On Measuring, Inferring, and Modeling Internet Connectivity: A Guided Tour across the TCP/IP Protocol Stack

Walter Willinger
AT&T Labs-Research
walter@research.att.com

Acknowledgments

Main Collaborators

- John Doyle (Caltech)
- David Alderson (Caltech)
- Lun Li (Caltech)

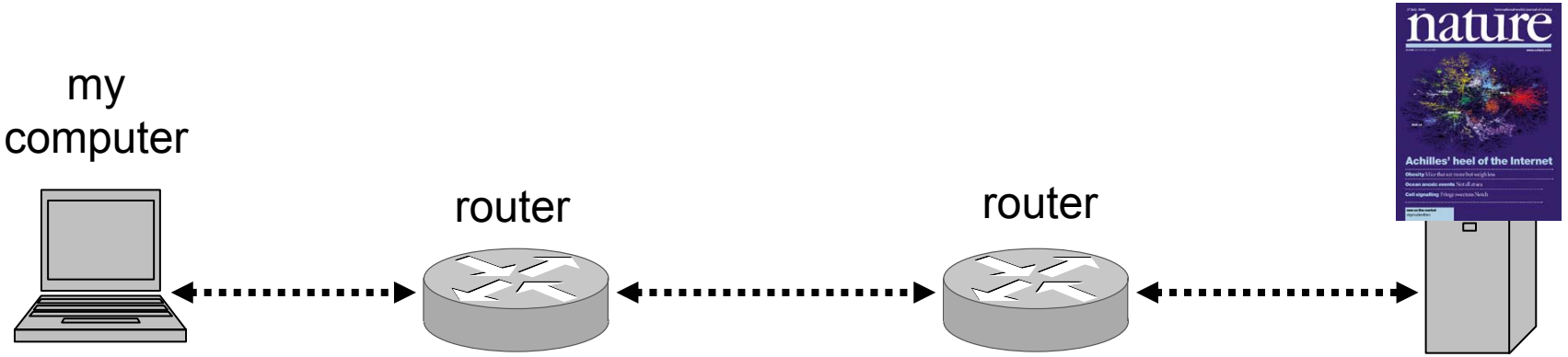
Contributions

- Reiko Tanaka (RIKEN, Japan)
- Matt Roughan (Univ. of Adelaide, Australia)
- Steven Low (Caltech)
- Ramesh Govindan (USC)
- Hyunseok Chang, Sugih Jamin, Morely Mao (Univ. of Michigan)
- Stanislav Shalunov (Abilene)
- Heather Sherman (CENIC)
- Daniel Stutzbach, Reza Rejaie (Univ. of Oregon)
- S. Sen, Nick Duffield (AT&T Labs-Research)

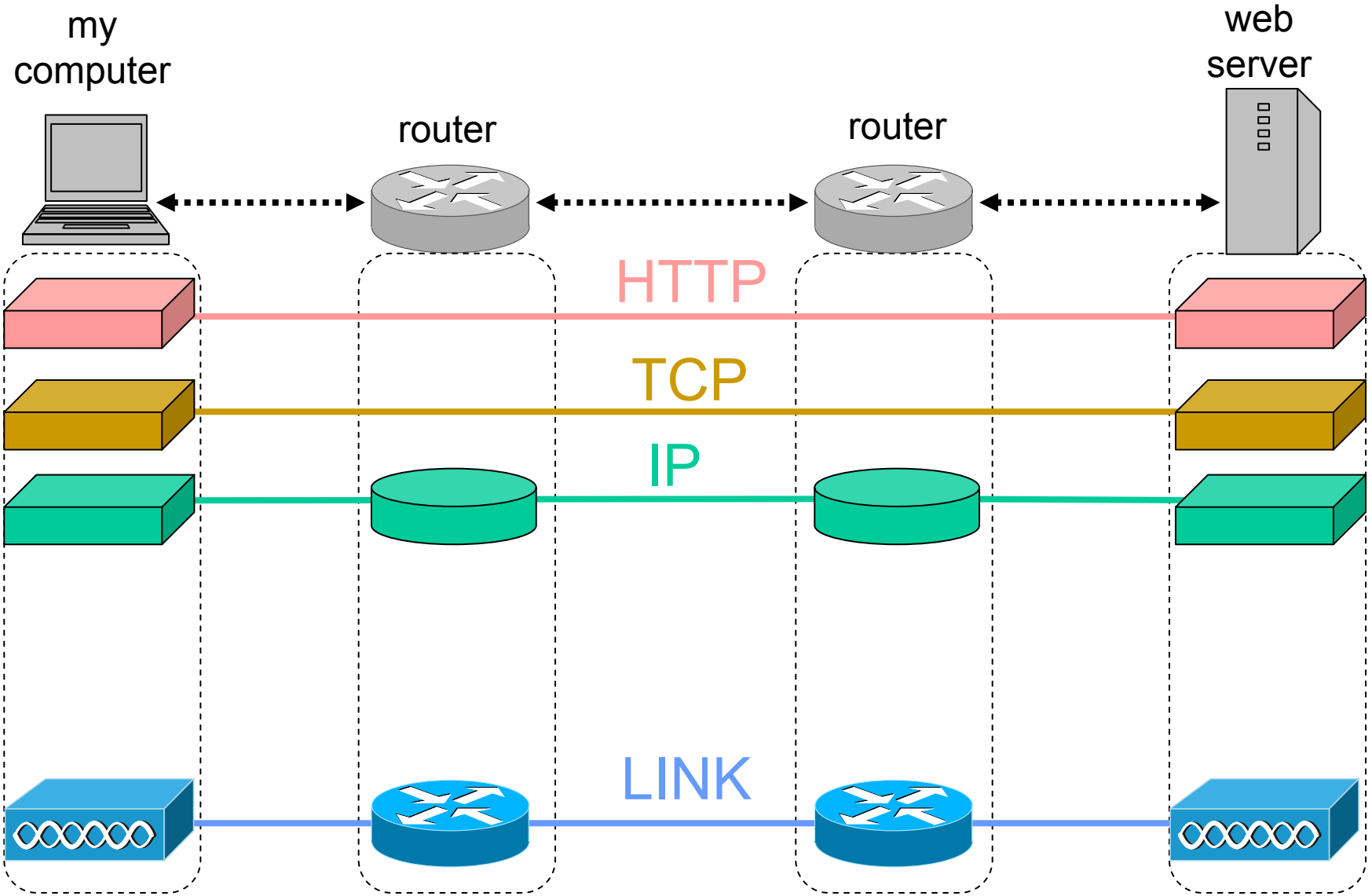
Outline

- Internet architecture and Internet topology
- On measuring Internet connectivity
- On inferring Internet connectivity structures
- On modeling Internet connectivity
- On validating models of Internet connectivity structures

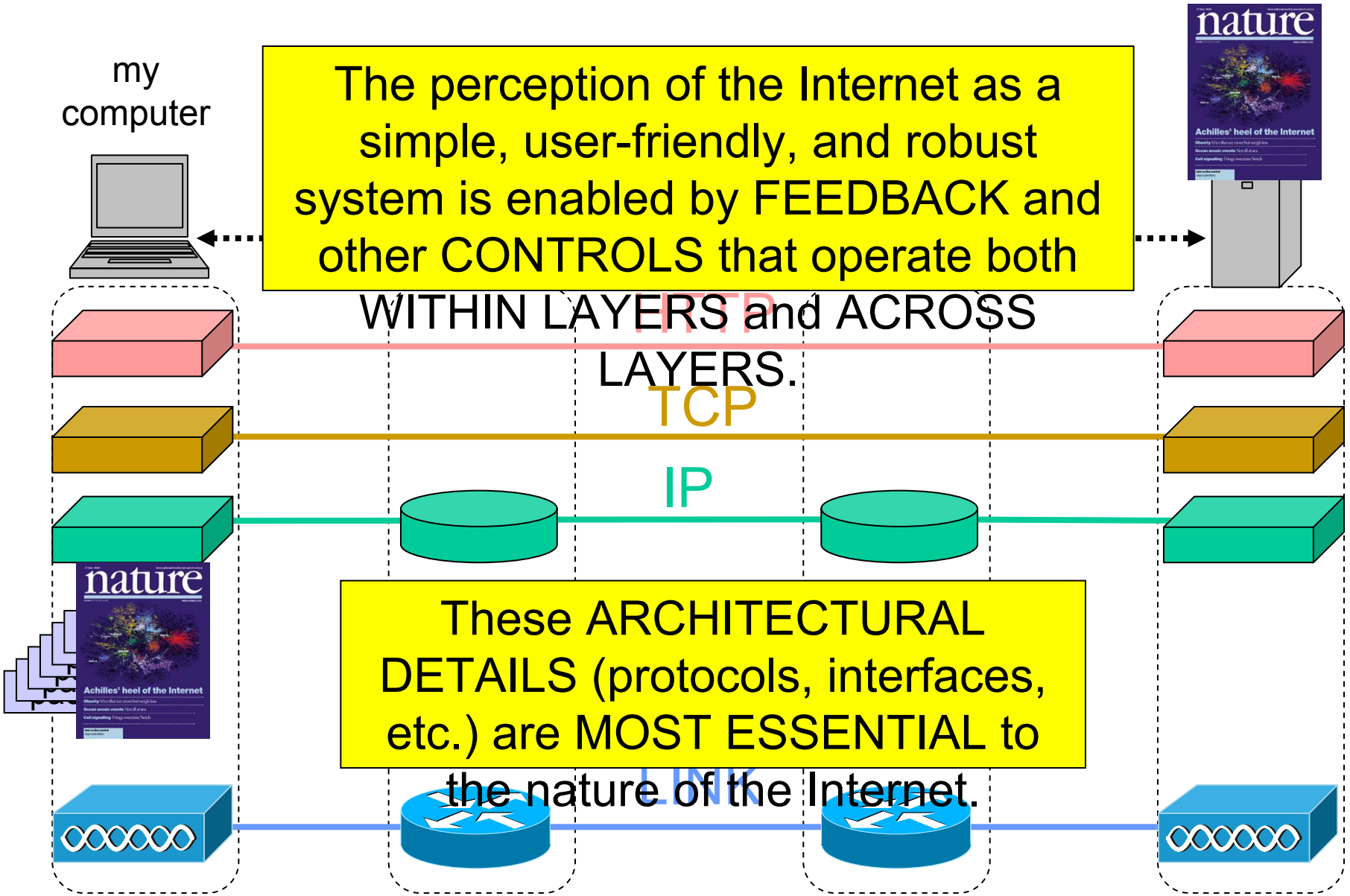
The Internet: The User Perspective



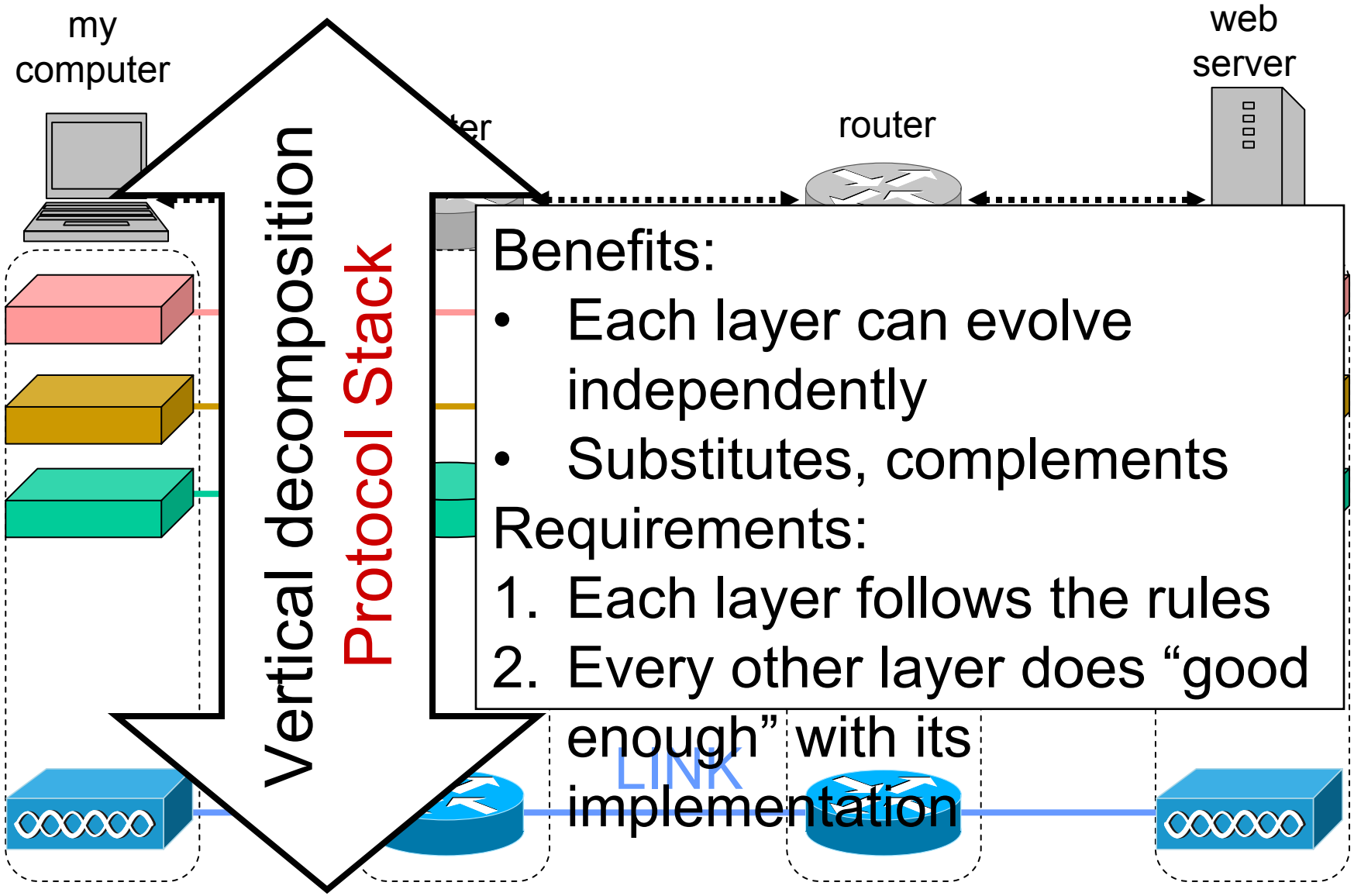
The Internet: The Engineering Perspective



The Internet is a LAYERED Network



Internet Architecture: Vertical Decomposition



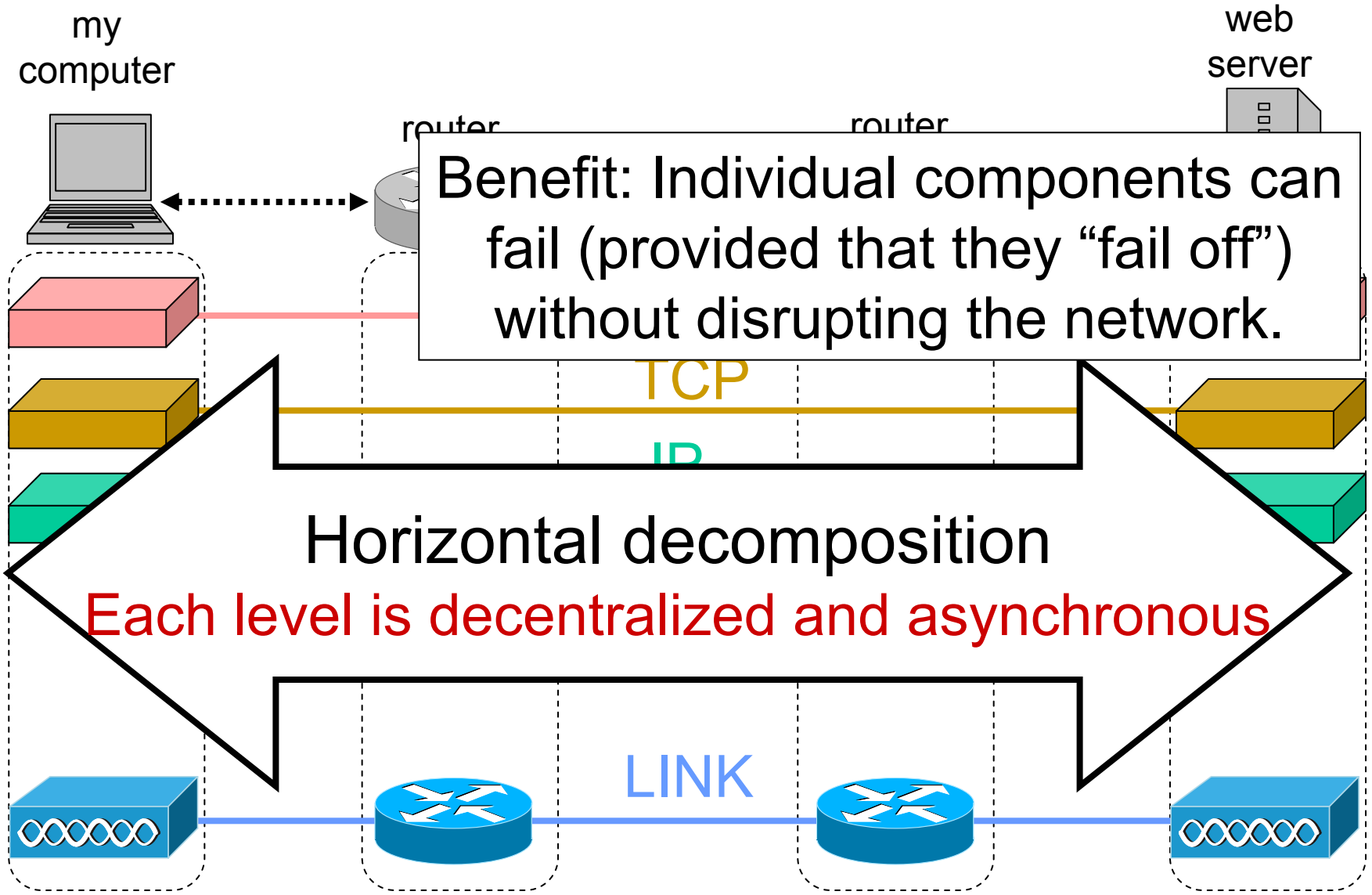
Benefits:

