Modeling of Complex Networks

Lecture 4: Internet

--Topology and Modeling

S8101003Q -01(Sem A, Fall 2021)

Instructor: Aaron, Haijun Zhang

Internet

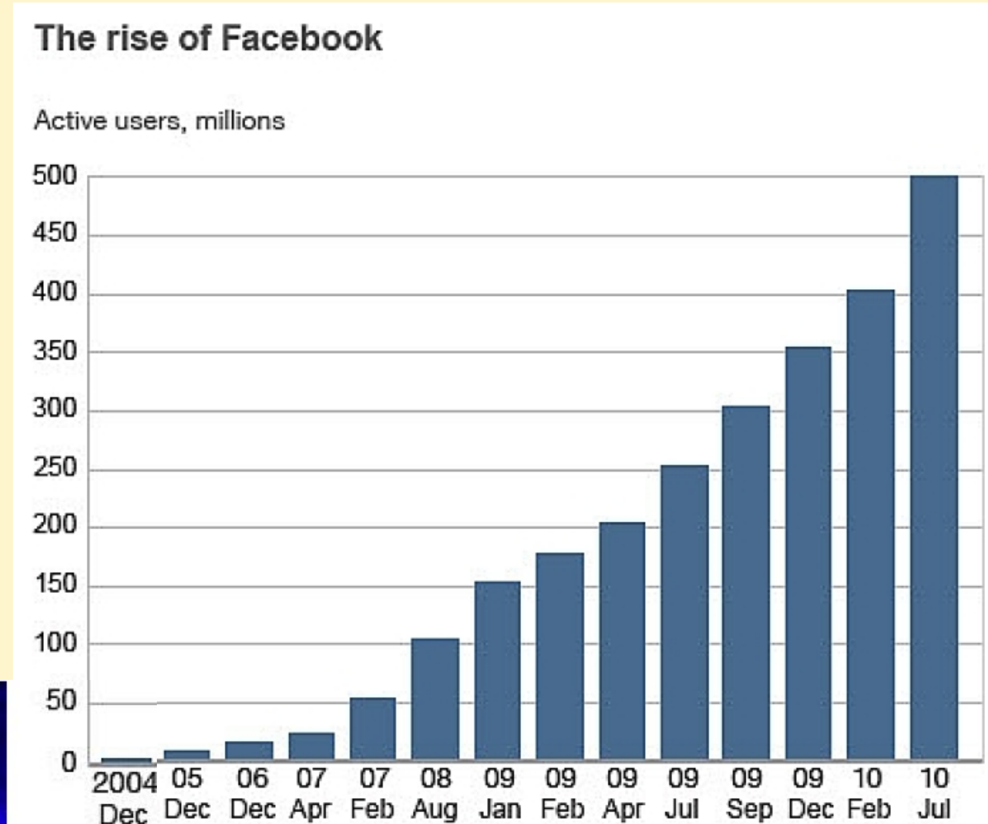
- Internet is a global system of interconnected computer networks that use the standard Internet Protocol Suite (TCP/IP) to serve billions of users worldwide (Wikipedia)

- Structure:
  - Autonomous Systems (AS): a collection of connected Internet Protocol (IP) routing prefixes under the control of one or more network operators that presents a common, clearly defined routing policy to the Internet (Wikipedia)
  - Wide Area Network (WAN)
  - Metropolitan Area Network (MAN)
  - Local Area Network (LAN)
- Software:
  - Transmission Control Protocol (TCP)
  - Internet Protocol (IP)
  - Other protocols at the application level:
    - Internet Control Message Protocol (ICMP)
    - File Transfer Protocol (FTP)
    - Simple Mail Transfer Protocol (SMTP)

Internet is a network of networks (structure)

