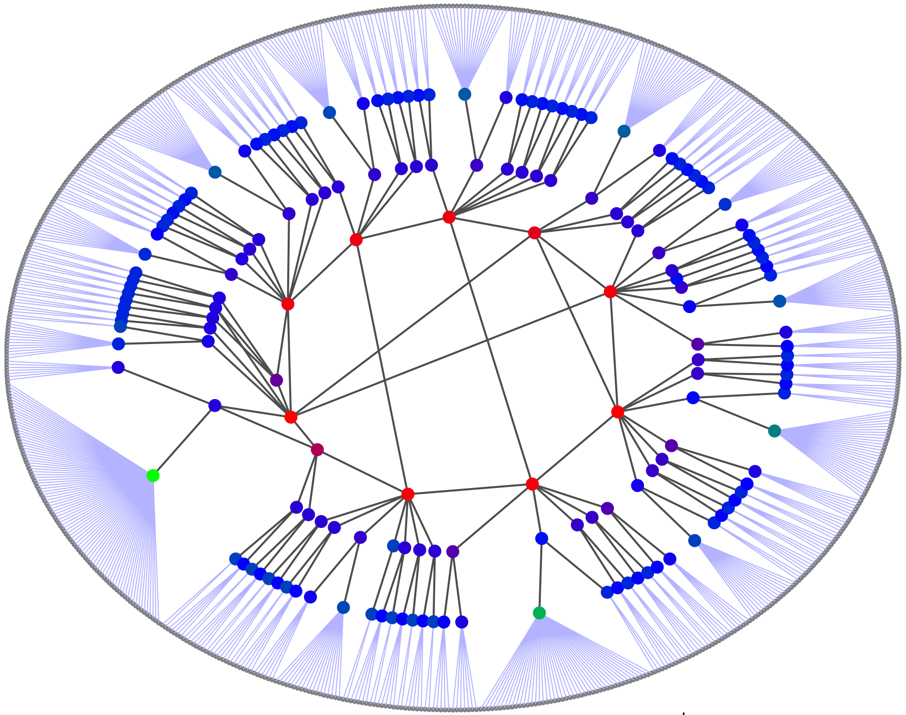
- application of “first principles” (e.g. router technology)
- alias resolution: discovery of duplicate nodes
- new graph annotation methods

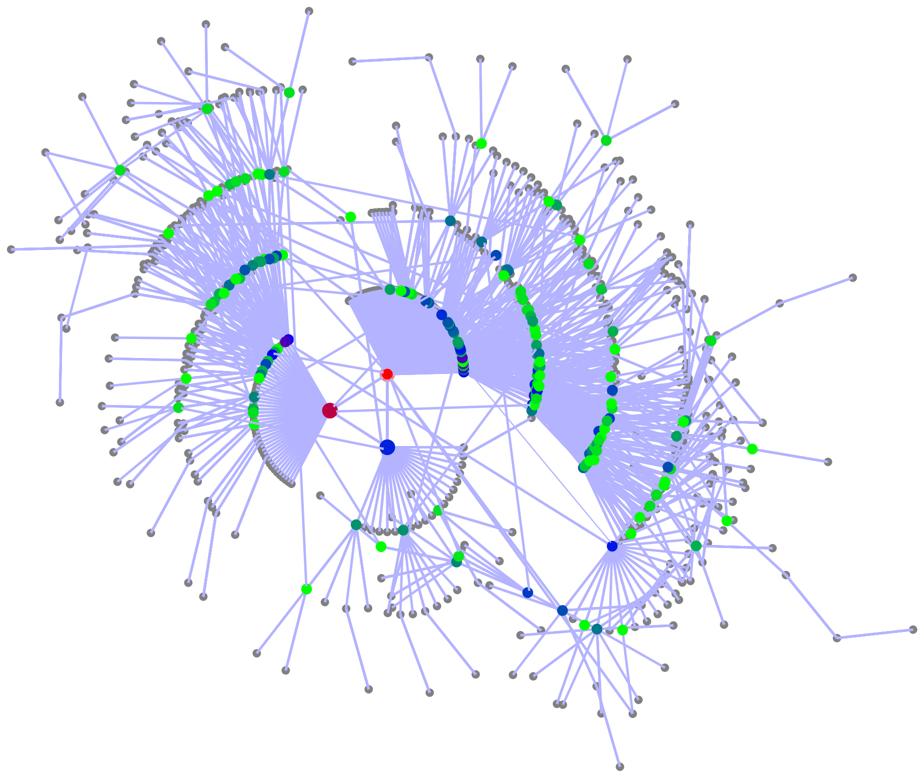
ISP Points of Presence (POPs) have <u>highly organized structure</u> (simplicity, hierarchy, redundancy)

AS 7018	9261 total nodes 640 core nodes 156 duplicates (24%) 484 unique core nodes
AS 1239	7043 total nodes 673 core nodes 215 duplicates (32%) 458 unique core nodes