- Each layer can evolve independently
- Substitutes, complements

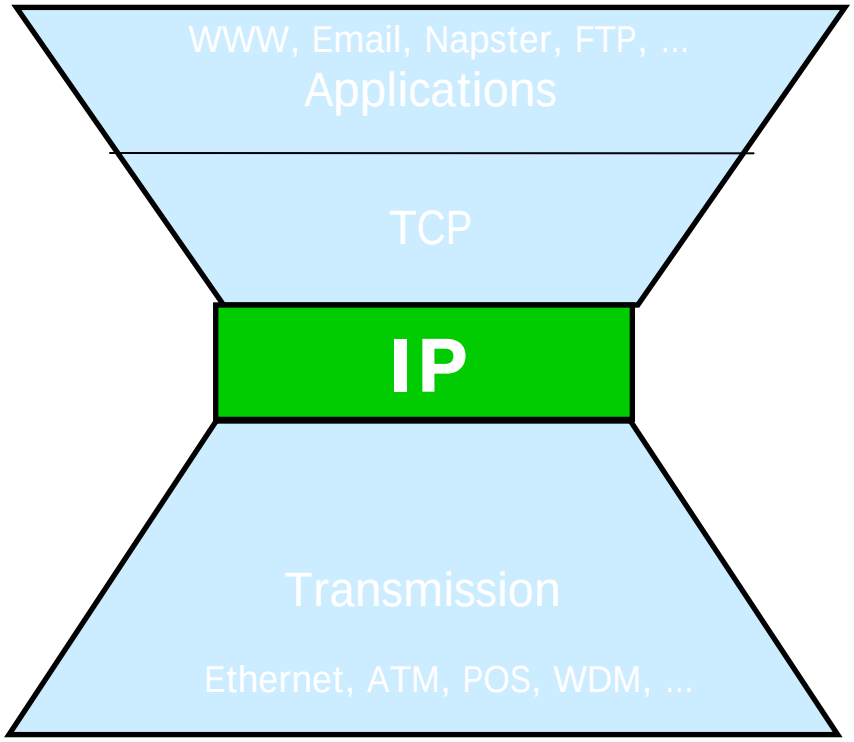
Requirements:

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

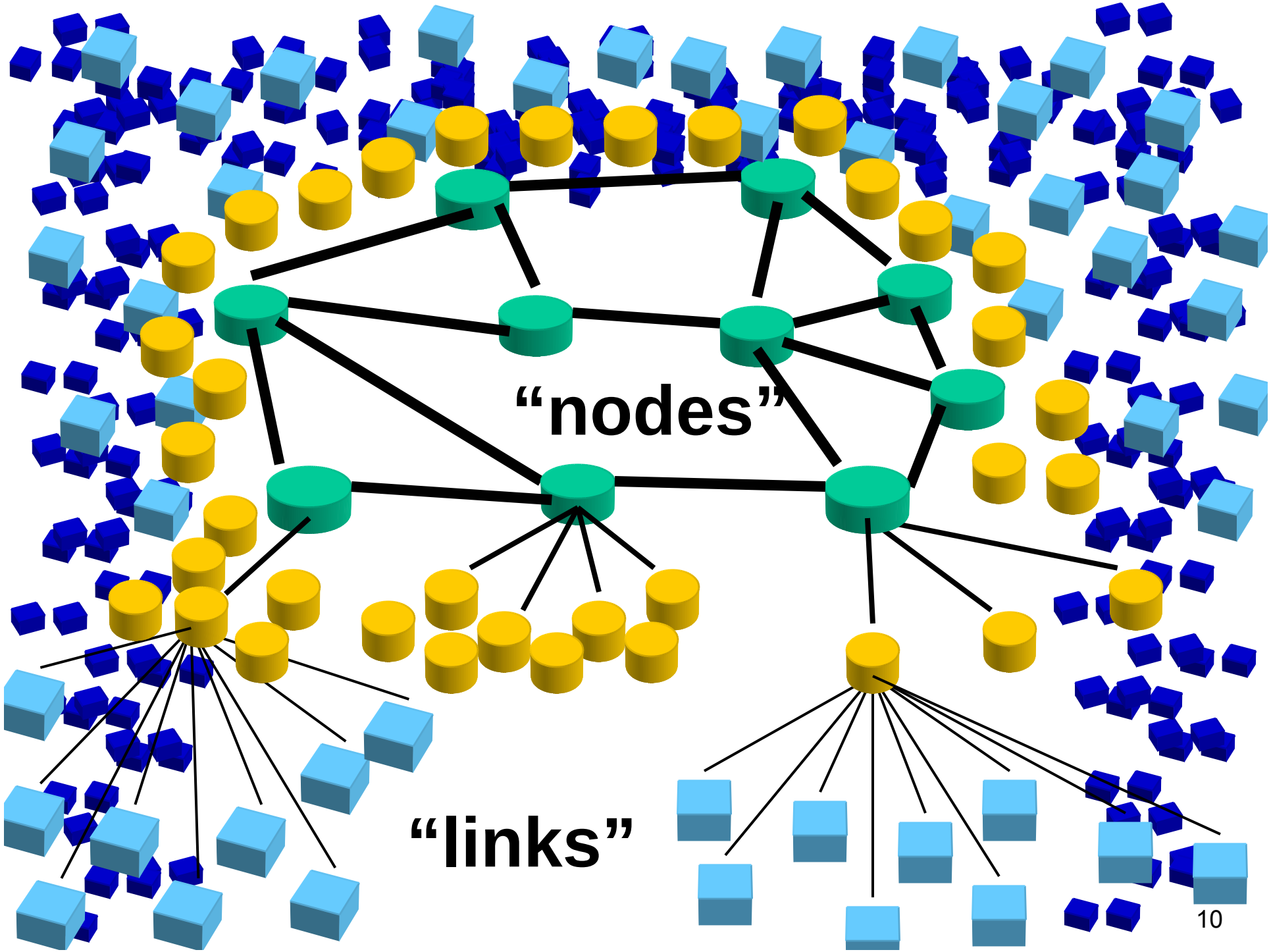
Internet Architecture: Horizontal Decomposition



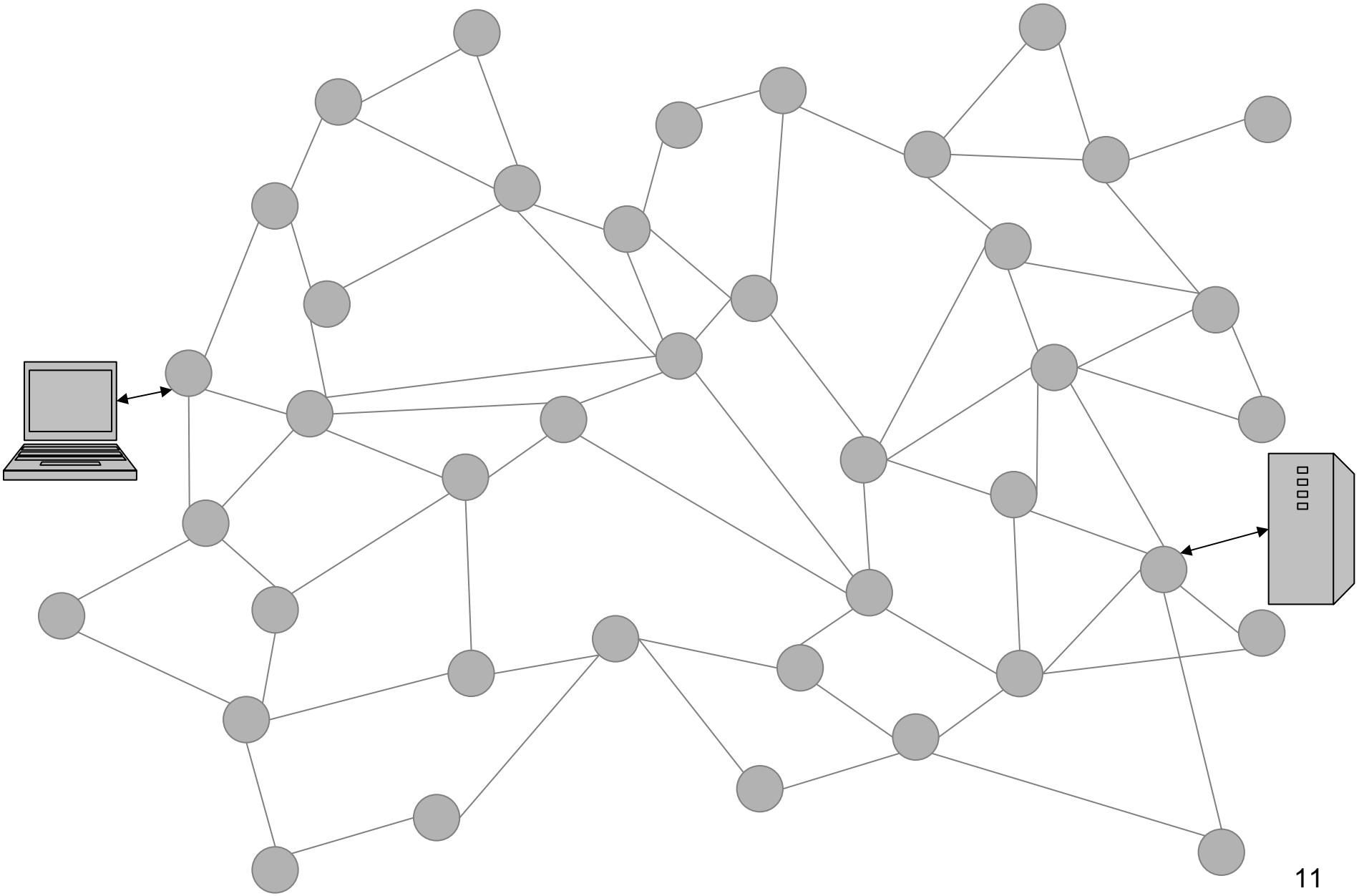
Internet Connectivity: The “hourglass” Architecture



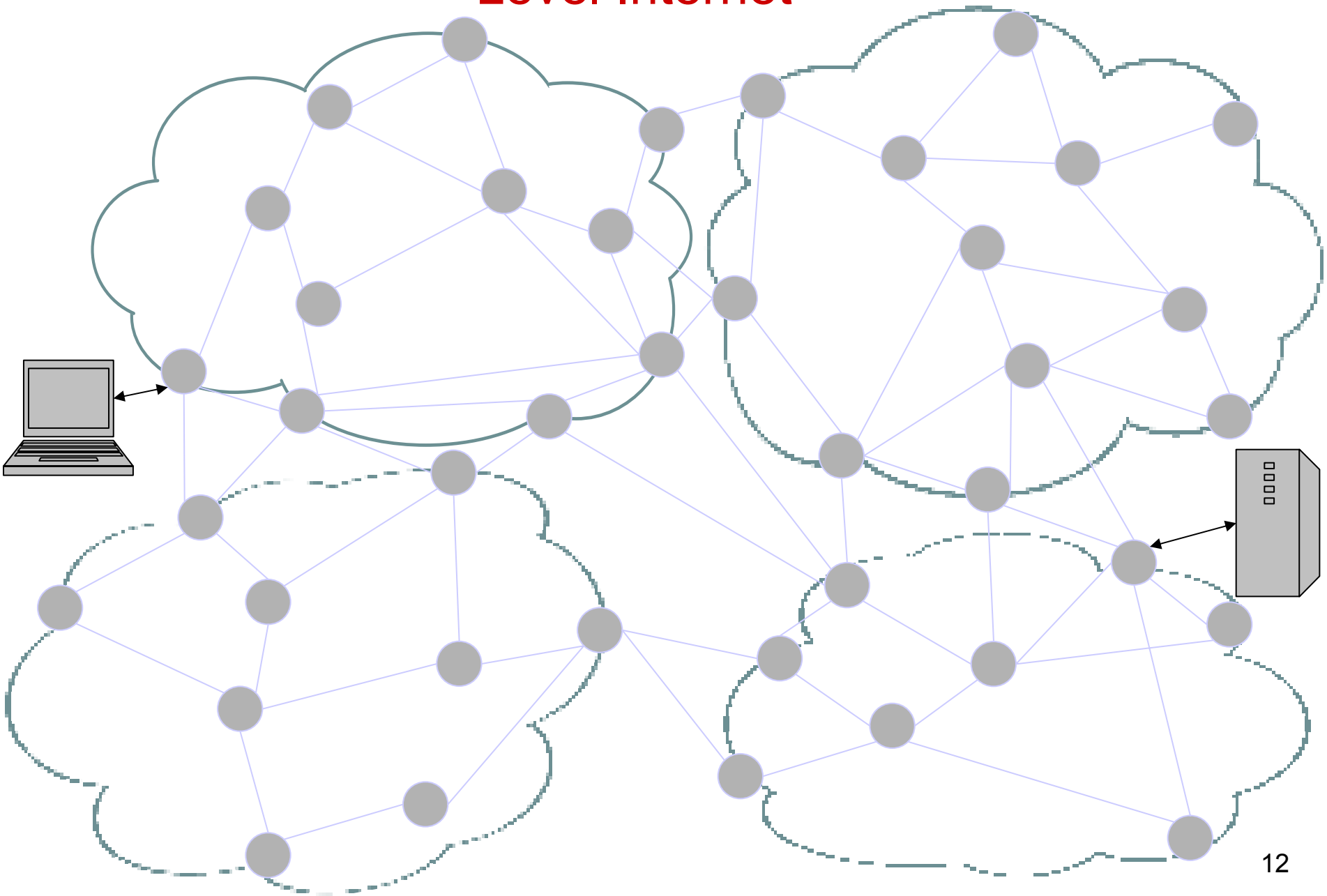
- Consider a (vertical) layer of the Internet hourglass
- Expand it horizontally
- Give layer-specific meaning to “nodes” and “links”


“nodes”
“links”

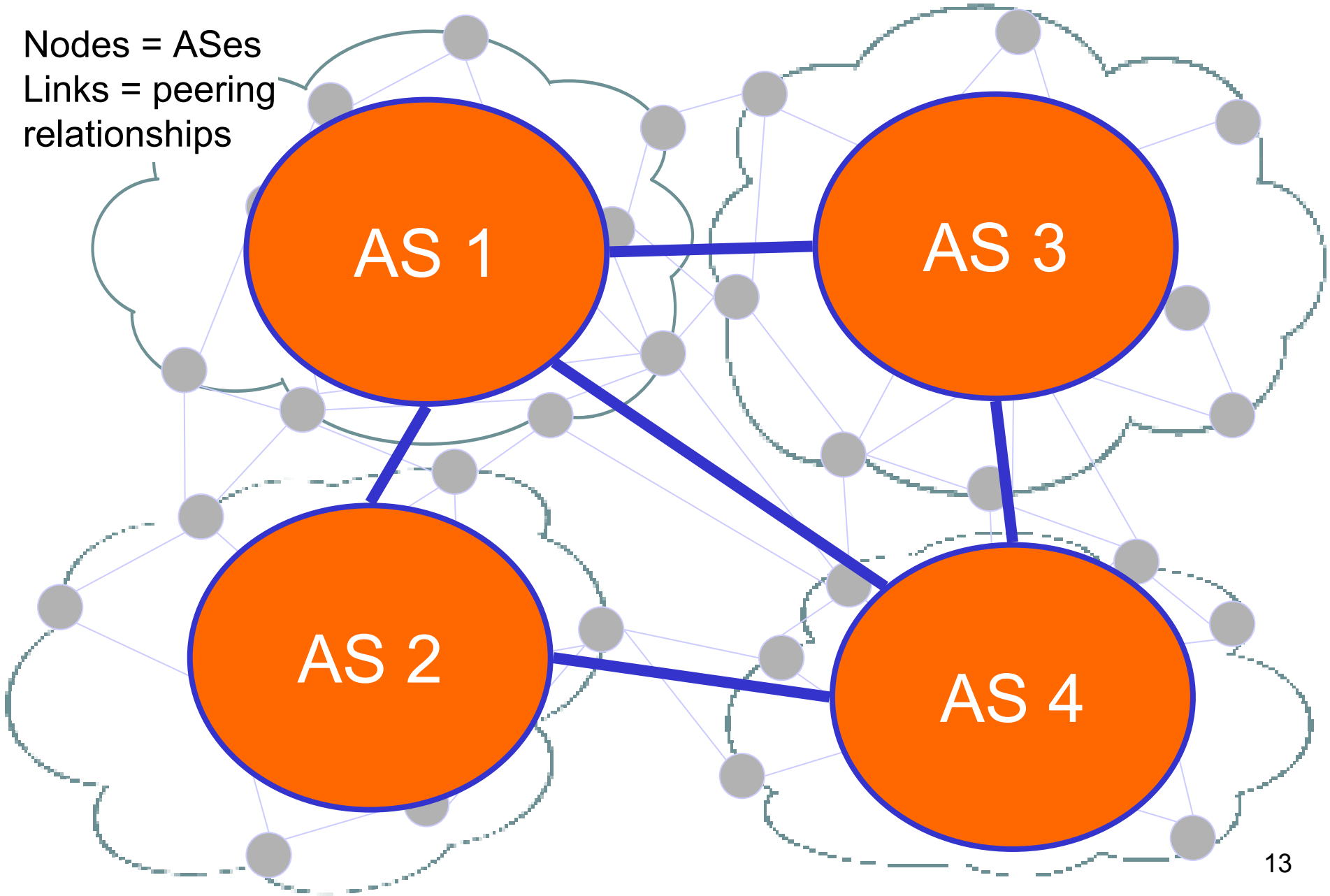
Router-Level Internet



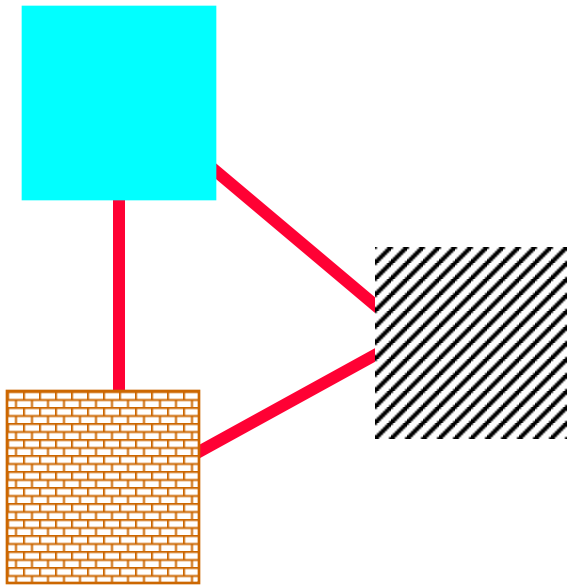
From Router-Level to Autonomous System (AS)-Level Internet



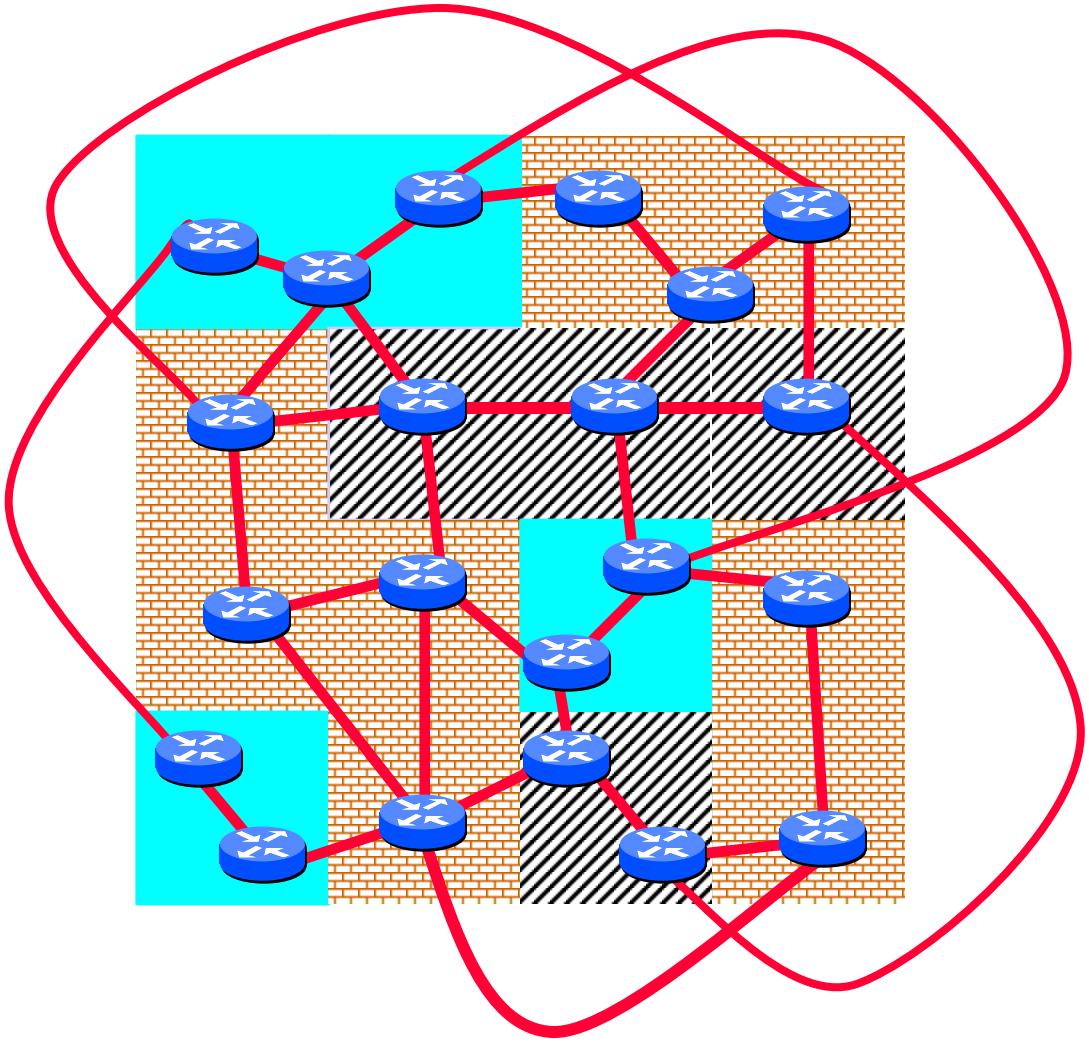
AS Graphs = Business Relationships



AS Graphs obscure Topology!