- Internet is the inter-connections of networks:
  - Internet Service Provider (ISP) networks
  - Corporate networks (e.g. Cisco and Microsoft)
    - ✓ A corporate network is a group of computers, connected together in a building or in a particular area, which are all owned by the same company.
  - Campus networks (e.g. HITSZ)
  - Mobile operator networks (e.g. one2free)
  - Government networks
  - Internet Exchanges (e.g. HKIX)
    - ✓ An Internet exchange point (IX or IXP) is a physical infrastructure through which Internet service providers(ISPs) exchange Internet traffic between their networks (autonomous systems).
- through optical fibers (telephone lines)

A Typical Application Example - Facebook



www.bbc.co.uk

Data Source

Protocols

- Protocol originates from the Greek word protocollon, which was a piece of paper glued to a manuscript to describe its contents.
- In computing and communication systems, a protocol is a set of rules, a convention, or a standard, that controls the connection and data transfer between endpoints.
- Telecommunication systems:
  - data interchange protocols at the hardware device level
  - data interchange protocols at the application program level
- Internet:
  - Transmission Control Protocol (TCP), for exchanging messages with other Internet points at the information packet level
  - Internet Protocol (IP), for sending and receiving messages at the Internet address level
  - Other protocols: Hypertext Transfer Protocol (HTTP), File Transfer Protocol (FTP), Border Gateway Protocol (BGP), and Dynamic Host Configuration Protocol (DHCP), ...

Brief History of the Internet

- 1968 - DARPA (Defense Advanced Research Projects Agency) contracts with BBN (Bolt, Beranek & Newman) to create ARPAnet