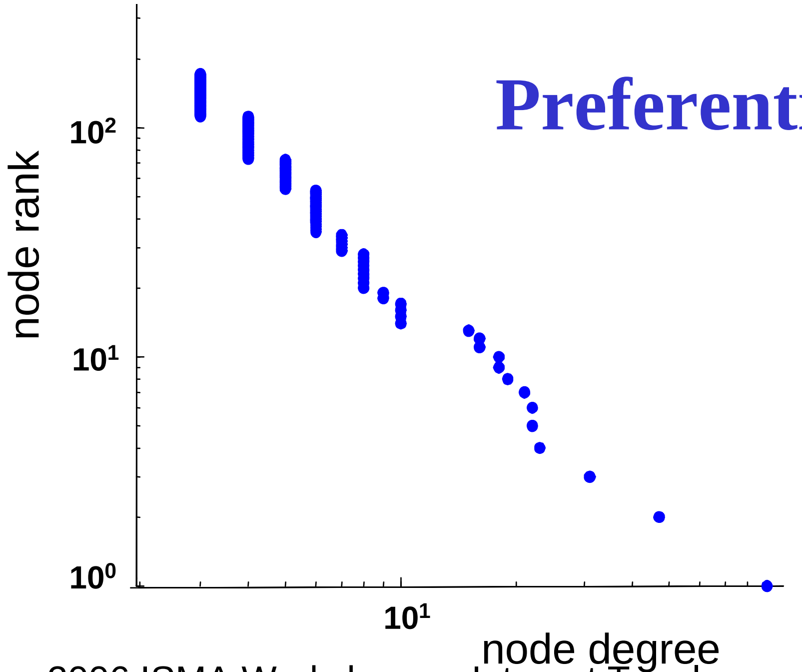
AS 7018: Austin, TX



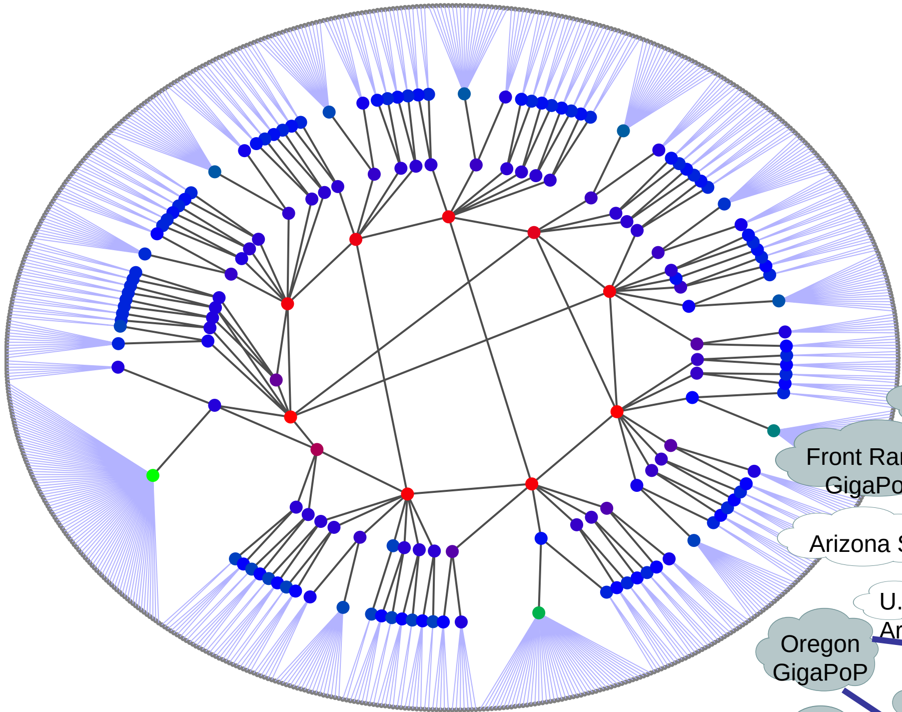




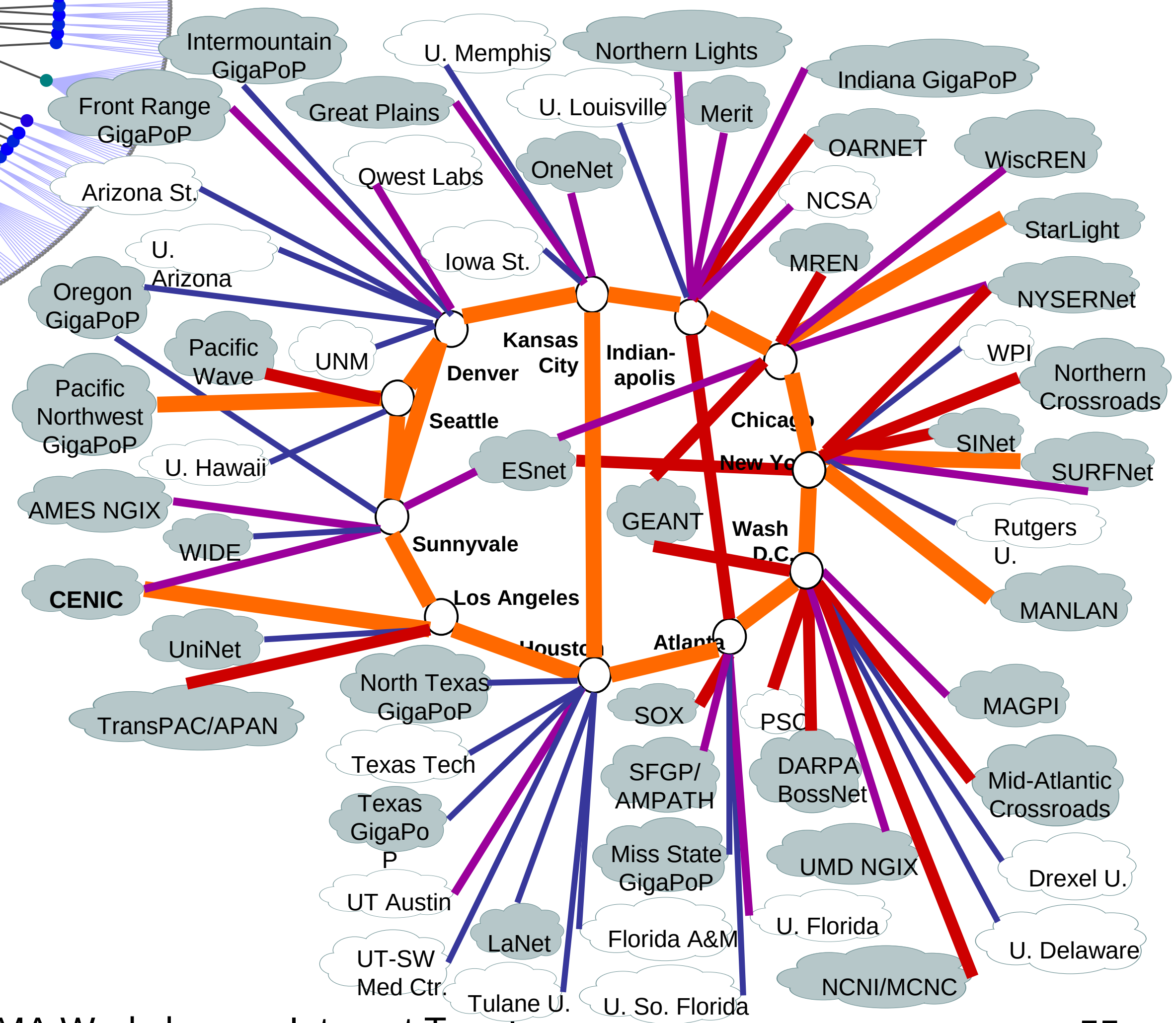
Preferential Attachment



Abilene Backbone Physical Connectivity



Abilene-inspired

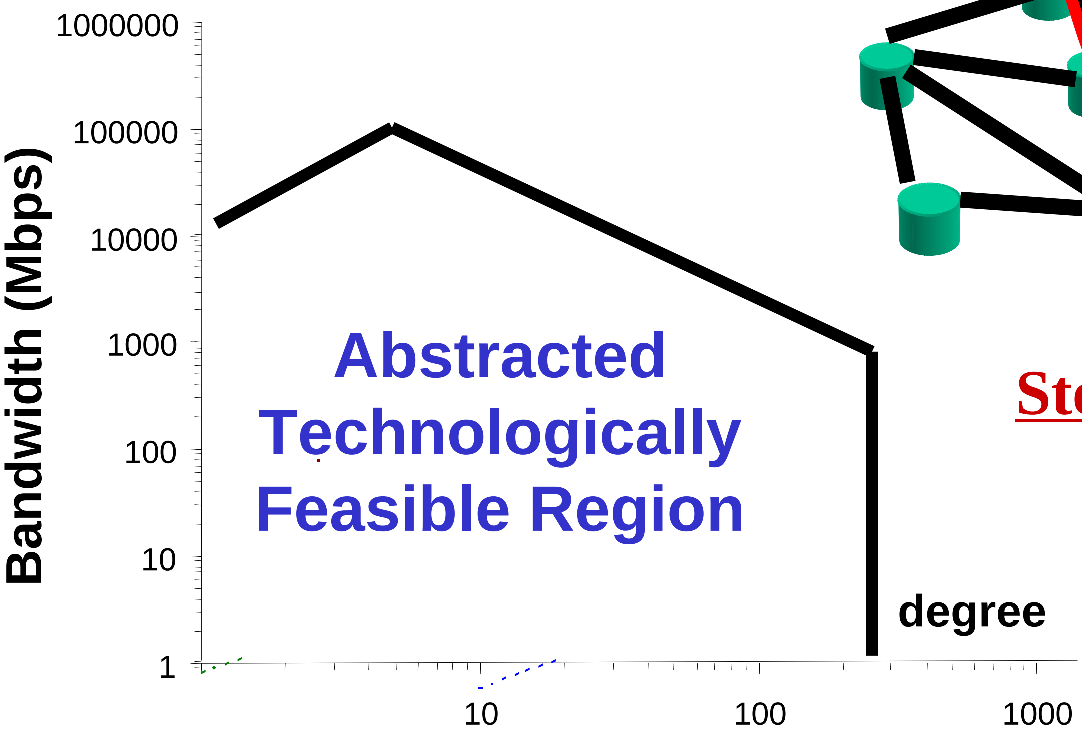


- 0.1-0.5 Gbps
- 0.5-1.0 Gbps
- 1.0-5.0 Gbps
- 5.0-10.0 Gbps

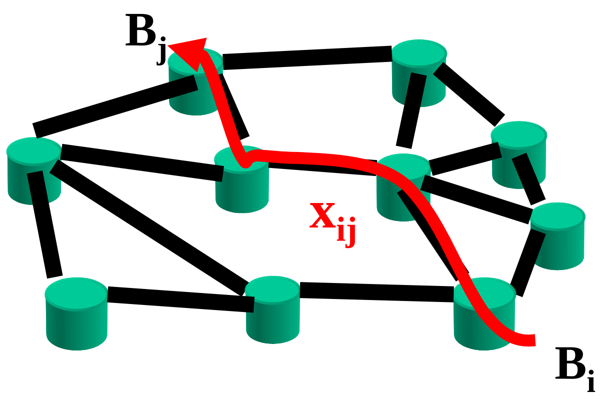
Network Performance