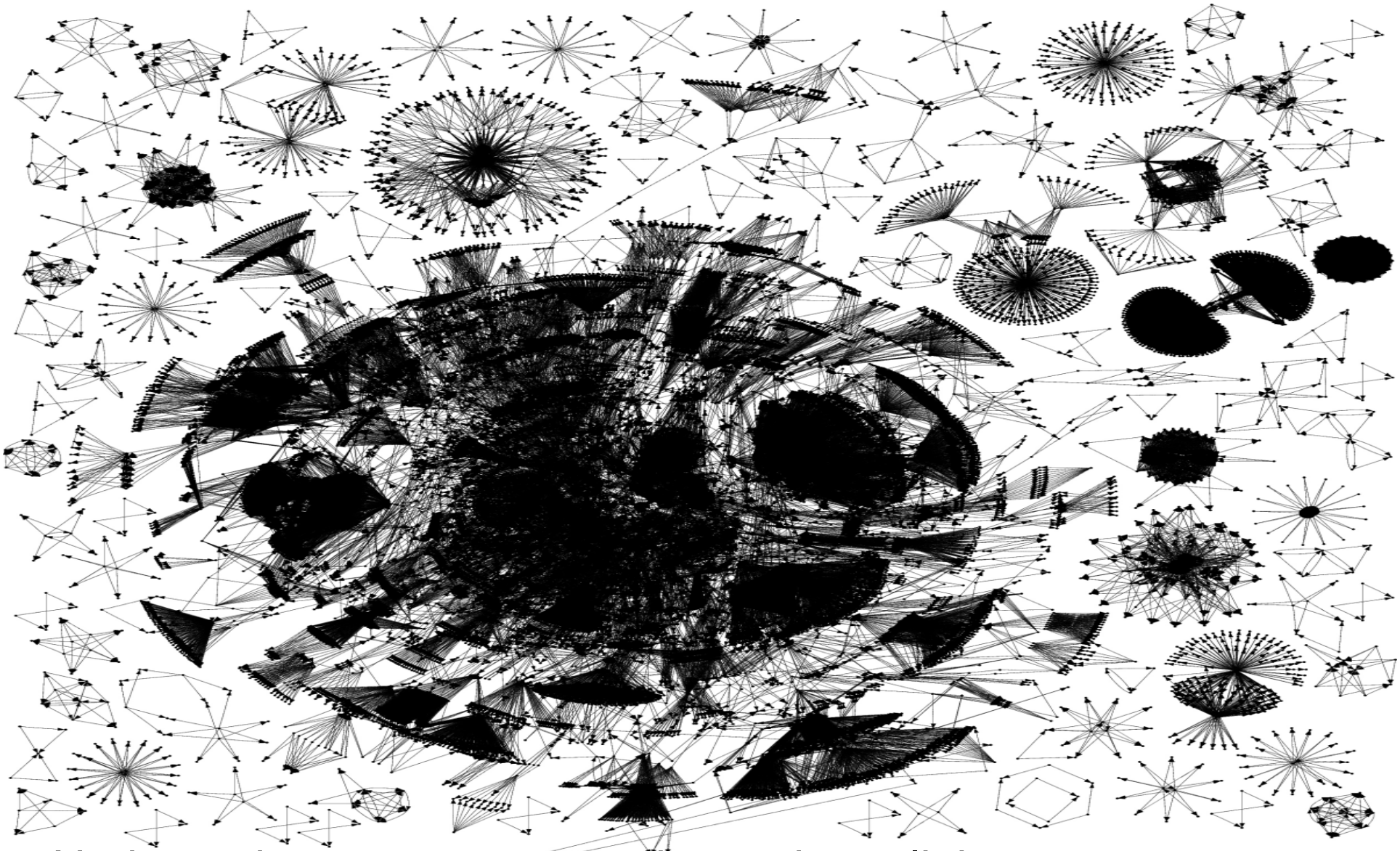
The AS graph may look like this.



Reality may be closer to this…

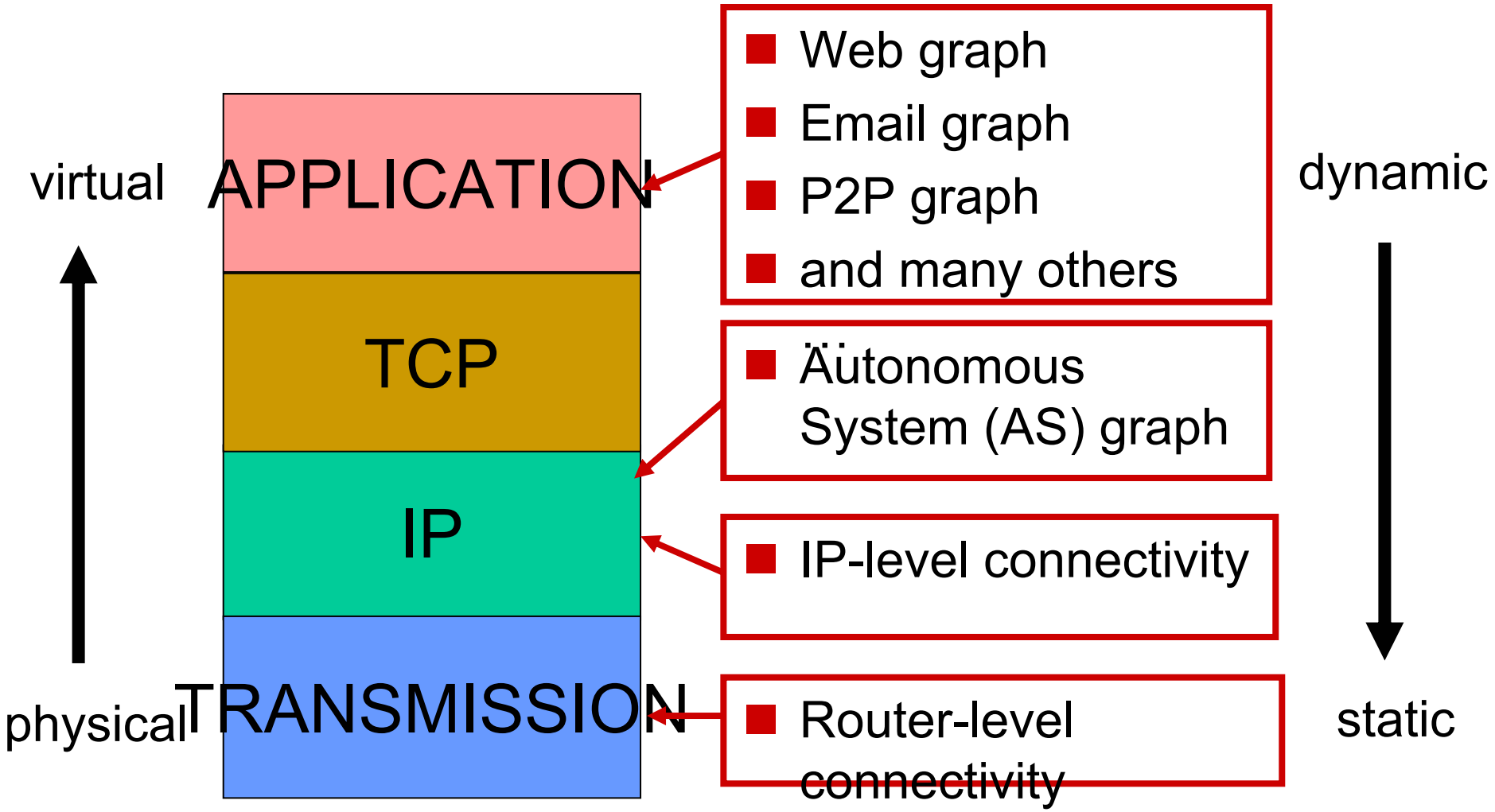
Courtesy Tim Griffin

(Part of the) Web Graph



Nodes = documents, connections = hyperlinks

The many Facets of Internet Topology



MESSAGE #1: Specify WHICH aspect of Internet topology

- There is no "generic" Internet topology
- The many facets of Internet topology
  - Router-level (physical)
  - IP-level (logical)
  - AS-level (logical)
  - Application-level (logical)
  - …
- Details of each make a big difference
- Lack of specificity can cause 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??

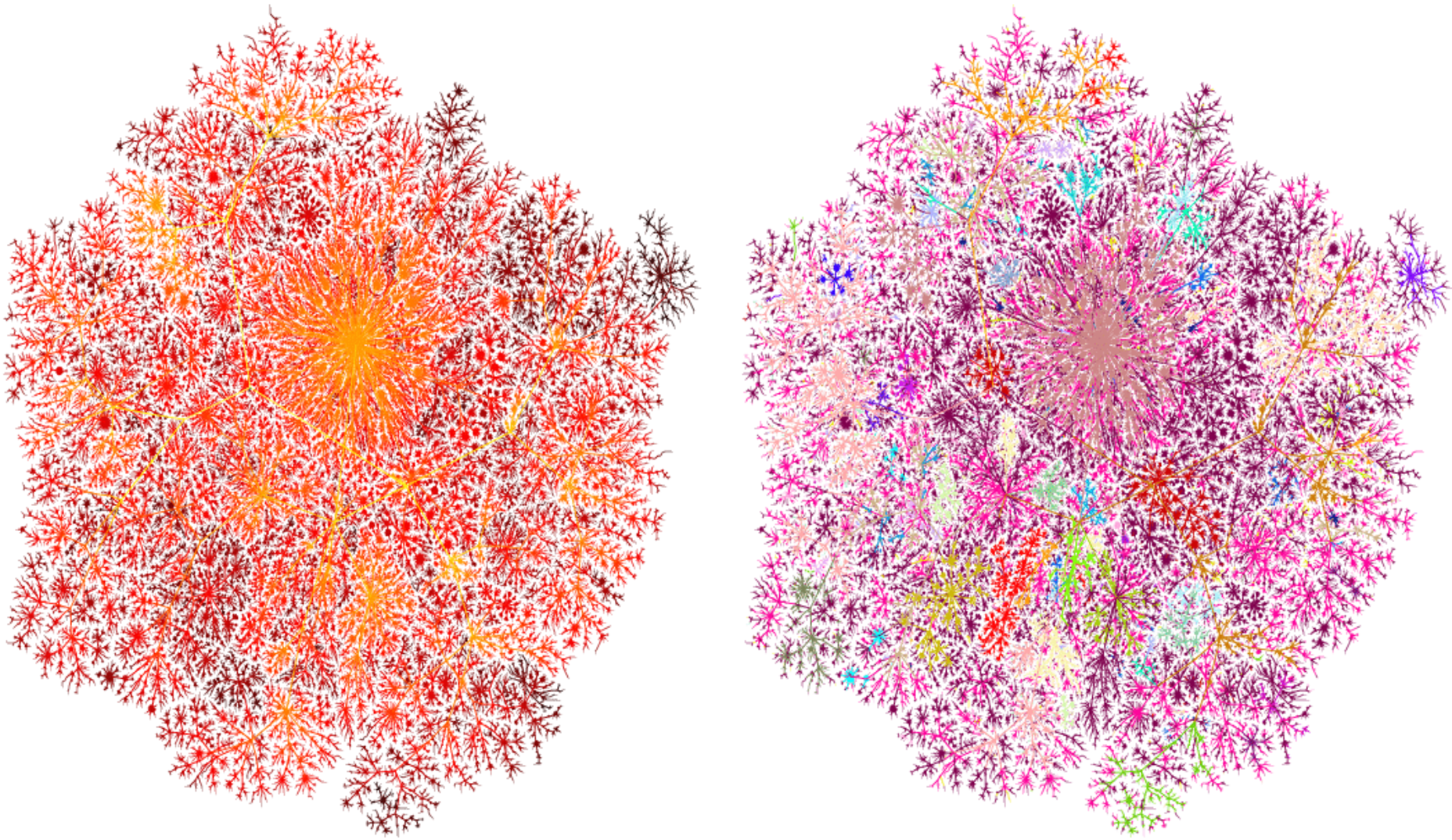
On Measuring Internet Connectivity

- No central agency/repository
- Economic incentive for ISPs to obscure network structure
- Direct inspection is typically not possible
- Based on measurement experiments, hacks
- Mismatch between what we want to measure and can measure

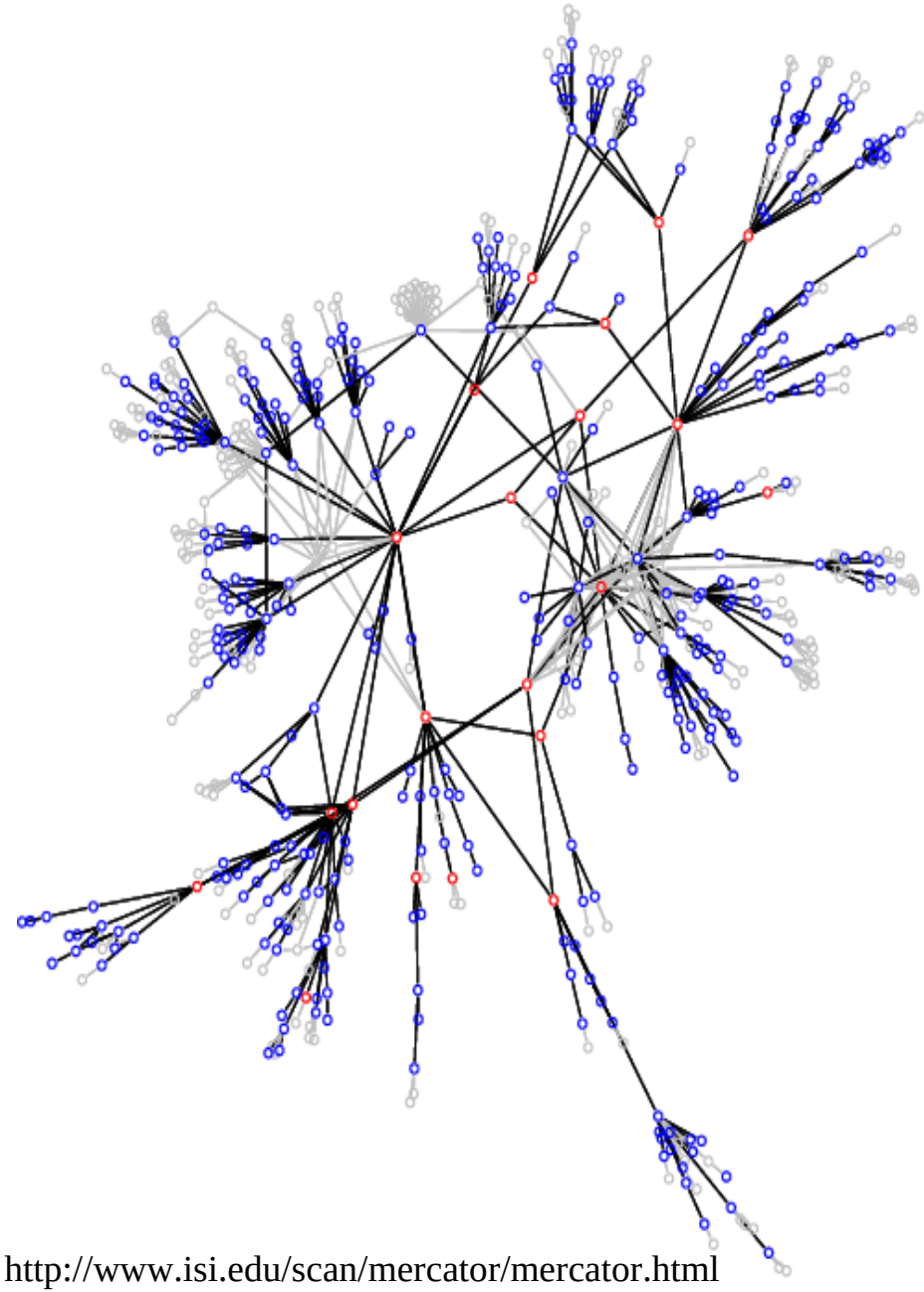
On Measuring the Internet's Router-level Topology