- 1970 - First five nodes:
  UCLA, Stanford, UC Santa Barbara, U of Utah, BBN

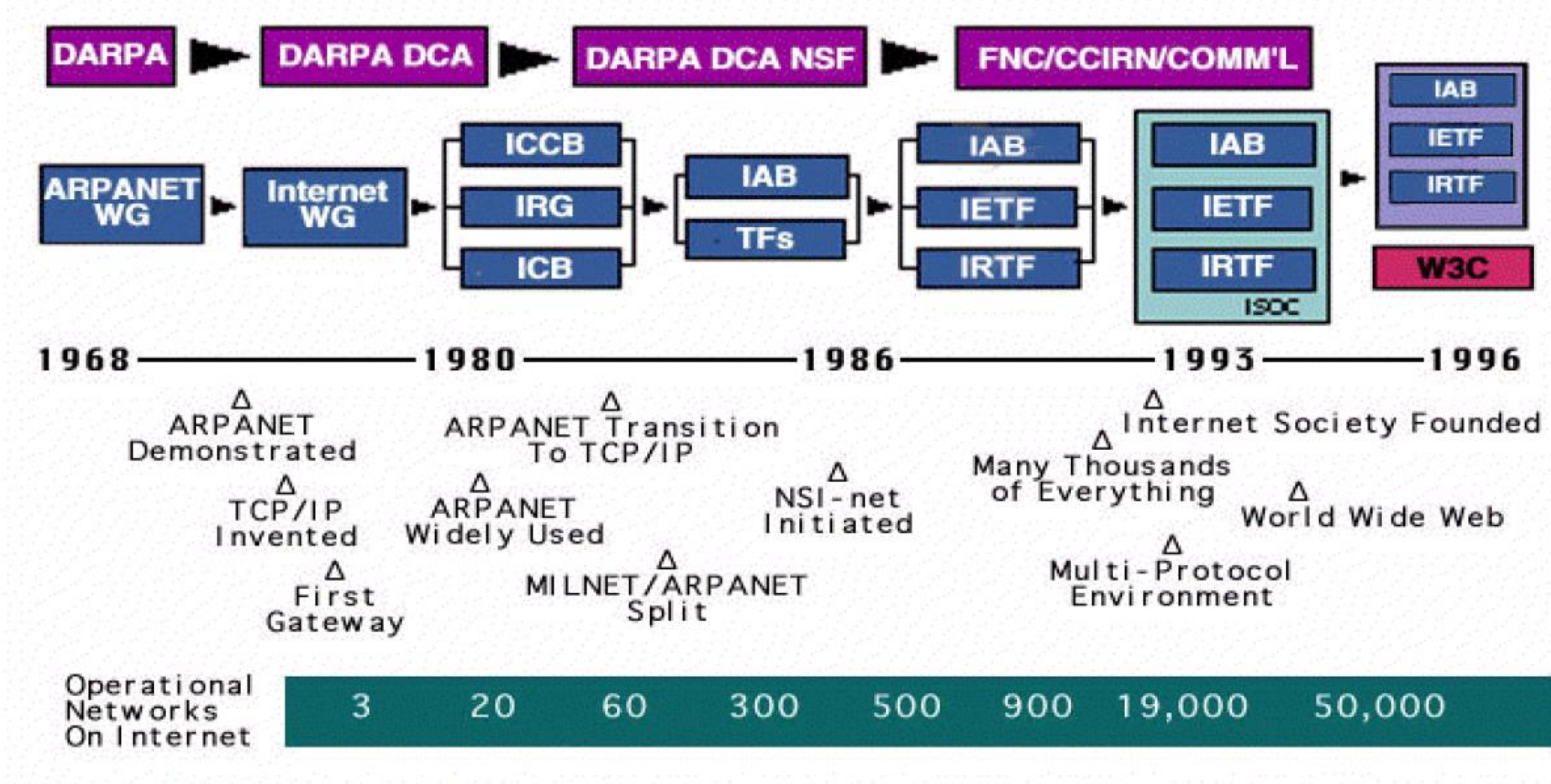
- 1974 - TCP specification by Vint Cerf

- 1984 - January 1,1984, the Internet with its 1000 hosts converts altogether to using TCP/IP for its messaging