Given realistic technology constraints on routers, how well is the network able to carry traffic?

<u>Step 1</u>: Constrain to be feasible



<u>Step 2</u>: Compute traffic demand



$$x_{ij} \propto B_i B_j$$

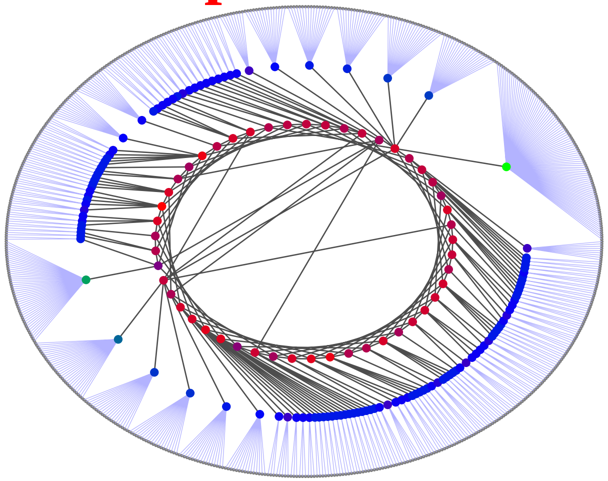
<u>Step 3</u>: Compute max flow

$$\max_{\alpha} \sum_{i,j} x_{ij} = \max \sum_{i,j} \alpha B_i B_j$$

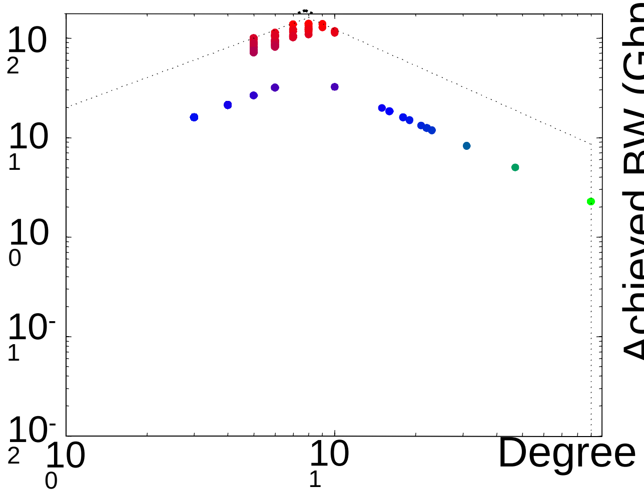
$$s.t. \sum_{i,j:k \in r_{ij}} x_{ij} \leq B_k, \forall k$$