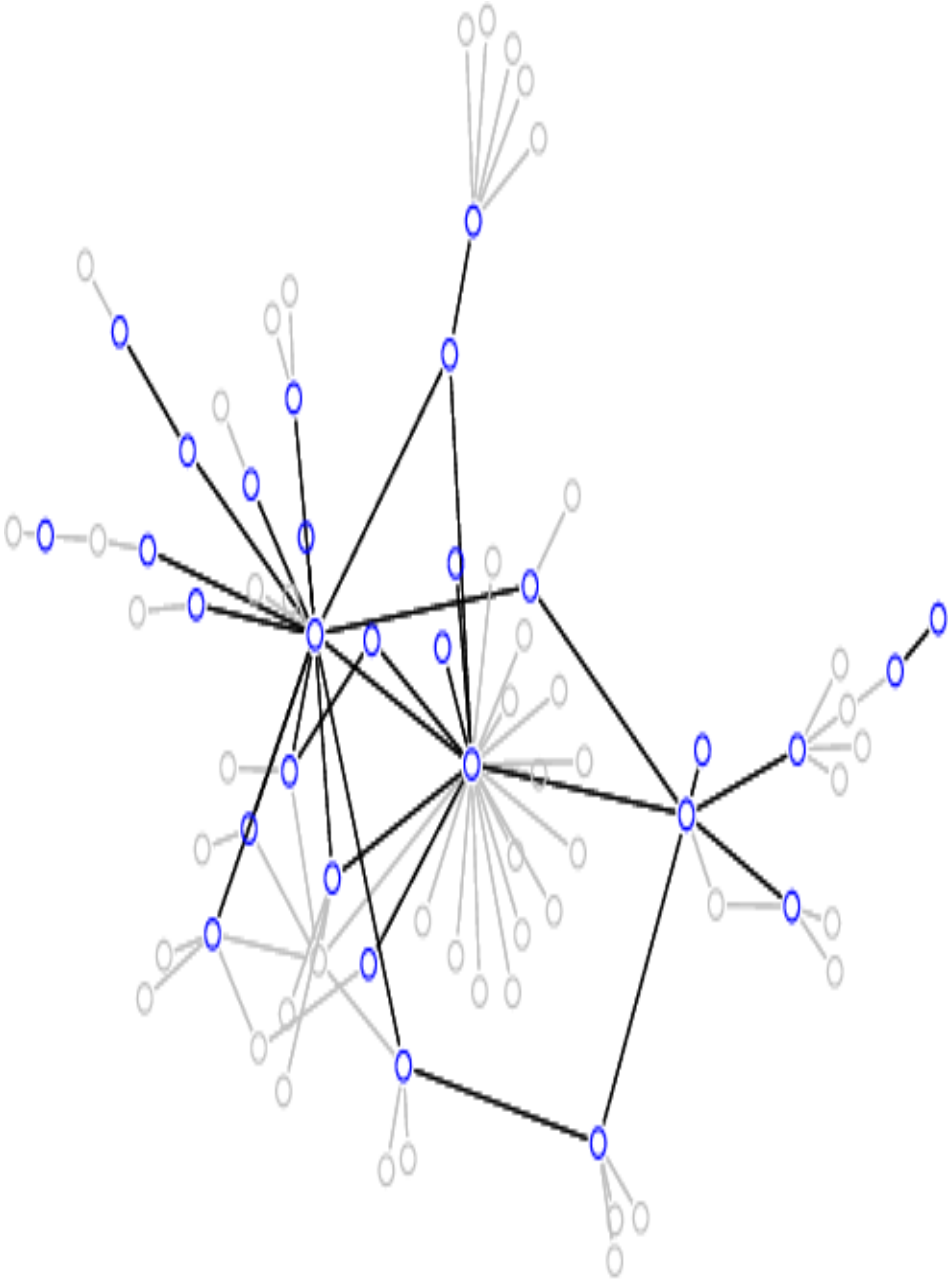
- **`traceroute`** tool
  - Discovers compliant (i.e., IP) routers along path between selected network host computers

- 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)
  - Dimes (EU project)