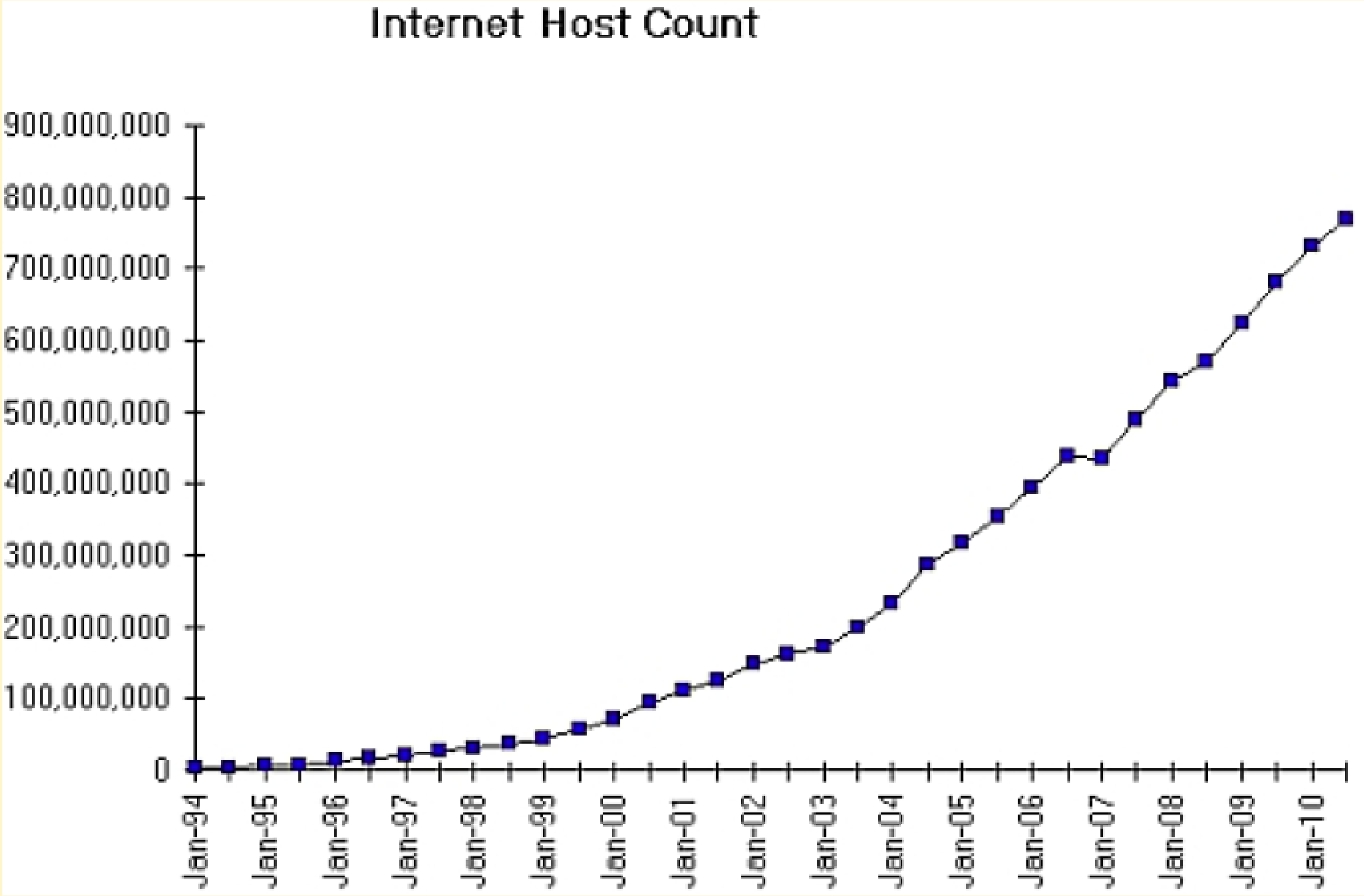
- ……

Earlier History of Internet

DARPA



Internet Statistics