Structure Determines Performance

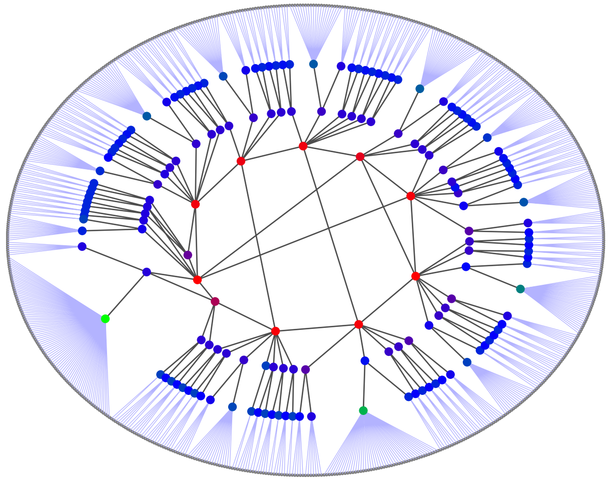
Heuristically Optimized



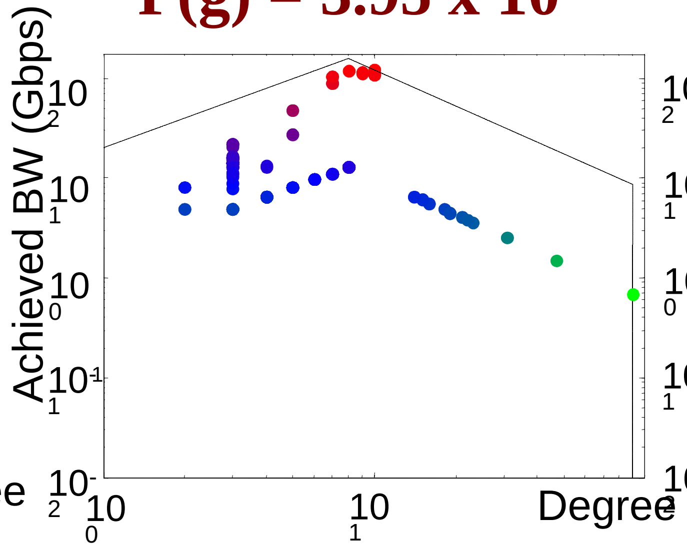
P(g) = 1.13 x 10^{12}



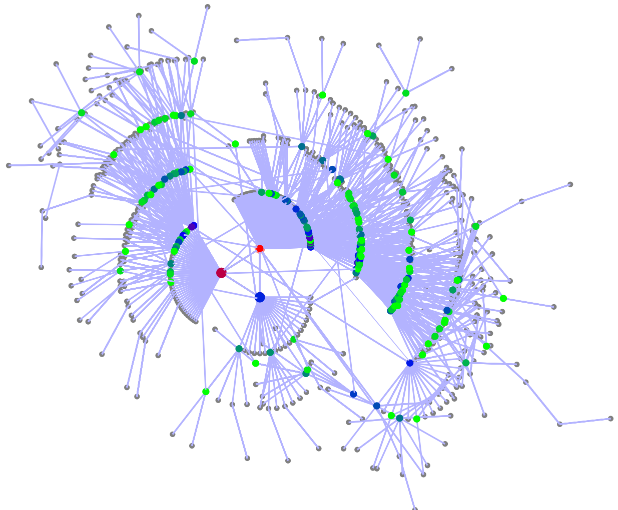
Abilene-inspired



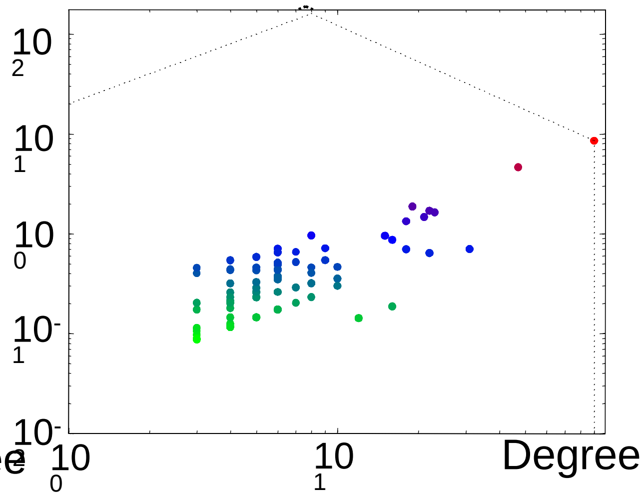
P(g) = 3.95 x 10^{11}



PA



P(g) = 1.19 x 10^{10}



Structural Metric

Define the metric $$s(g) = \sum_{\substack{i,j \\ connected}} d_i d_j$$ (d_i = degree of node i)

- Introduced in SIGCOMM'04
- Easily computed for any graph
- Depends on the structure of the graph, not the generation mechanism
- Measures how "hub-like" the network core is

We can renormalize so that $0 \leq S(g) \leq 1$

$$S(g) = \frac{s(g)}{s_{\max}}$$

where $s_{\max}$ has the largest value of $s(g)$ among all graphs g having the same degree distribution.

Properties of the S(g) metric

- Captures all of the information in degree correlation statistics [Dorogovtsev and Mendes, 2003]
- Closely related to notion of assortativity [Newman, 2002]
- Graphs with high-S(g) have certain self-similar properties as defined by notions of rewiring, coarse graining, trimming
- For graphs resulting from probabilistic construction (e.g. PLRG/GRG), LogLikelihood (LLH) $\propto S(g)$

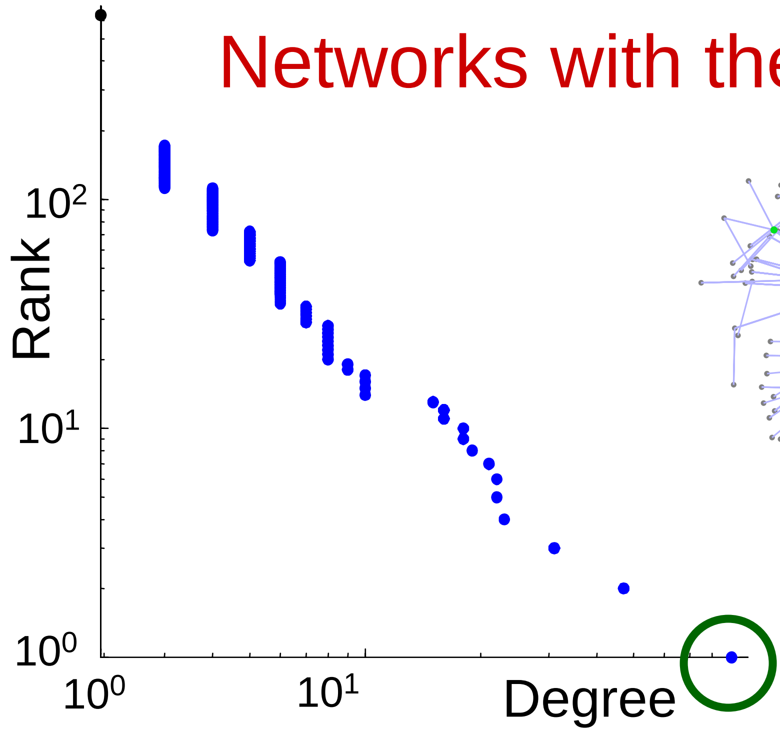
Relevant Interpretation: How likely is a particular graph (having given degree sequence) to arise *at random*?