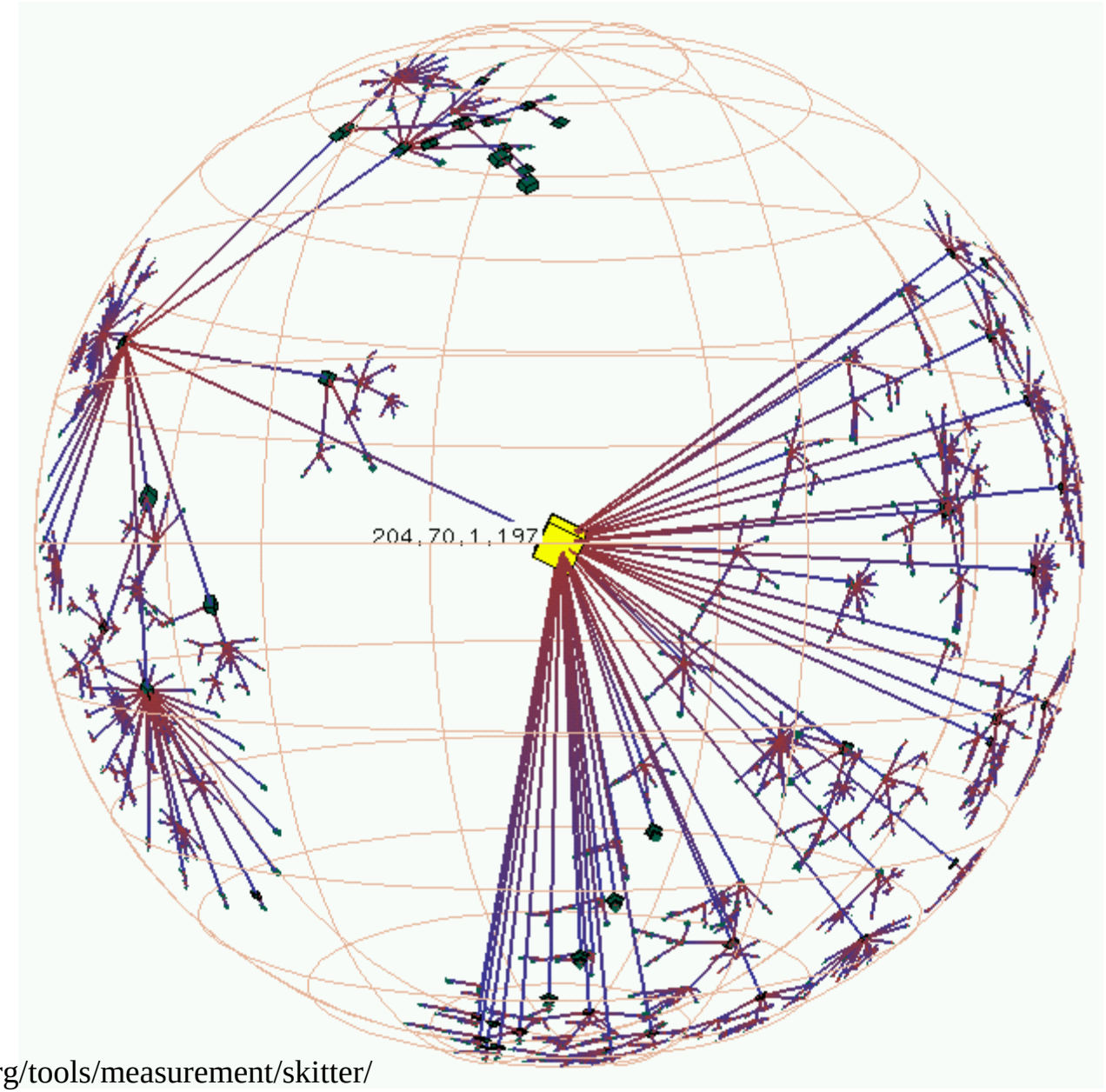


http://research.lumeta.com/ches/map/

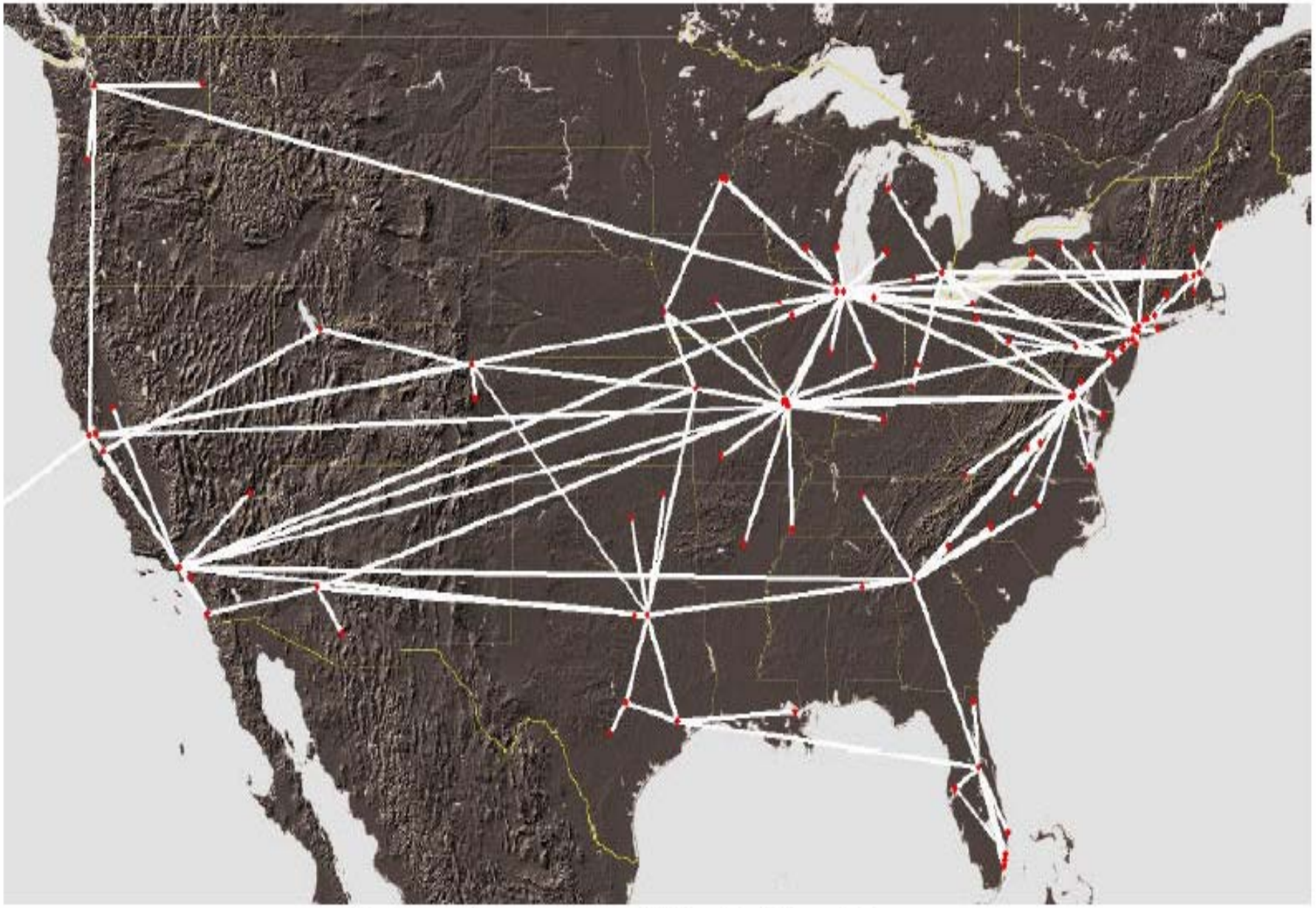




http://www.isi.edu/scan/mercator/mercator.html

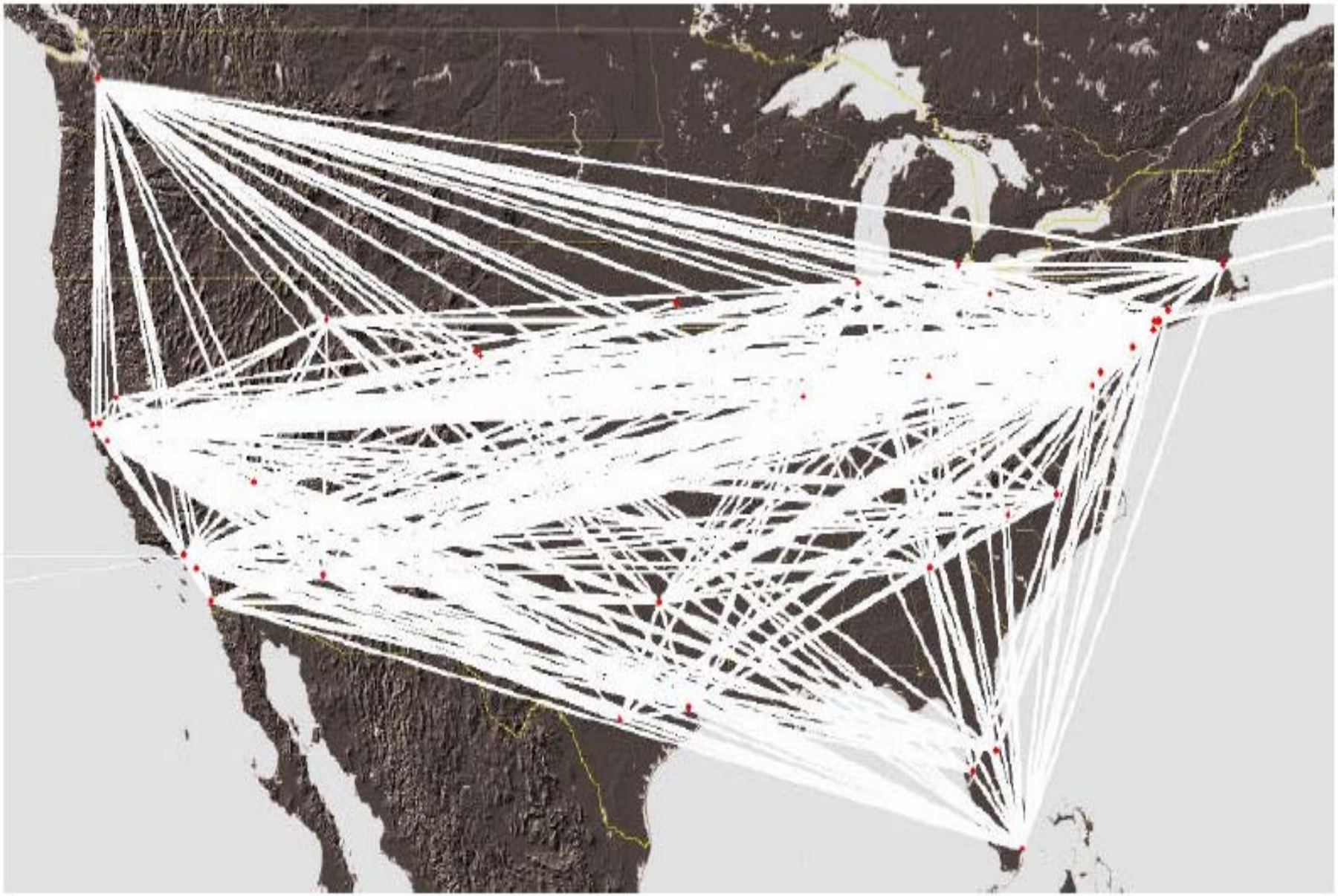


http://www.caida.org/tools/measurement/skitter/



Background image courtesy JHU, applied physics labs

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

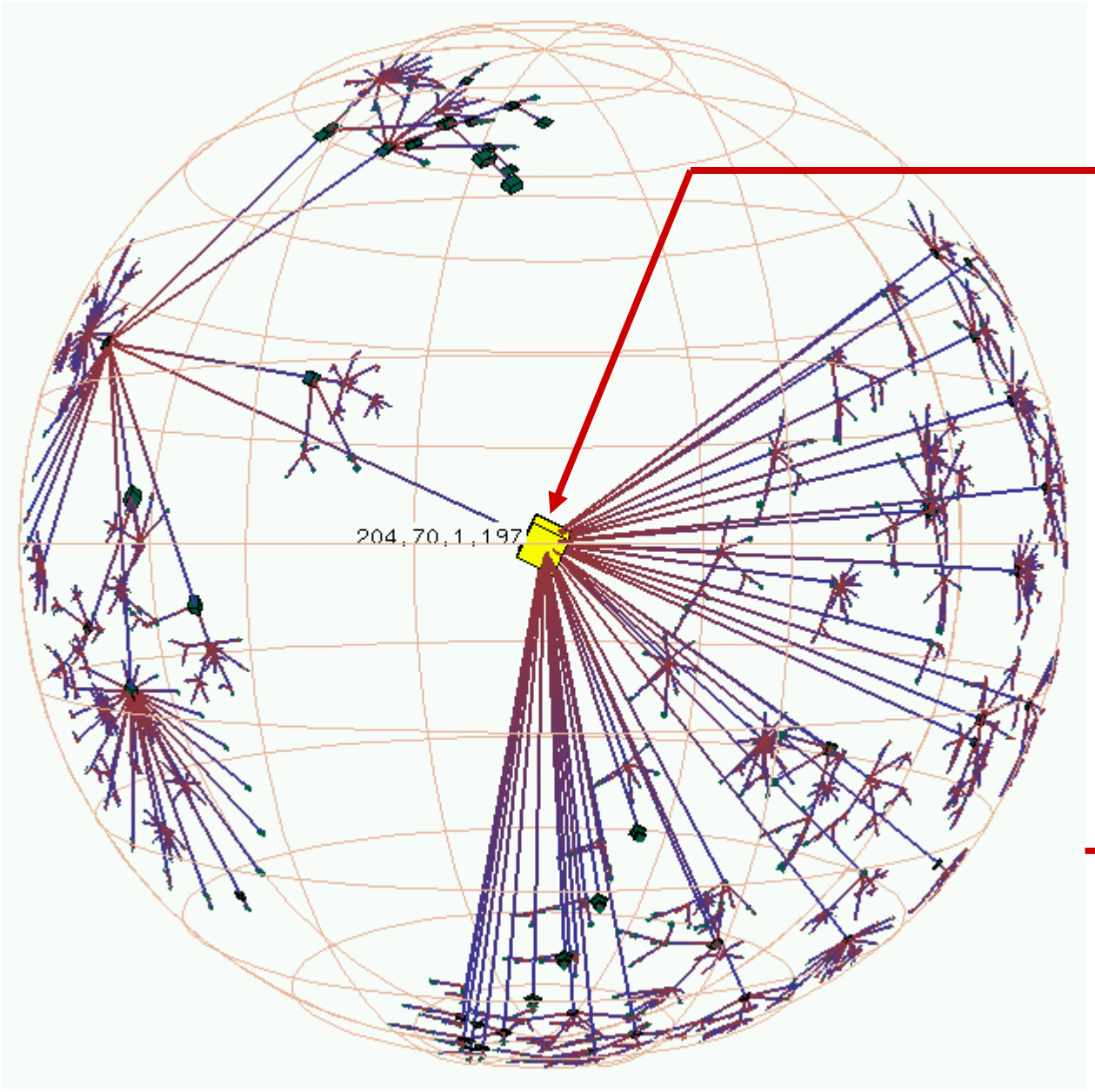


Background image courtesy JHU, applied physics labs

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

HOWEVER: Problems with existing measurements

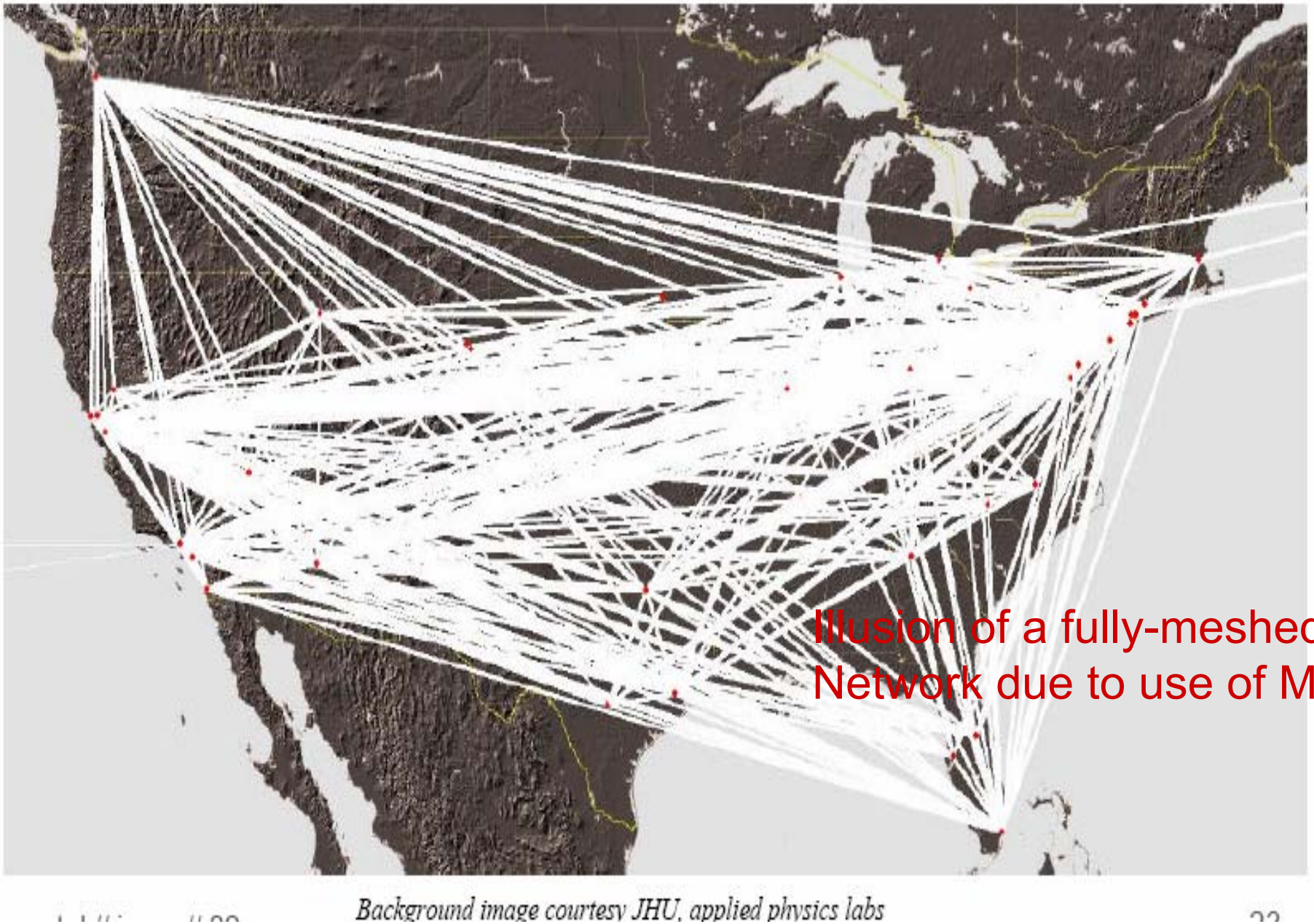
- traceroute-based measurements are ambiguous
  - traceroute is strictly about IP-level connectivity
  - traceroute cannot distinguish between high connectivity nodes that are for real and that are fake and due to underlying Layer 2 (e.g., Ethernet, ATM) or Layer 2.5 technologies (e.g., MPLS)



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

This “node” is an entire network! (not just a router)

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



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

HOWEVER: Problems with existing measurements

- 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

On Measuring the Internet's AS-level Topology

- BGP routing tables/updates
  - RouteViews (Univ. of Oregon)
  - RIPE (Europe)
  - E.g., 129.223.224.0/19 7018 701 4637 1221
- Traceroute measurements
  - Skitter (CAIDA)
  - Dimes
- Other available sources
  - Public databases (WHOIS)
  - Looking glass sites, additional routing tables

HOWEVER: Problems with existing measurements

- BGP-based measurements are incomplete
  - Contains most nodes (ASes)
  - Might miss up to 40-50% of existing links
- BGP-based measurements are ambiguous
  - Dynamics of AS-level Internet
  - Requires some guesswork in deciding whether a "new" node or link is genuine
- BGP-based measurements are inaccurate
  - Use of heuristics for inferring peering relationships

On Measuring the Internet's Overlay Topologies

- P2P networks
  - Structured (e.g., Kad DHT): Central control
  - Unstructured (e.g., Gnutella): Crawler
  - Sampling
- World-Wide-Web (WWW)
  - AltaVista crawls (Broder et al,) in 1999
  - Duration is a couple of weeks

HOWEVER: Problems with existing measurements

- High degree of dynamics of overlay networks
  - Connectivity structure changes underneath the crawler
  - Fast vs. slow crawls
- Enormous size of overlay networks
  - Complete crawls take too long
  - Alternative approach: Sampling
- Issues of sampling bias
  - Due to temporal dynamics of nodes (peers)
  - Due to spatial features of overlay topology

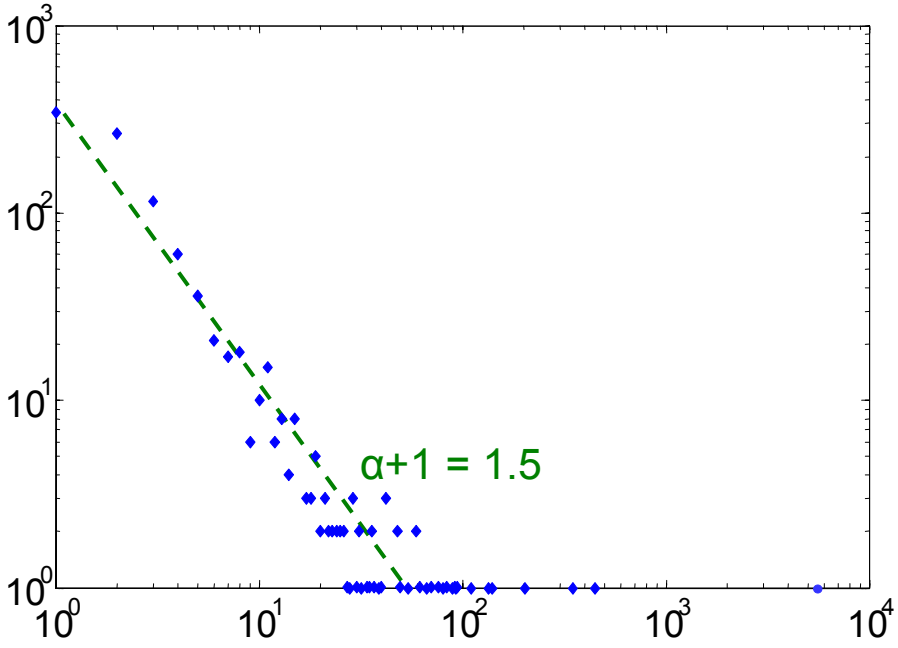
MESSAGE #2: Internet connectivity measurements should never be taken at face value

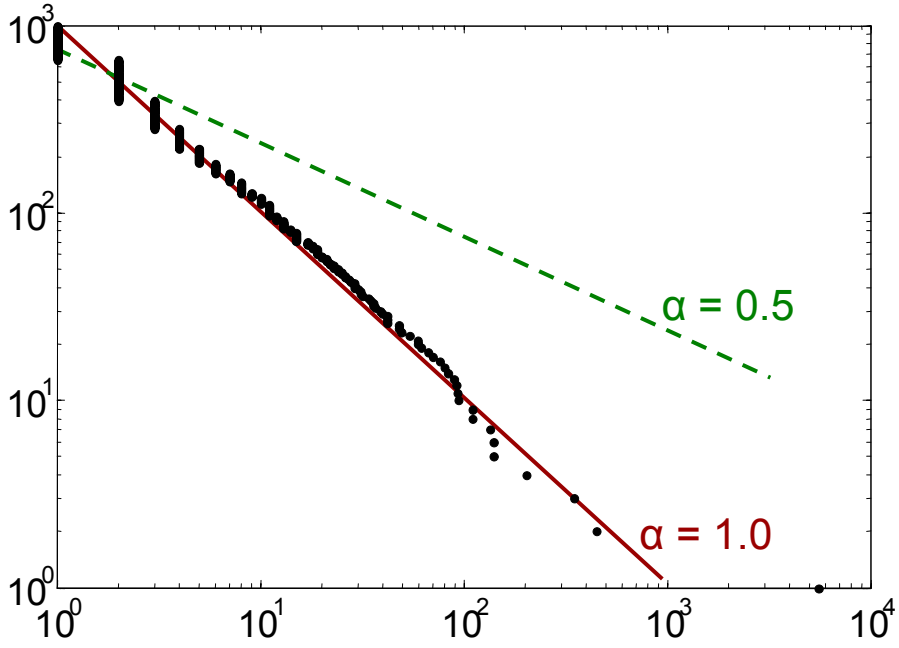
- Each technique is typically specific to the network of interest (e.g., traceroute for IP-level, BGP tables for AS-level)
- It is important to understand the process by which the measurements were obtained and collected
- Even best-of-breed measurement data is ambiguous, inaccurate, and incomplete
- Taking (someone else's) data at face value may provide a false basis for results

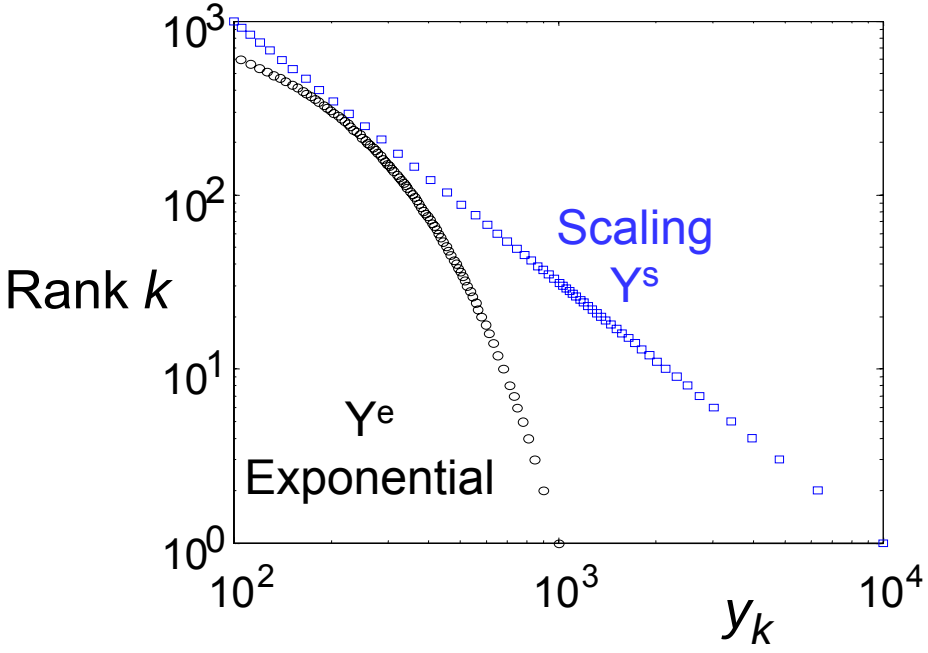
On Inferring Internet Connectivity

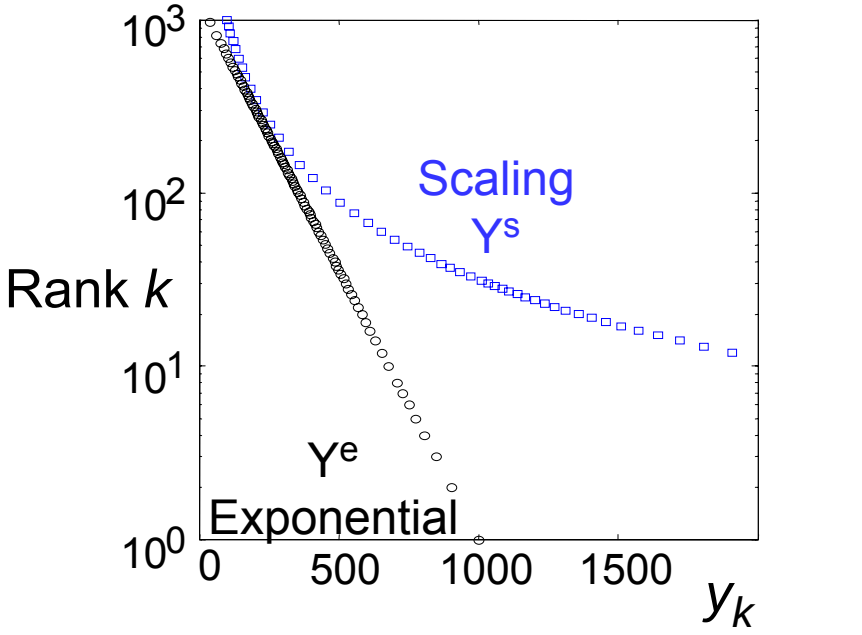
- Quality of available data
  - See earlier
- Quality of data analysis
  - Doing specious analysis with specious data
- Sensitivity of inferred properties to known imperfections of the underlying data
  - See later

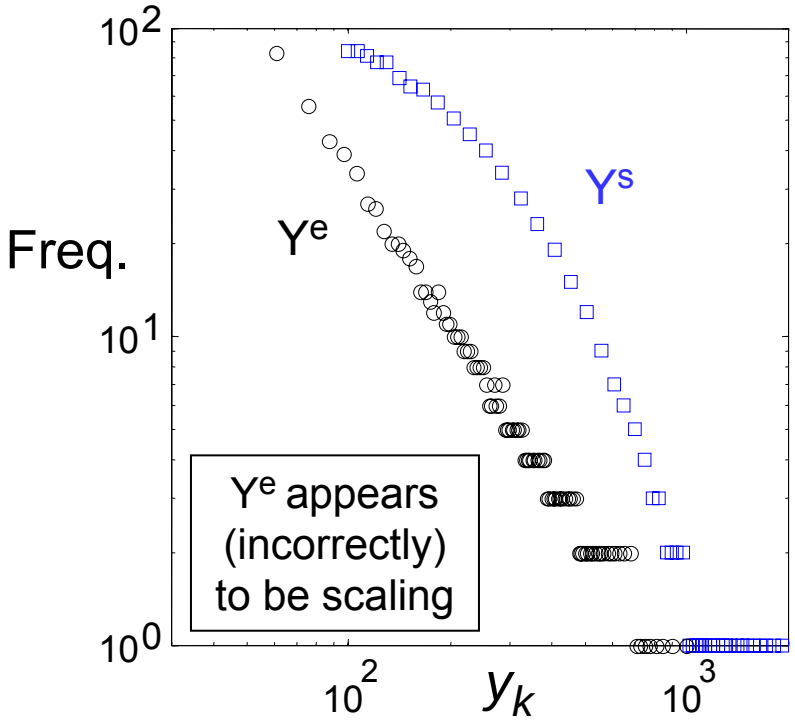
Size-Frequency vs. Size-Rank Plots
or
Non-cumulative vs. Cumulative