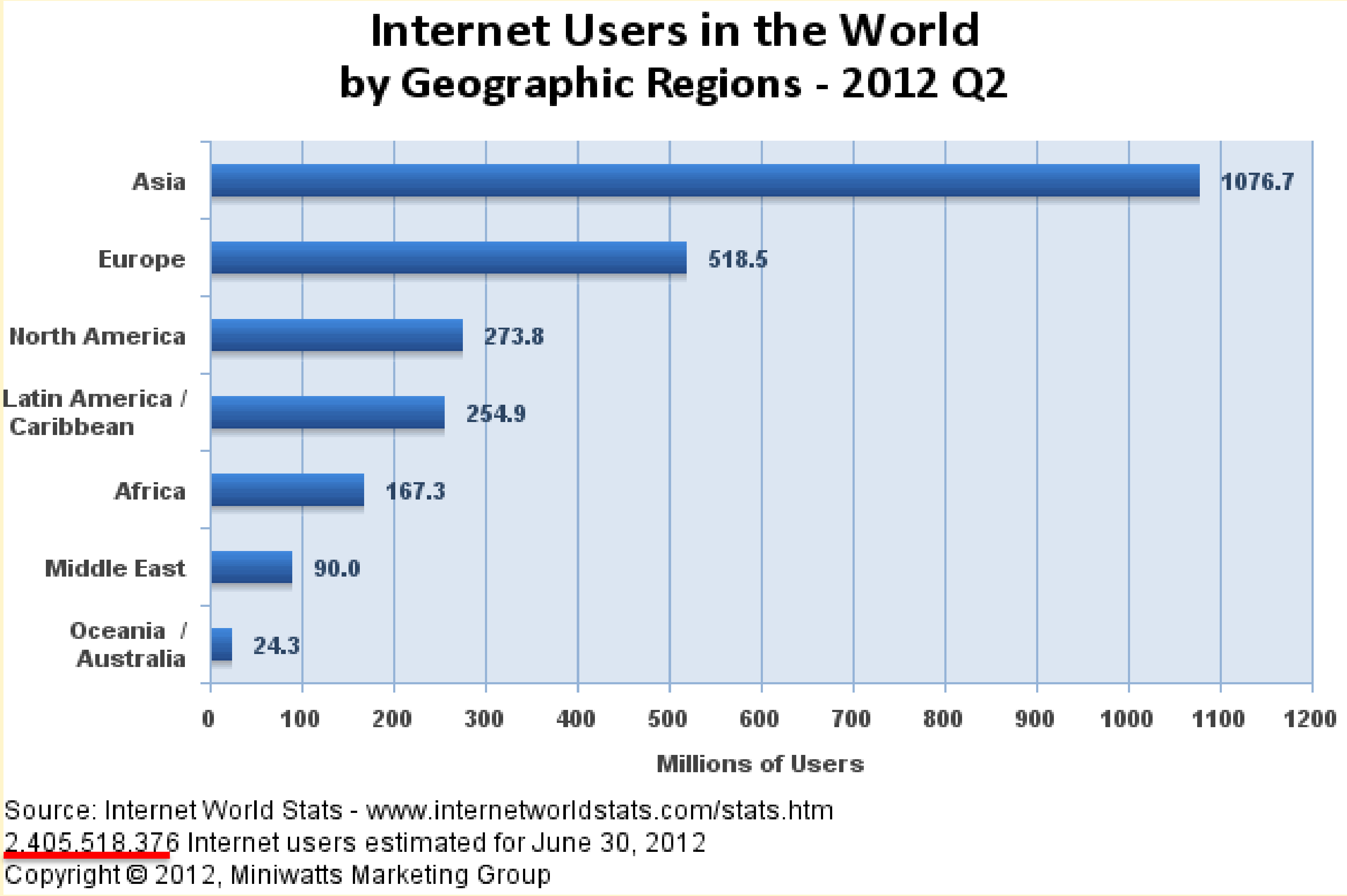
Host count is 768,913,036 in 2010
93,047,785 in 2000

826.36%

[Data Source](#)