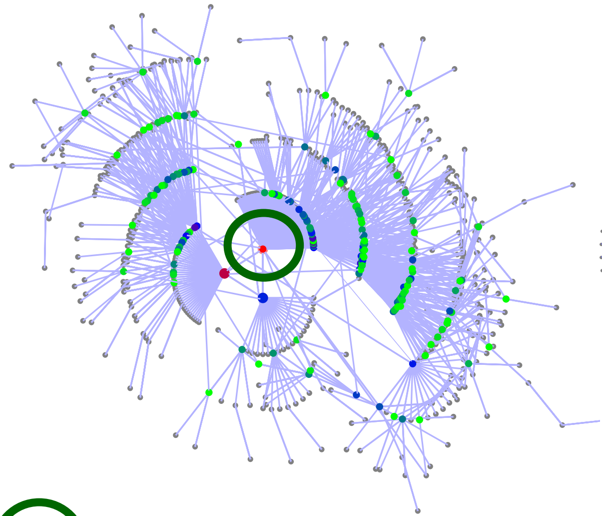
see also the talk by Lun Li on Thurs I

For details:

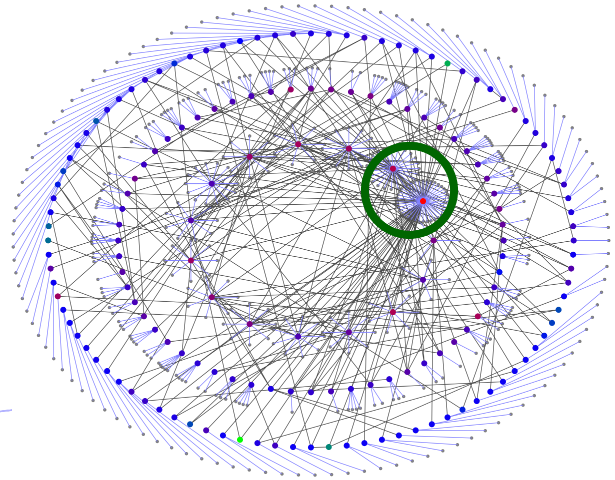
L. Li, D. Alderson, J.C. Doyle, W. Willinger. Toward a Theory of Scale-Free Networks: Definition, Properties, and Implications. *Internet Mathematics*, In Press (2006).