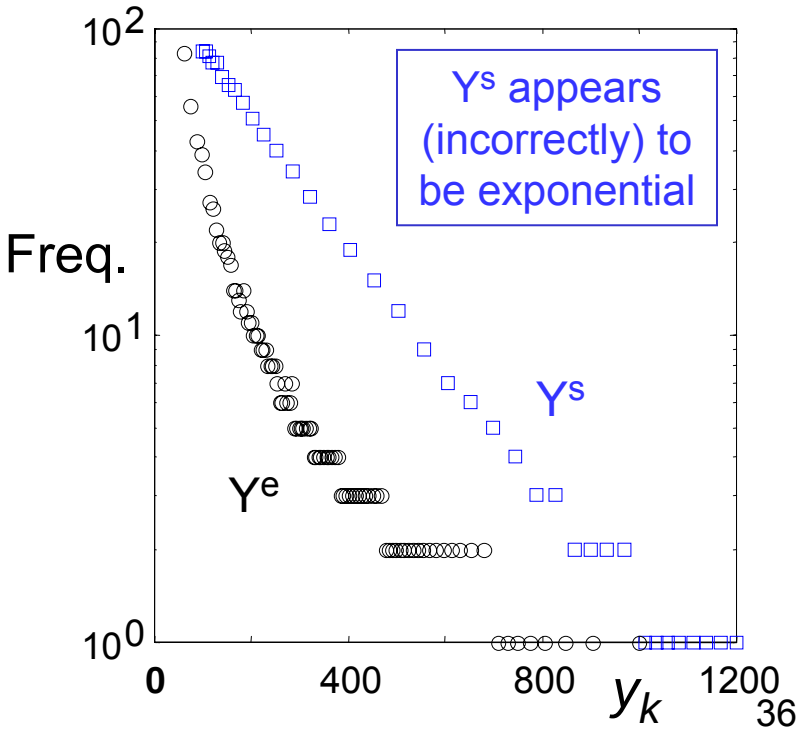


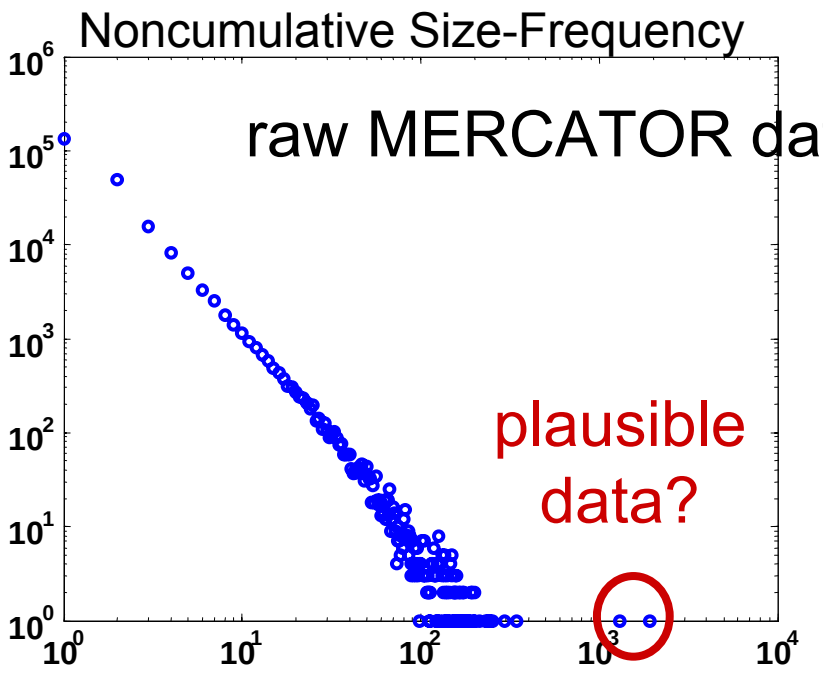


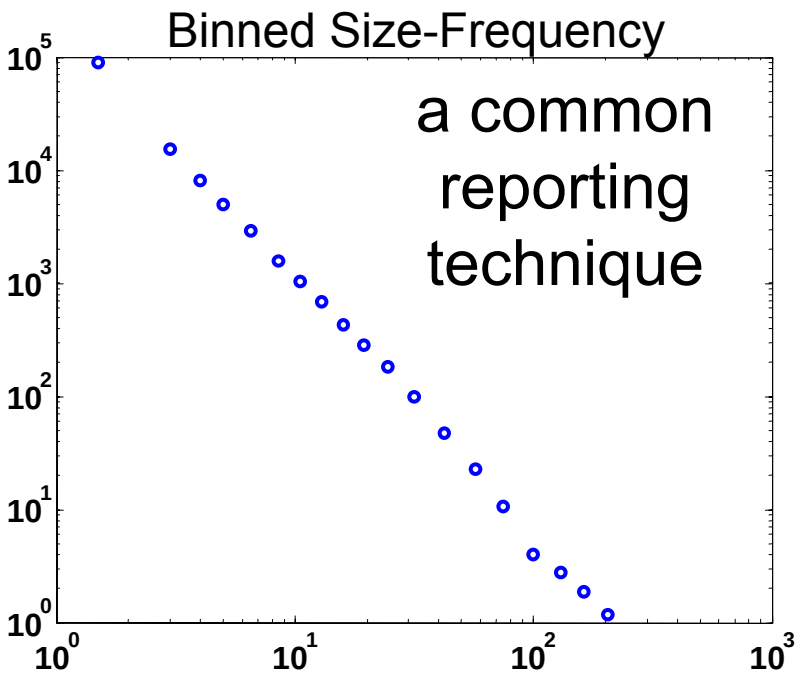

Rank k
10^3
10^2
10^1
10^0
Scaling
Y^s
Y^e
Exponential
10^2
10^3
10^4
y_k

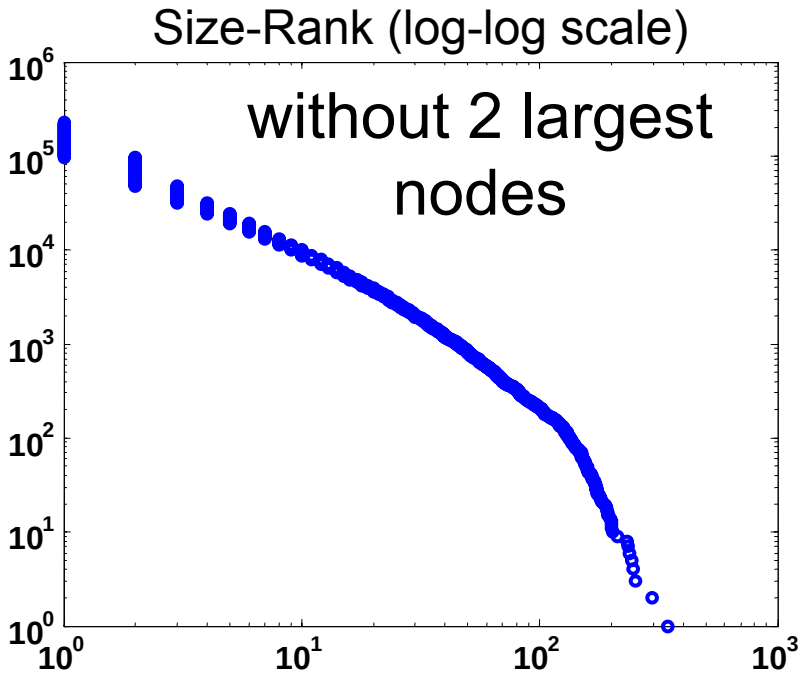

Rank k
10^3
10^2
10^1
10^0
Scaling
Y^s
Y^e
Exponential
0
500
1000
1500
y_k

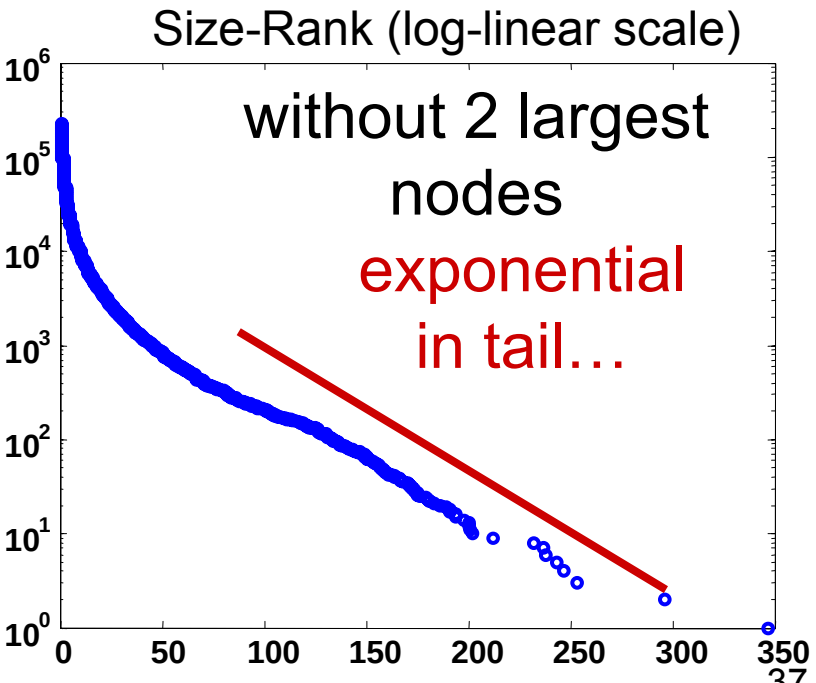

Freq.
10^2
10^1
10^0
Y^s
Y^e
Y^e appears (incorrectly) to be scaling
10^2
y_k
10^3


Freq.
10^2
10^1
10^0
Y^s appears (incorrectly) to be exponential
Y^s
Y^e
0
400
800
1200
y_k


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…

MESSAGE #3: Be aware of specious analysis of specious measurements

- **Know your data**
- **Avoid (non-cumulative) size-frequency plots**
- **Rely on (cumulative) size-rank plots**

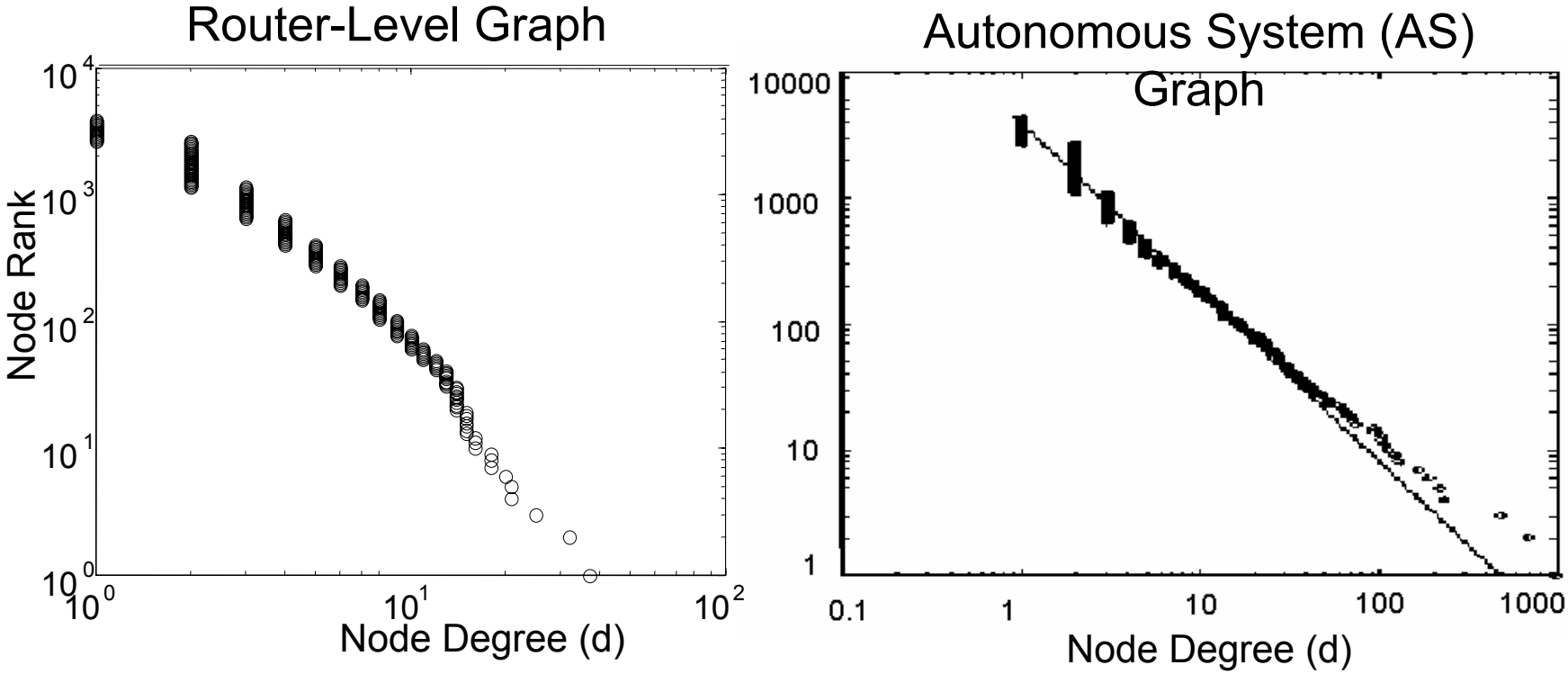
On Modeling Internet Connectivity and Model Validation

- All models are wrong …
- There are in general many different explanations/models for one and the same phenomenon
- Role of randomness vs. design
- To argue in favor of any particular model typically requires additional information
  - In the form of domain knowledge
  - In the form of new or complementary data
- Reproducing a given graph statistics is a data-fitting exercise and does not validate a chosen model

Internet Topology and Power Laws

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

Source: Faloutsos et al. (1999)



Node Degree: d = # connections

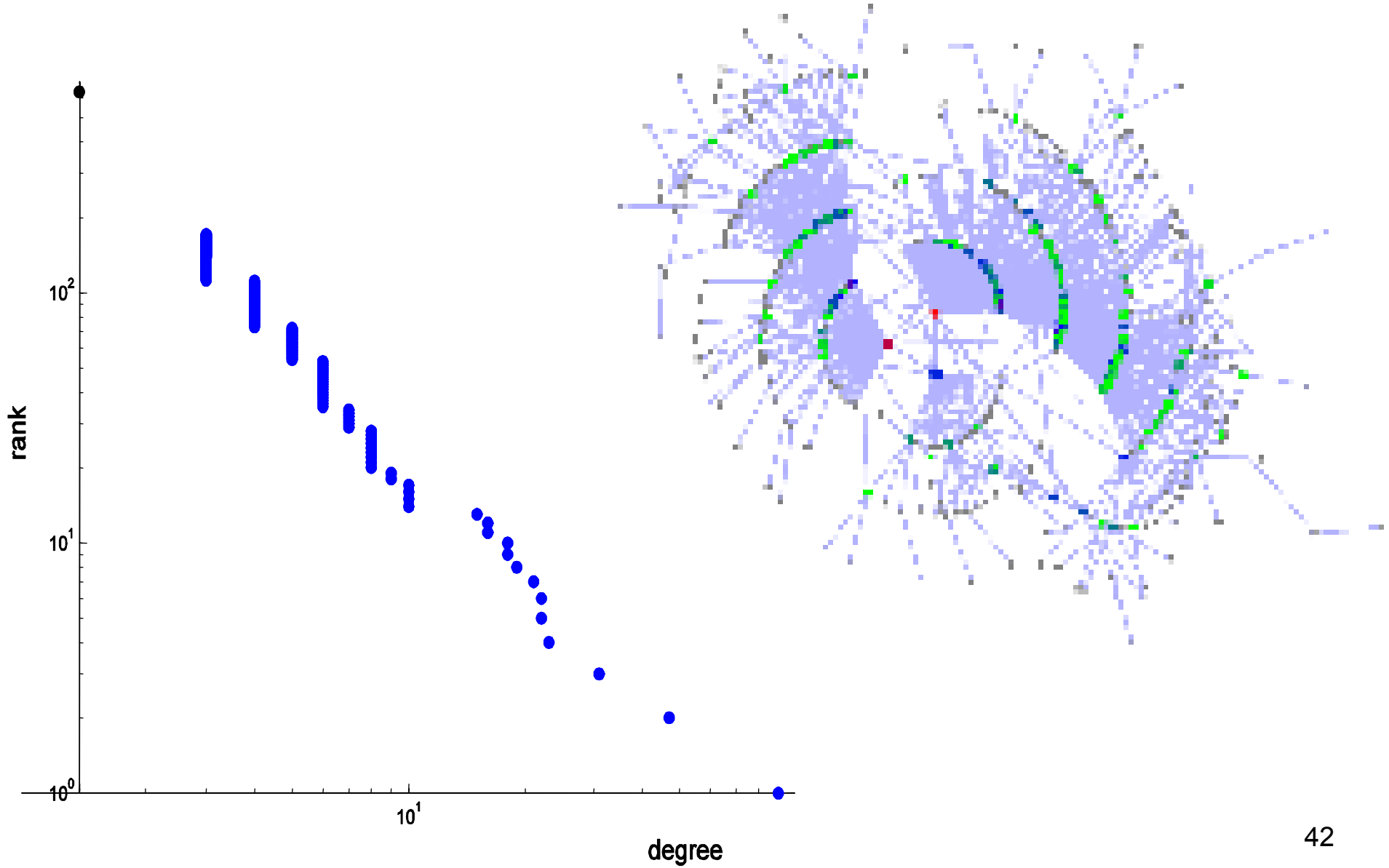
- A random variable X is said to follow a *power law* with *index* $\alpha > 0$ if $P[X > x] \approx cx^{-\alpha}, \text{ as } x \to \infty$

- Has led to active research in *degree-based* network models

Degree-Based (Random) Graph 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]**

Preferential Attachment

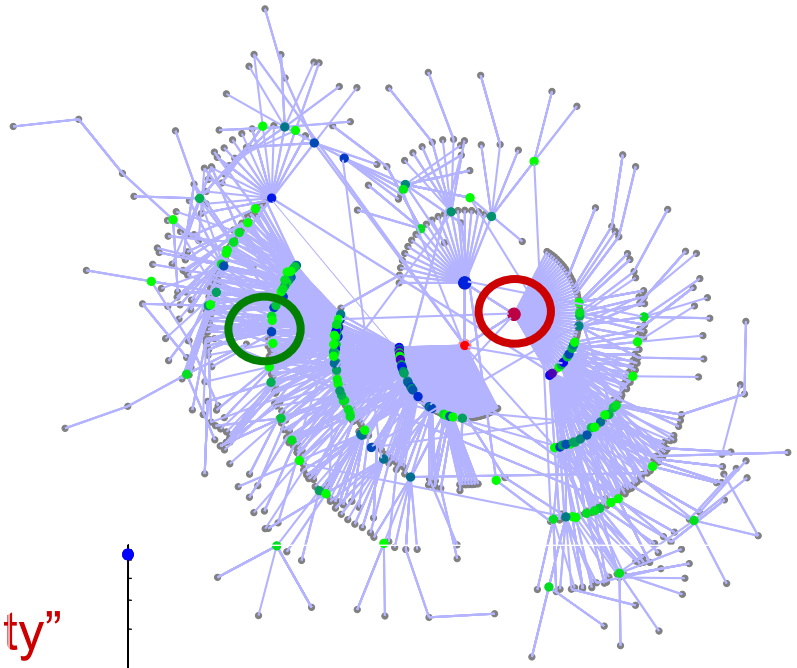


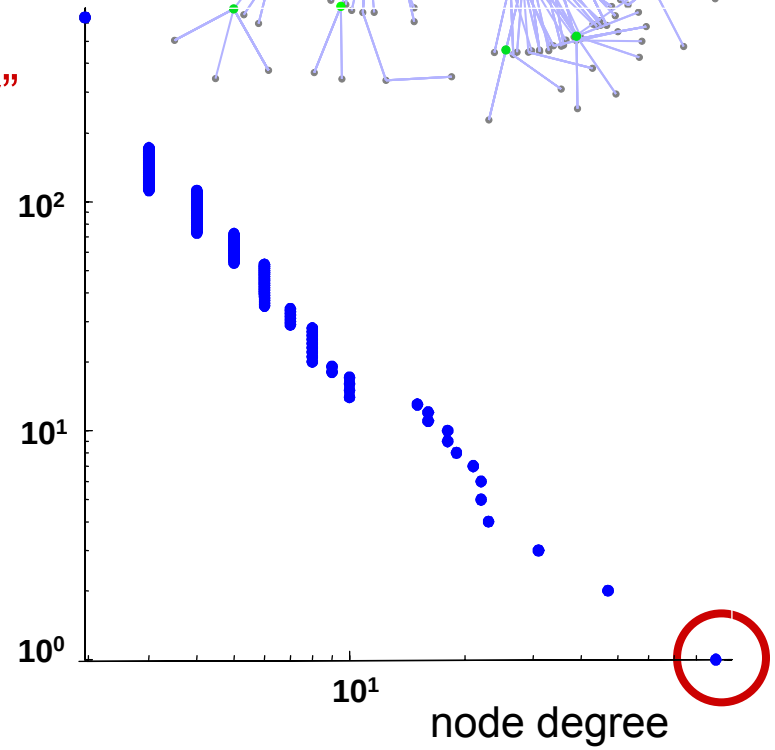
"Scale-free" networks and the "Achilles' heel" of the Internet

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





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

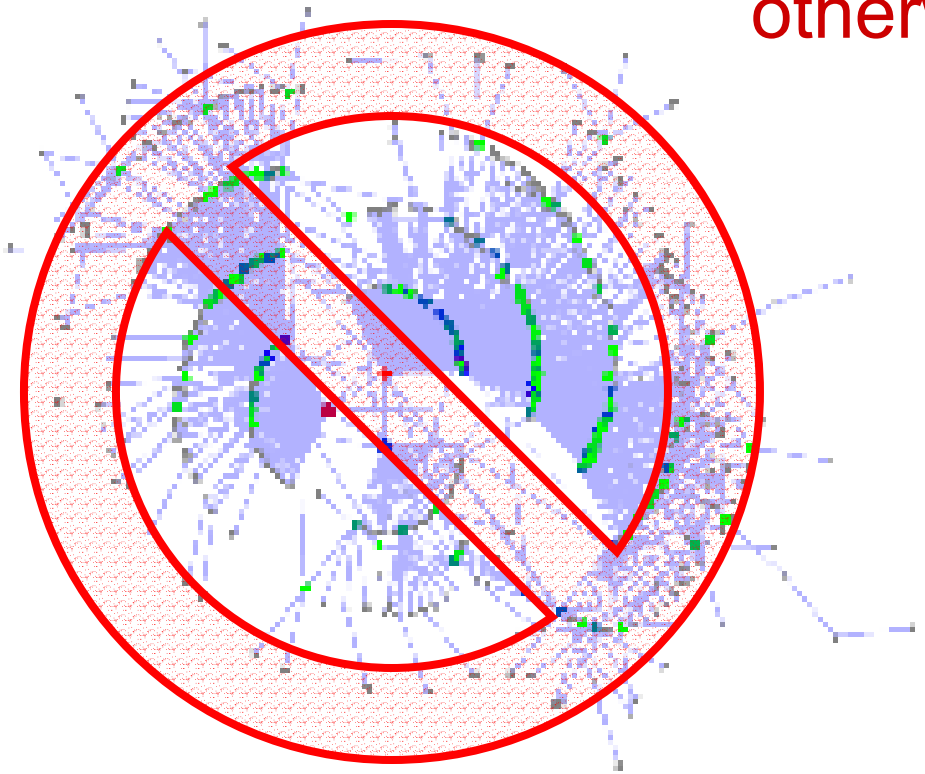
⇒ Yield a zero epidemic threshold for contagion