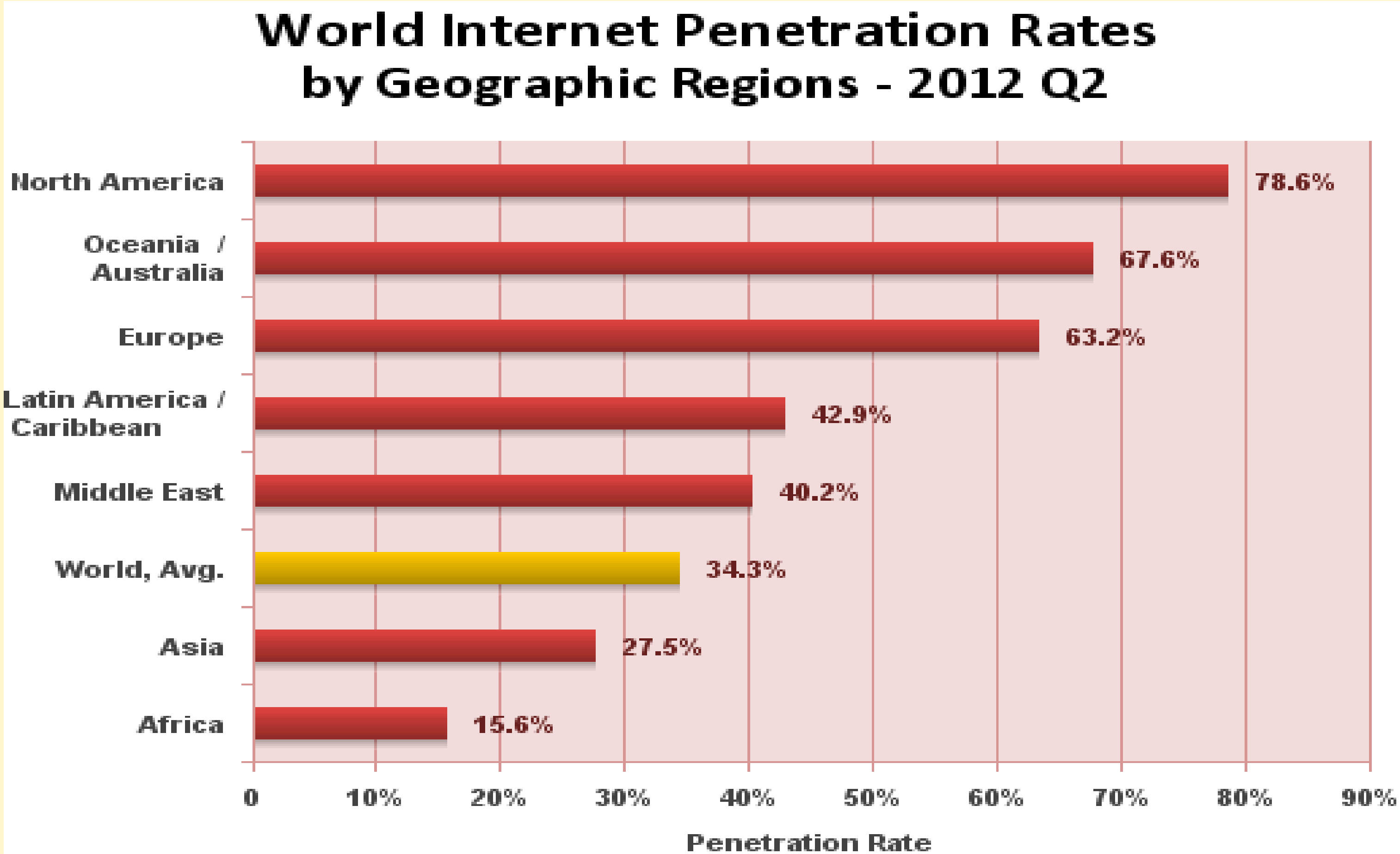
Recent Internet Statistics

www.internetworldstats.com



Recent Internet Statistics

www.internetworldstats.com



Source: Internet World Stats - www.internetworldststs.com/stats.htm
Penetration Rates are based on a world population of 7,017,846,922
and 2,405,518,376 estimated Internet users on June 30, 2012.