Networks with the Same Degree Sequence

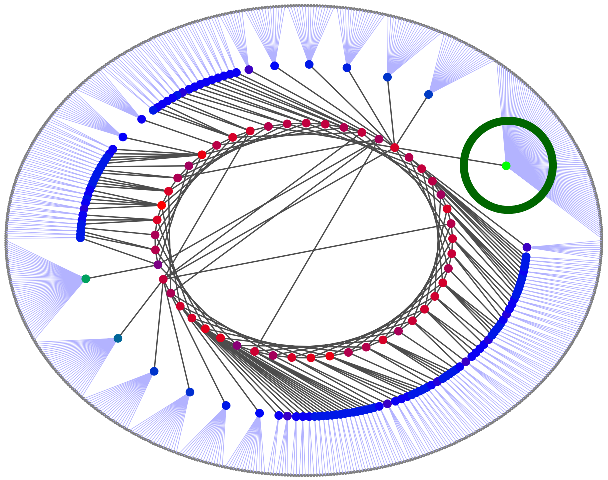




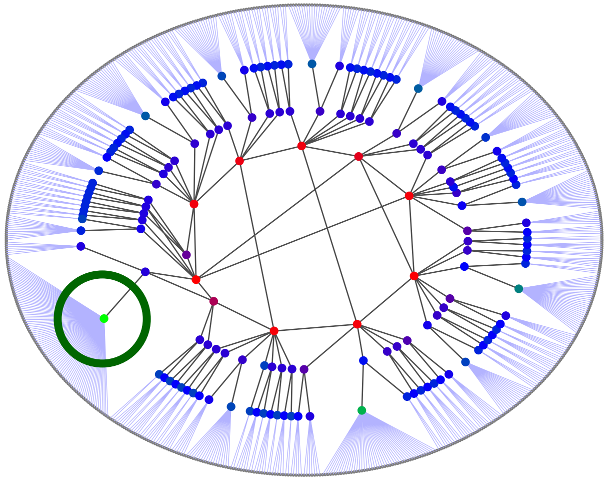
PA



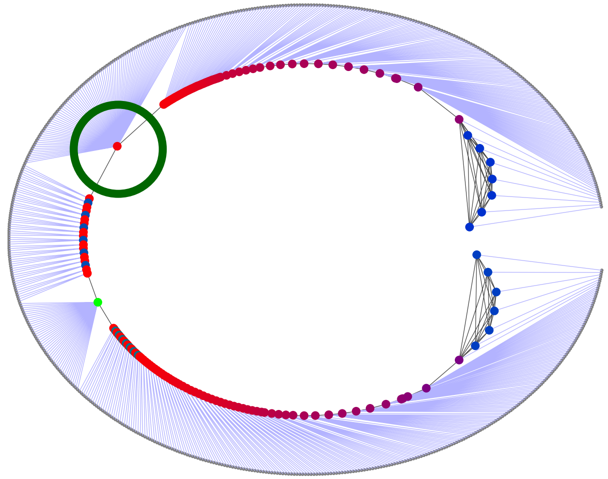
PLRG



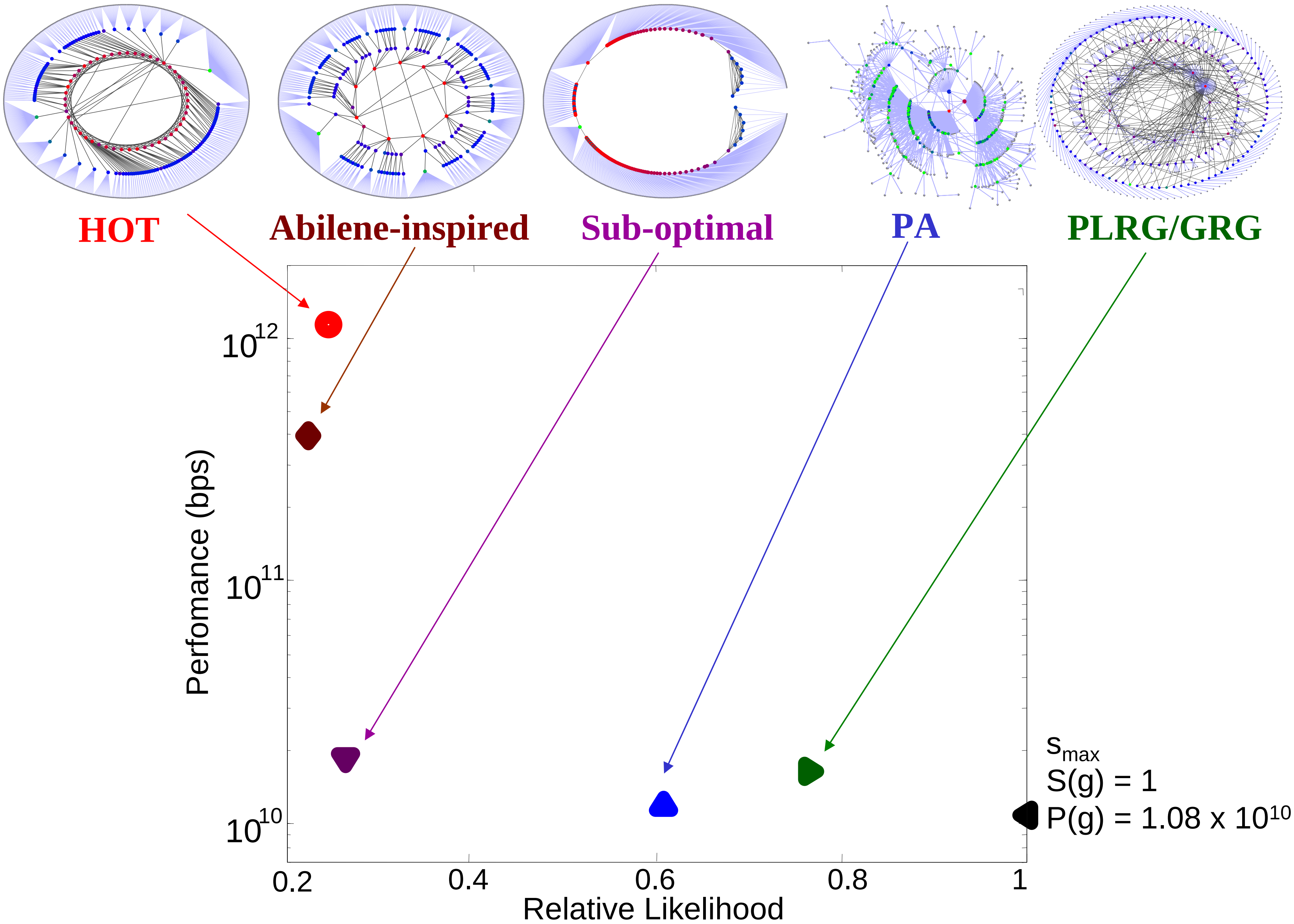
Heuristically Optimal

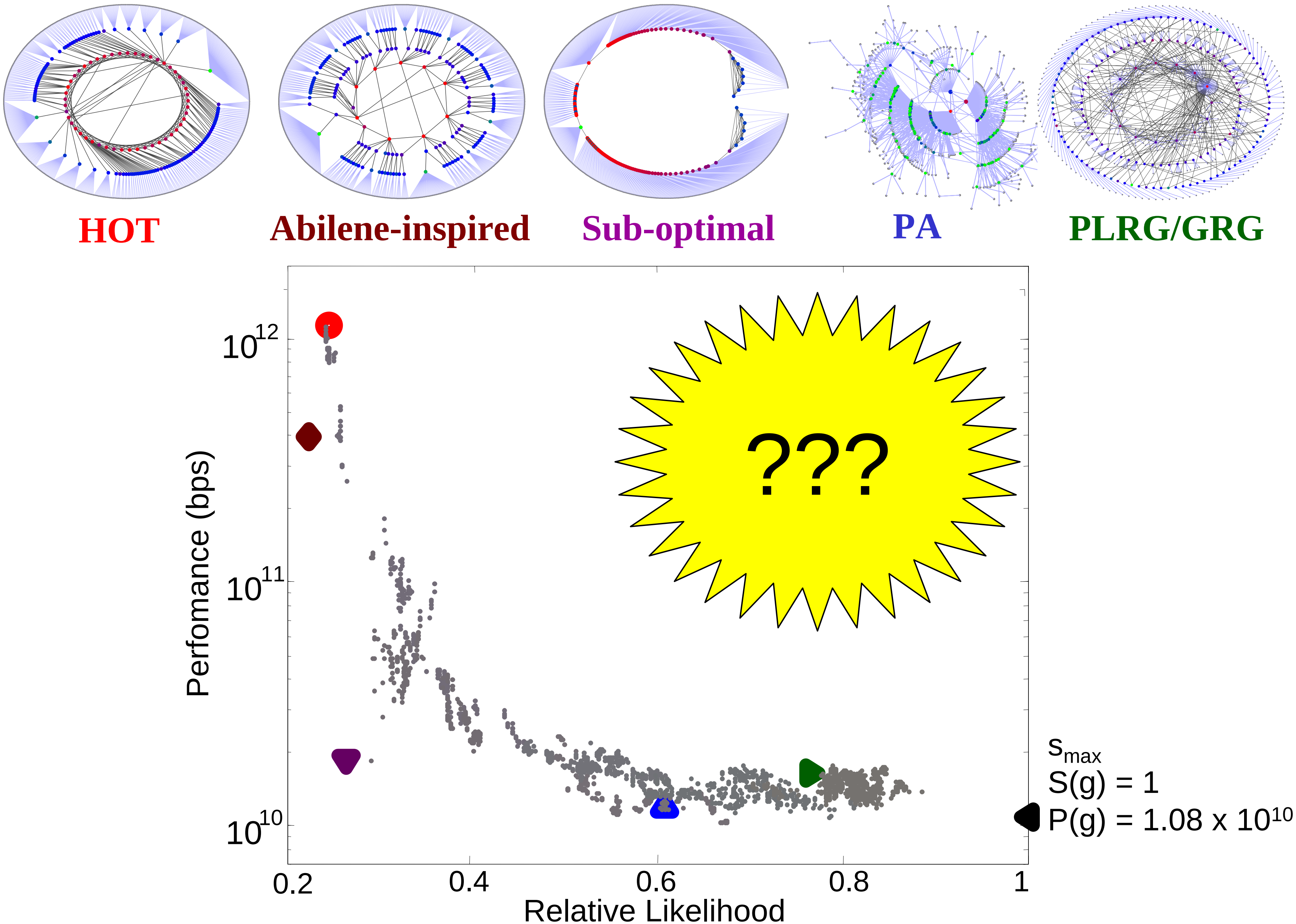


Abilene-inspired

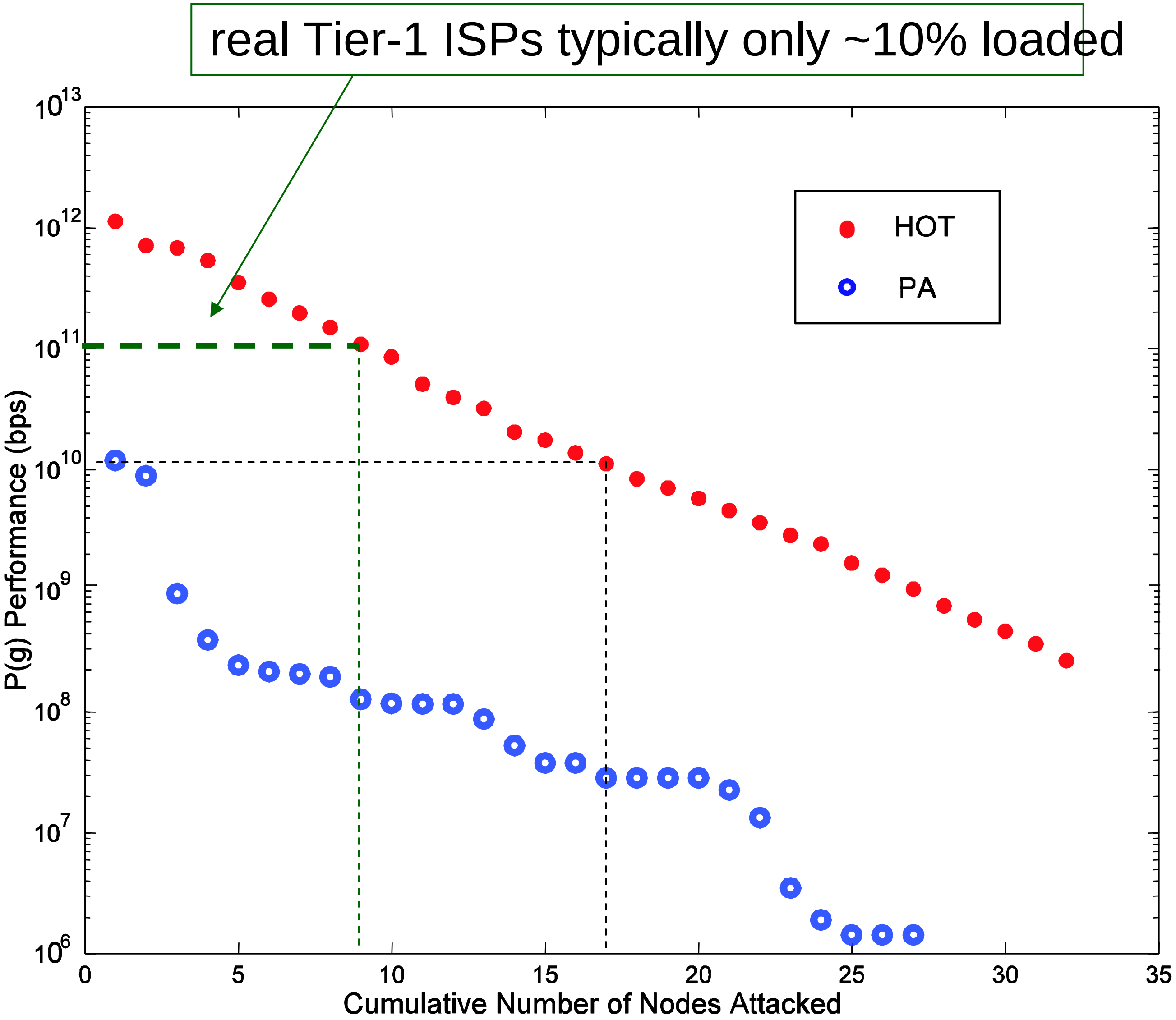


Sub-optimal

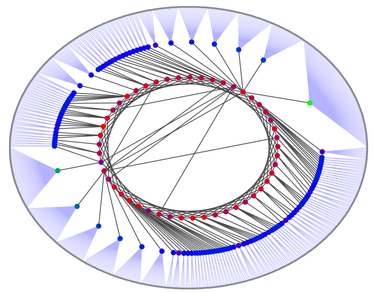

HOT
Abilene-inspired
Sub-optimal
PA
PLRG/GRG
Perfomance (bps)
Relative Likelihood
10^{12}
10^{11}
10^{10}
0.2
0.4
0.6
0.8
1
s_{max}
S(g) = 1
P(g) = 1.08 x 10^{10}


HOT
Abilene-inspired
Sub-optimal
PA
PLRG/GRG
10^{12}
10^{11}
10^{10}
Perfomance (bps)
0.2
0.4
0.6
0.8
1
Relative Likelihood
???
s_{max}
S(g) = 1
P(g) = 1.08 x 10^{10}