⇒ Are evidence of fundamental self-organization in networks

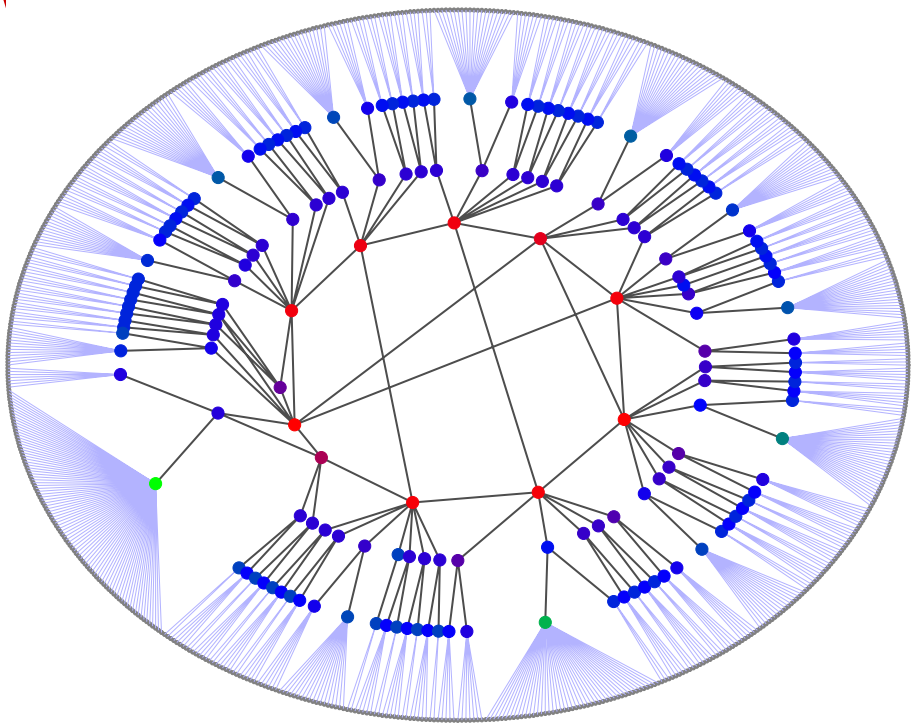
⇒ This self-organization is believed by some to be a *universal feature* of technological, biological, social and business networks

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

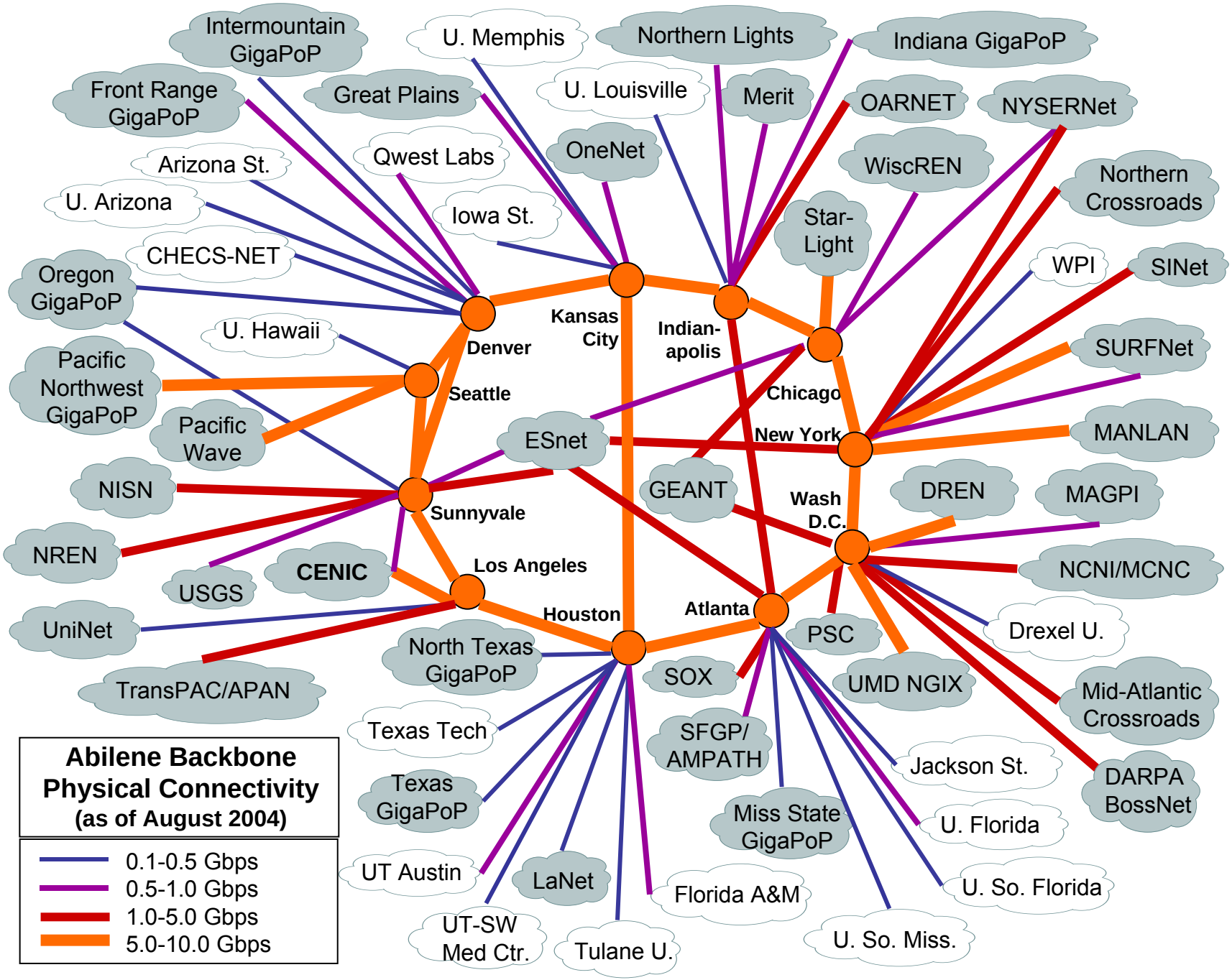
MESSAGE #3: Can construct networks that have the same node degree distribution but are OPPOSITES otherwise

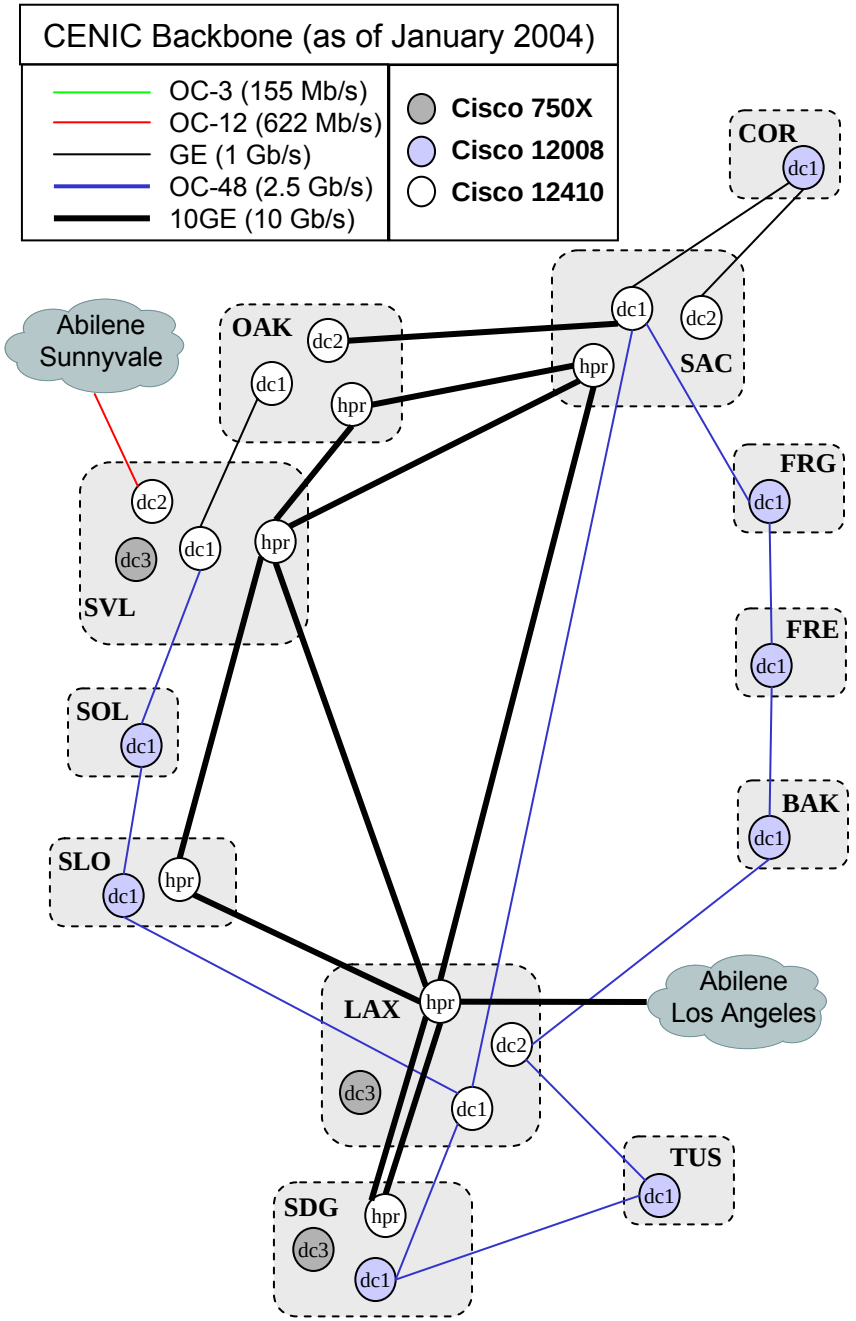


- High degree central “hubs”
- From random construction
- Poor performance and robustness



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


Intermountain GigaPoP
U. Memphis
Northern Lights
Indiana GigaPoP
Front Range GigaPoP
Great Plains
U. Louisville
Merit
OARNET
NYSERNet
Qwest Labs
OneNet
WiscREN
Arizona St.
U. Arizona
Iowa St.
Star-Light
Northern Crossroads
CHECS-NET
WPI
SINet
Oregon GigaPoP
Kansas City
Indian-apolis
U. Hawaii
Denver
SURFNet
Pacific Northwest GigaPoP
Seattle
Chicago
Pacific Wave
ESnet
New York
MANLAN
NISN
GEANT
DREN
MAGPI
Sunnyvale
Wash D.C.
NREN
CENIC
Los Angeles
NCNI/MCNC
USGS
UniNet
Houston
Atlanta
PSC
Drexel U.
North Texas GigaPoP
SOX
UMD NGIX
TransPAC/APAN
Mid-Atlantic Crossroads
Texas Tech
SFGP/ AMPATH
Jackson St.
Texas GigaPoP
DARPA BossNet
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

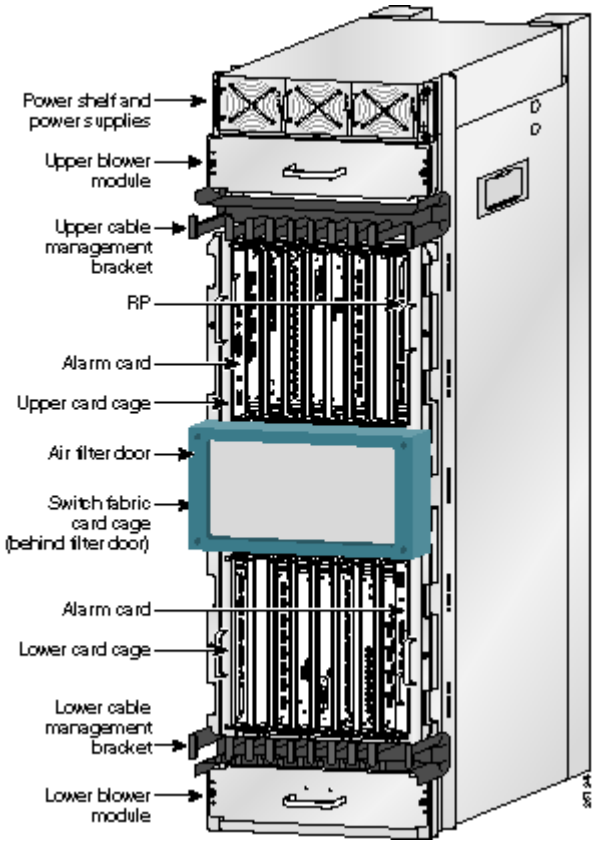
Like Abilene, its backbone is a sparsely-connected mesh, with relatively low connectivity and minimal redundancy.

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

Cisco 12000 Series Routers

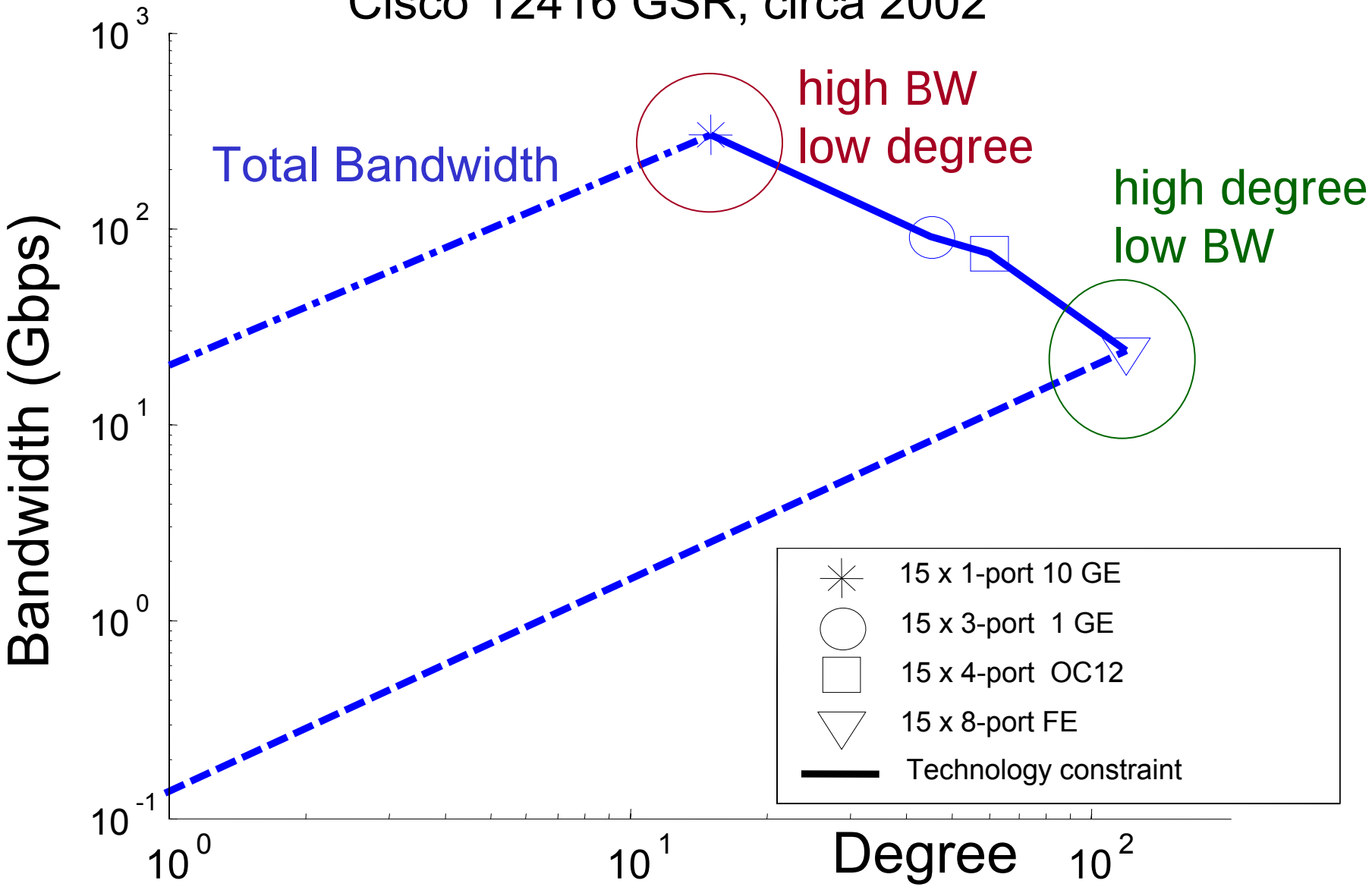
- Modular in design, creating flexibility in configuration.
- Router capacity is constrained by the number and speed of line cards inserted in each slot.