Response to router attack

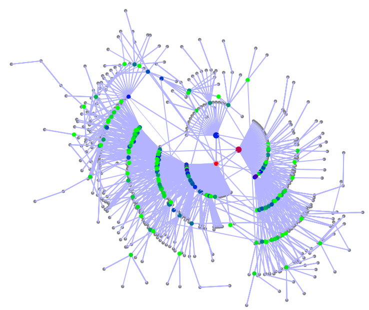


HOT



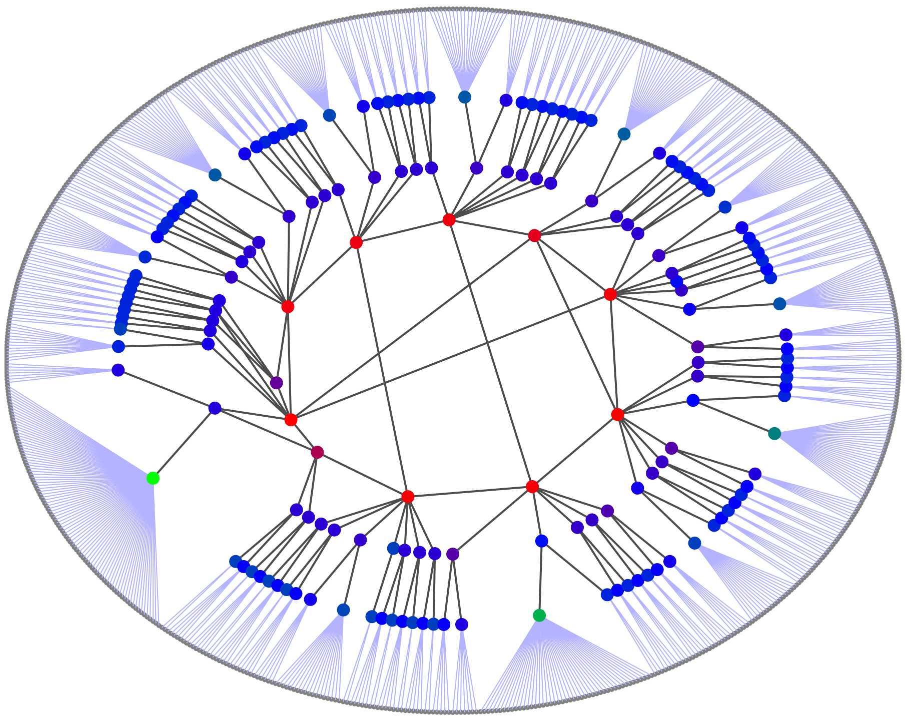
Worst case = low-degree core router

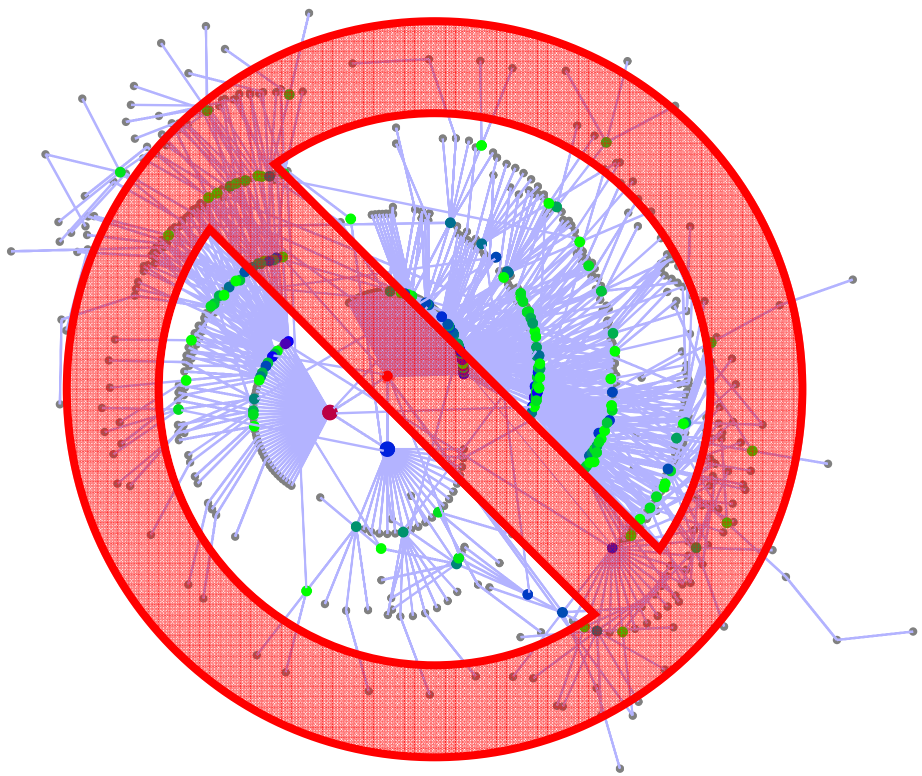
PA



Worst case = high degree central hub

MESSAGE #4: importance of model validation





- Descriptive modeling that replicates statistical features is no more than an exercise in “data fitting”
- Emphasis on “closing the loop” (using complementary measurements and domain expertise)

Summary: What "Matters" For Router-Level Topologies?

- The "nodes" and "links" are physical things that have hard constraints (technology).
- ISPs are constrained in what they can afford to build, operate, and maintain (economics).
- Decisions of ISPs are driven by objectives (performance) and reflect tradeoffs between what is feasible and what is desirable (heuristic optimization)
- Many important questions (robustness) only make sense in the context of the broader system (protocol stack)

PITFALL: Emphasis on power laws

- "Full of sound and fury, signifying nothing?" (Strogatz)
- Power laws as "more normal than Normal" (ask

http://hot.caltech.edu/topology.html

- L. Li, D. Alderson, J.C. Doyle, W. Willinger. *Toward a Theory of Scale-Free Networks: Definition, Properties, and Implications.* *Internet Math*. In Press (2006).
- D. Alderson, L. Li, W. Willinger, J.C. Doyle. Understanding Internet Topology: Principles, Models, and Validation. *IEEE Trans. on Networking.* 13(6): Dec 2005.
- J.C. Doyle, D. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger. The "robust yet fragile" nature of the Internet. *PNAS.* October 4, 2005.
- D. Alderson, W. Willinger, L. Li, and J. Doyle. The Role of Optimization in Understanding and Modeling Internet Topology. In *Telecommunications Planning: Innovations in Pricing, Network Design and Management*. S. Raghavan and G. Anandlingham, eds. Springer, 2005.
- D. Alderson and W. Willinger. A contrasting look at self-organization in the Internet and next-generation communication networks. *IEEE Comm. Magazine*. July 2005.
- W. Willinger, D Alderson, J.C. Doyle, and L. Li, *More "normal" than Normal: scaling distributions in complex systems. Proc. Winter Simulation Conf. 2004.*
- W. Willinger, D Alderson, and L. Li, *A pragmatic approach to dealing with high-variability in network measurements, Proc. ACM SIGCOMM IMC 2004.*
- L. Li, D. Alderson, W. Willinger, and J. Doyle, *A first-principles approach to understanding the Internet's router-level topology, Proc. ACM SIGCOMM 2004.*
- D. Alderson, J. Doyle, R. Govindan, and W. Willinger. *Toward an*