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

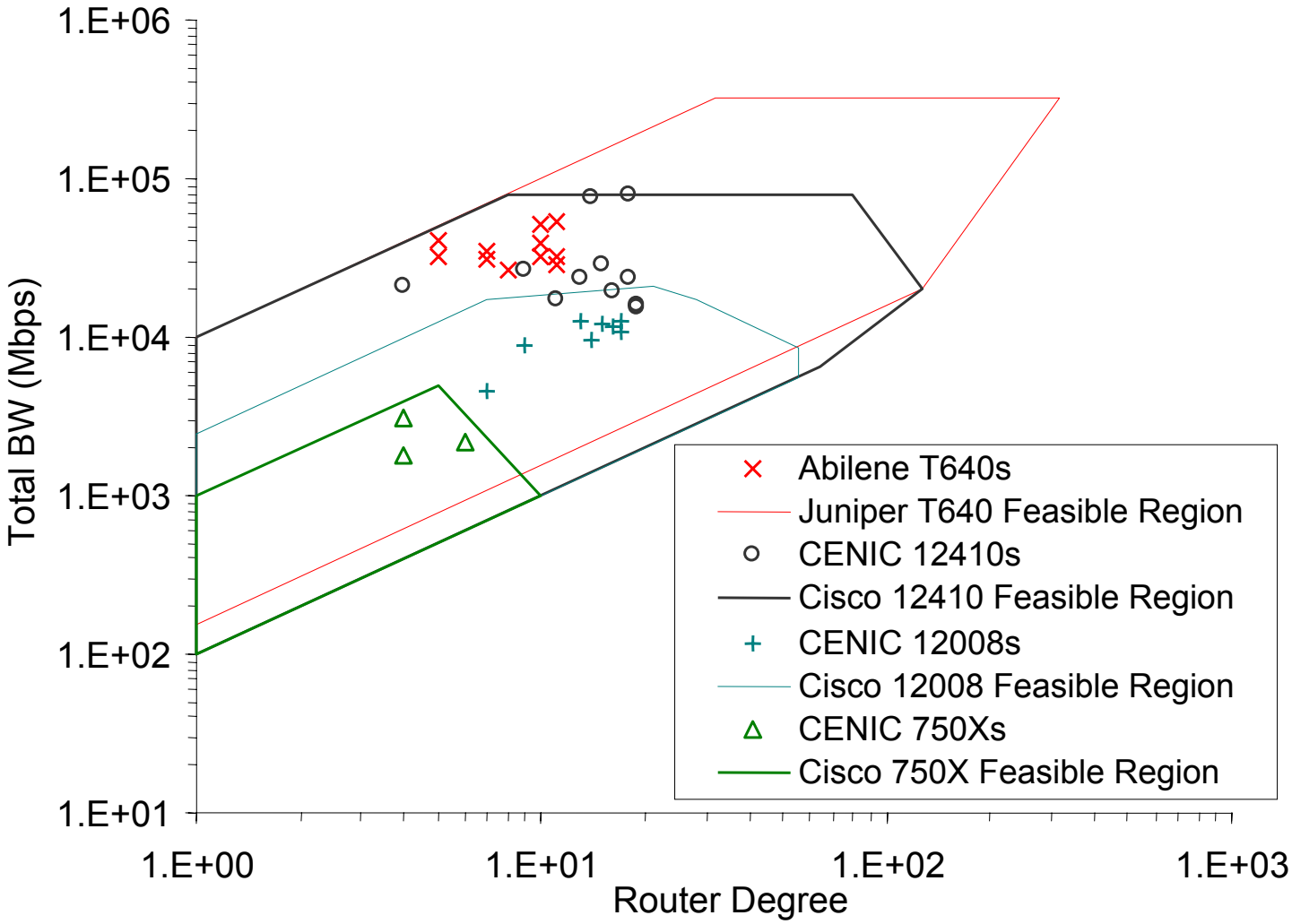


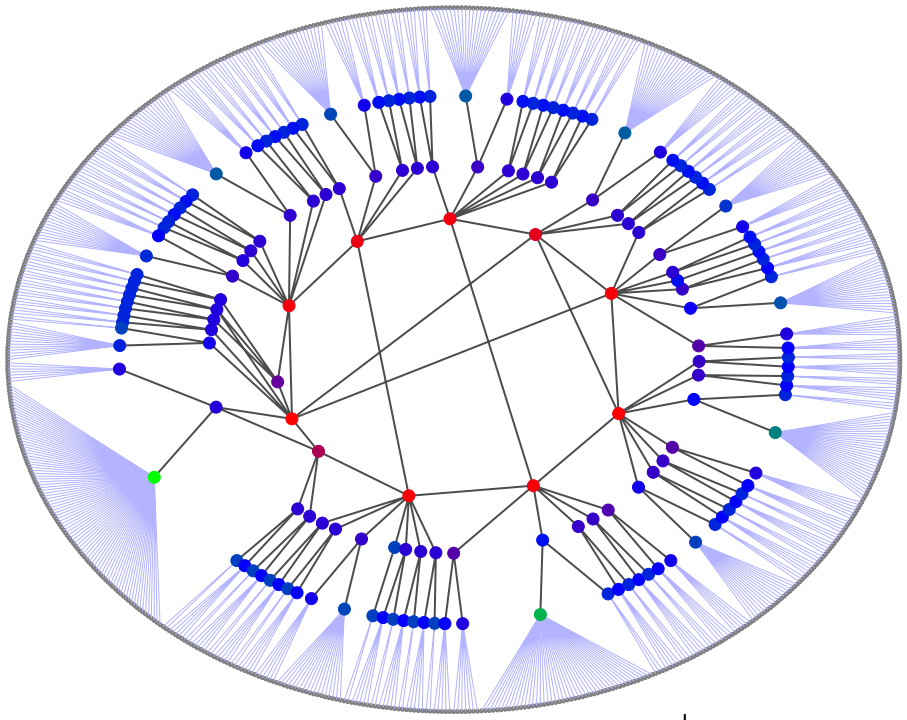
Source: www.cisco.com

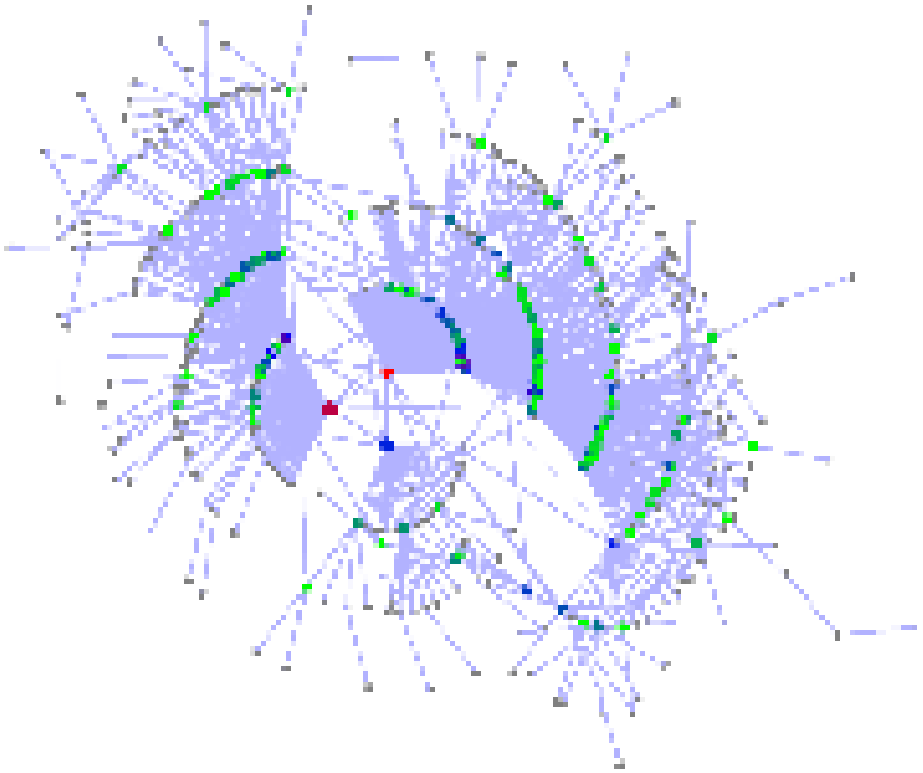
Router Technology Constraint



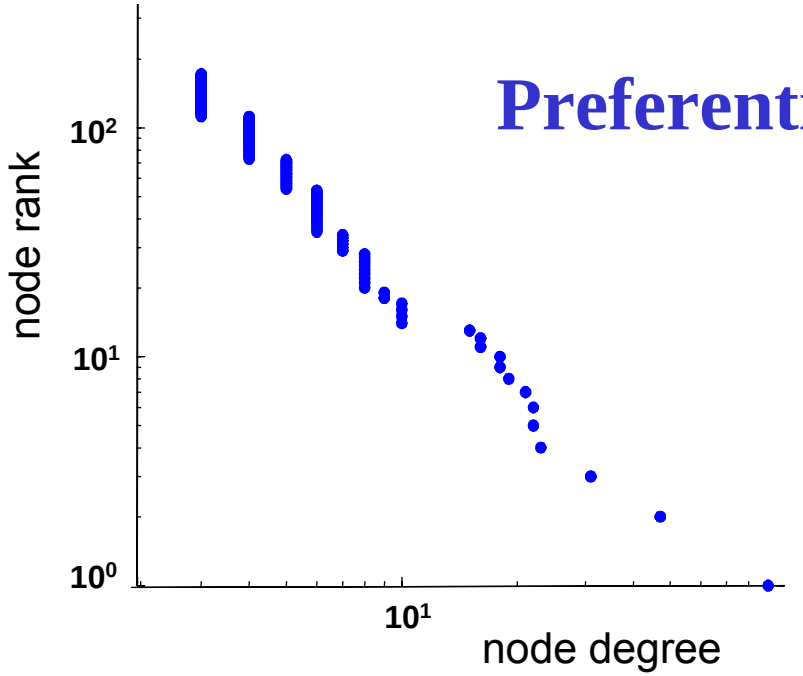
Router Deployment: Abilene and CENIC

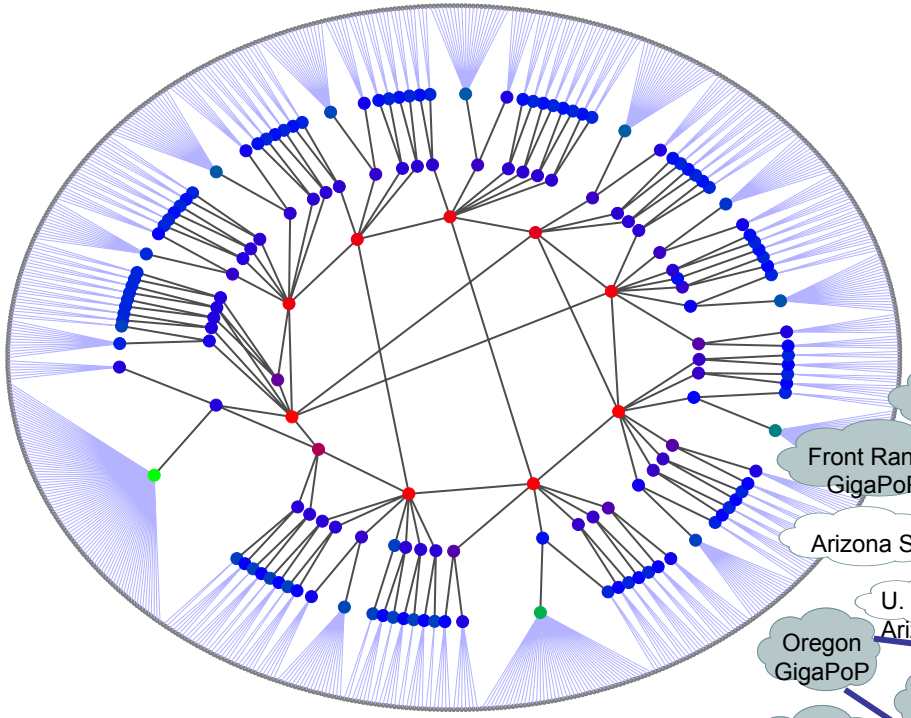






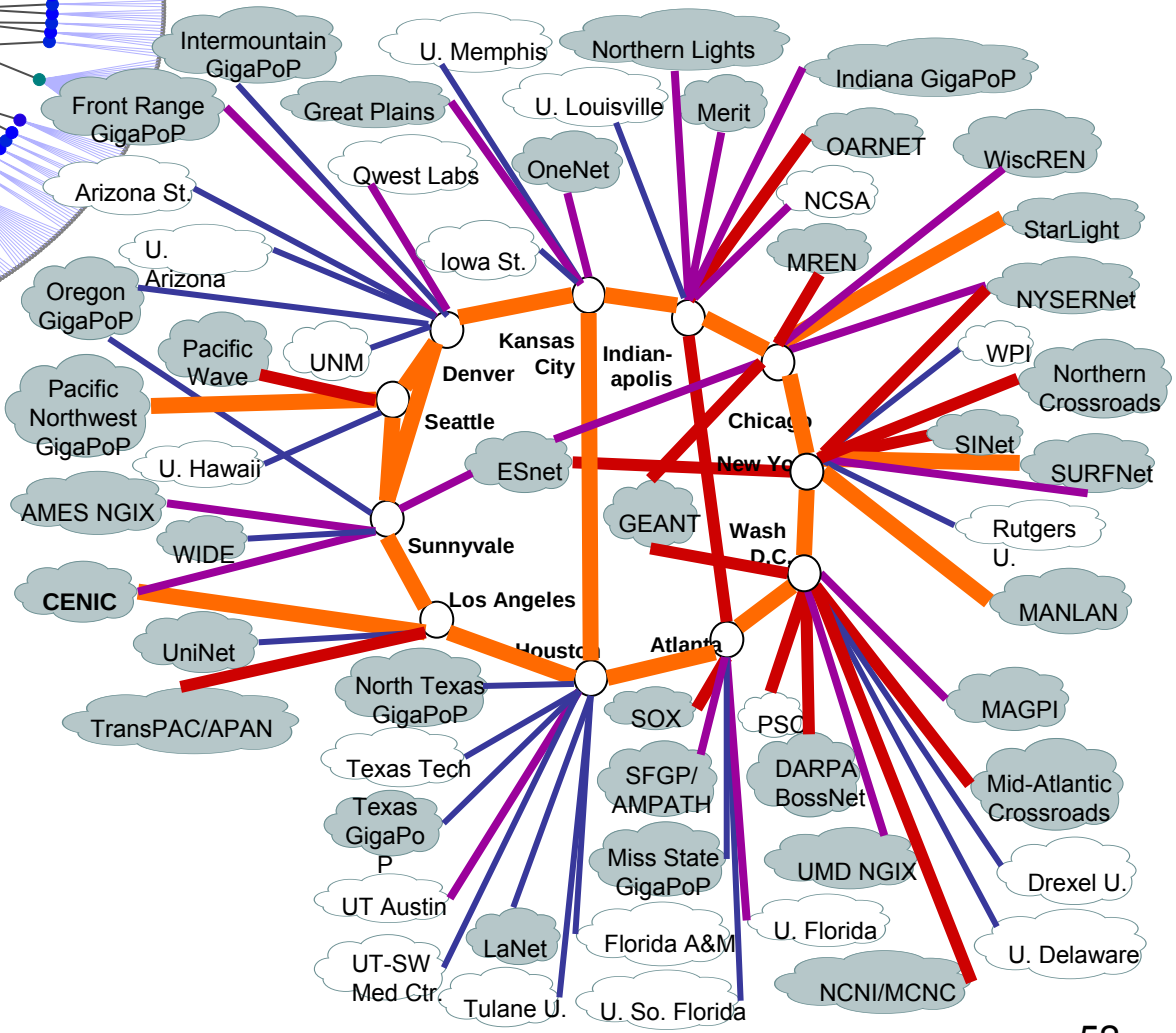
Preferential Attachment





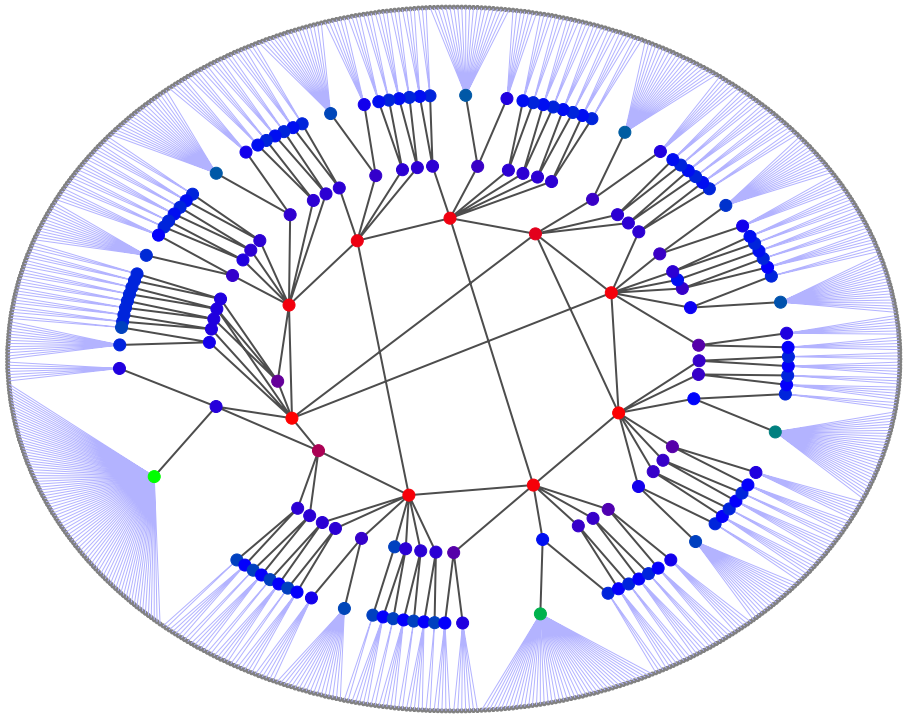
Abilene-inspired

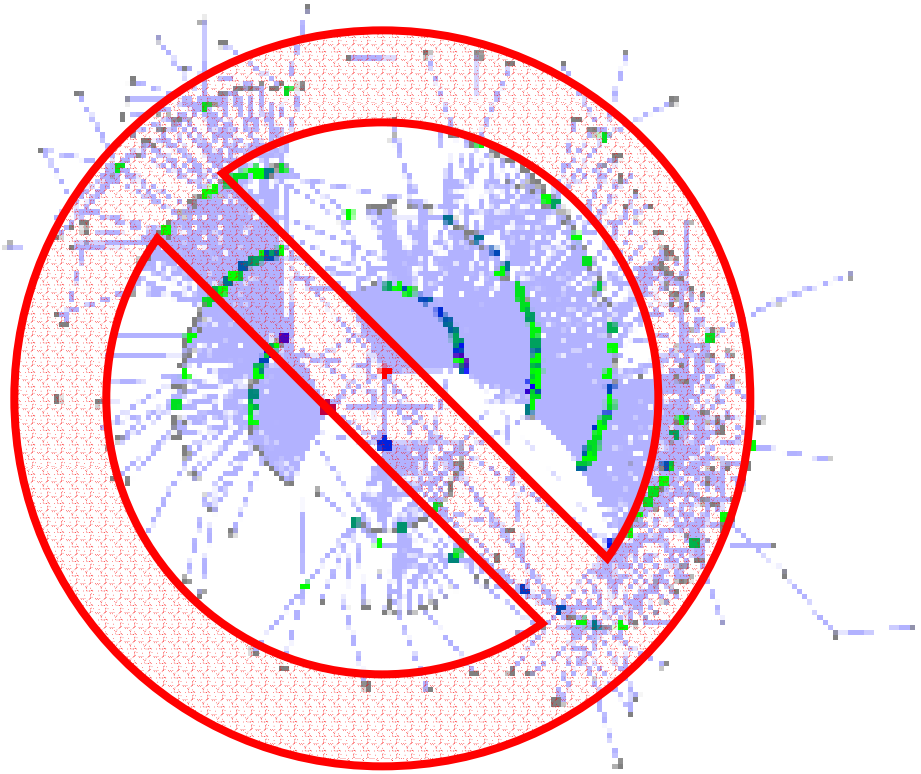
Abilene Backbone Physical Connectivity



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

MESSAGE #4: Importance of model validation





- Descriptive modeling that replicates statistical features is no more than an exercise in “data fitting”
- Matching given graph statistics should be a by-product and not a main focus of modeling
- Emphasis on “closing the loop” (using complementary measurements and domain expertise)

Take-Home Messages

- #1: Specify which aspect of Internet connectivity you are interested in
- #2: Internet connectivity measurements should never be taken at face value
- #3: Be aware of specious analysis of specious measurements
- #4: It is often easy to construct networks that agree with respect to certain graph statistics (e.g., same node degree distribution) but are otherwise completely different
- #5: Importance of model validation

Some open Questions

- ISPs design/evolve their networks for a purpose (see talk on Friday)
  - What is the purpose?
  - What are the constraints?
  - Where does randomness enter?
- What does the AS-level Internet as a whole try to achieve?
  - Objective, constraints, uncertainty?
- What is the purpose of a P2P network like Gnutella?
  - Objective, constraints, uncertainty?
- What does the Web graph as a whole try to achieve?
  - Hopeless, rely on randomness as main driver
- What about social networks?
  - An engineering perspective of social networks?

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 Mathematics 2(4),* 2006.
- D. Alderson, L. Li, W. Willinger, J.C. Doyle. Understanding Internet Topology: Principles, Models, and Validation. *ACM/IEEE Trans. on Networking* 13(6), 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 102(41)*, 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*.