Recent Internet Statistics

www.internetworldstats.com

The Internet Big Picture

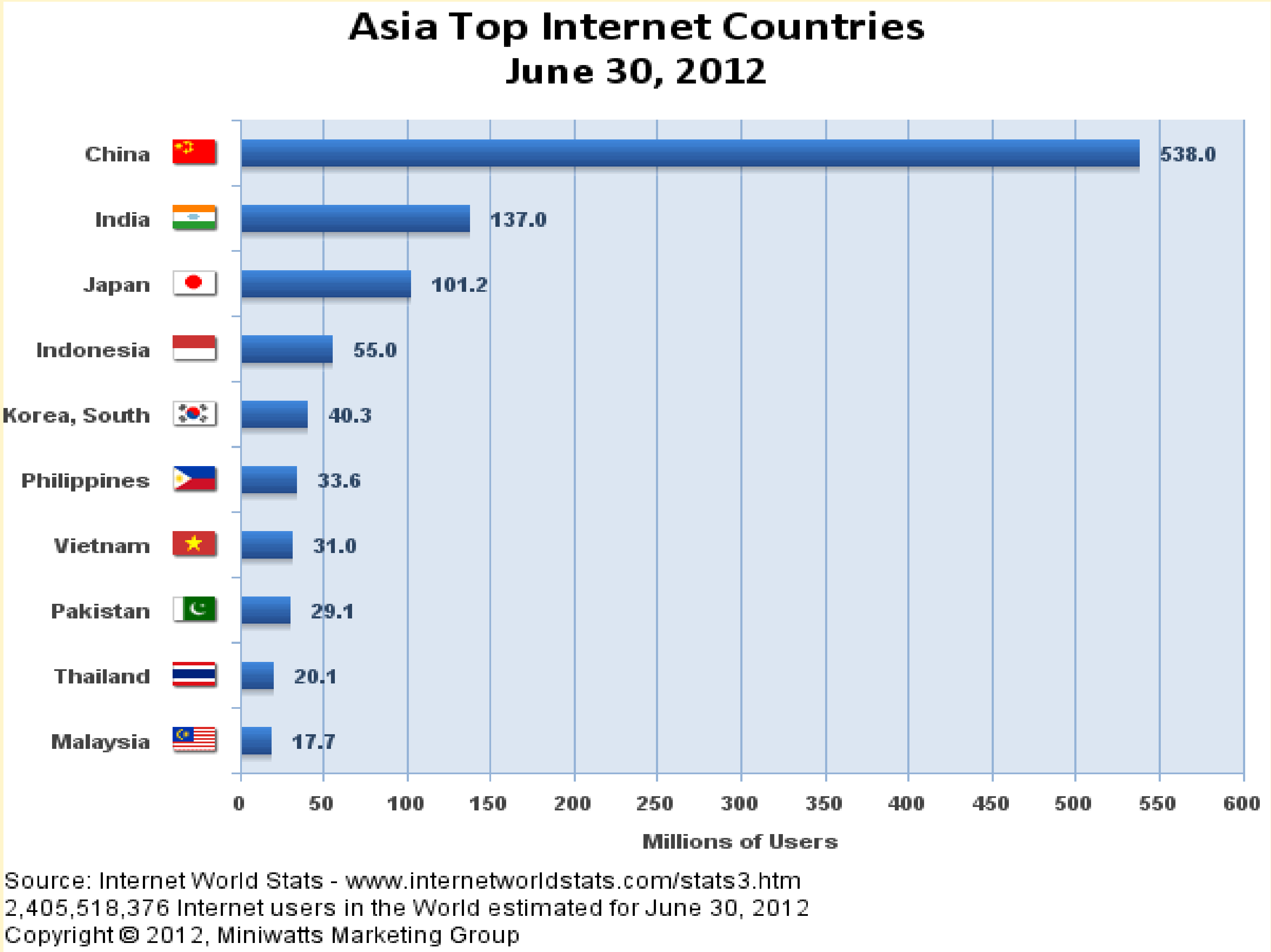
World Internet Users and Population Stats

WORLD INTERNET USAGE AND POPULATION STATISTICS June 30, 2012						
World Regions	Population (2012 Est.)	Internet Users Dec. 31, 2000	Internet Users Latest Data	Penetration (% Population)	Growth 2000-2012	Users % of Table
Africa	1,073,380,925	4,514,400	**167,335,676**	15.6 %	3,606.7 %	7.0 %
Asia	3,922,066,987	114,304,000	**1,076,681,059**	27.5 %	841.9 %	44.8 %
Europe	820,918,446	105,096,093	**518,512,109**	63.2 %	393.4 %	21.5 %
Middle East	223,608,203	3,284,800	**90,000,455**	40.2 %	2,639.9 %	3.7 %
North America	348,280,154	108,096,800	**273,785,413**	78.6 %	153.3 %	11.4 %
Latin America / Caribbean	593,688,638	18,068,919	**254,915,745**	42.9 %	1,310.8 %	10.6 %
Oceania / Australia	35,903,569	7,620,480	**24,287,919**	67.6 %	218.7 %	1.0 %
WORLD TOTAL	**7,017,846,922**	**360,985,492**	**2,405,518,376**	**34.3 %**	**566.4 %**	**100.0 %**

NOTES: (1) Internet Usage and World Population Statistics are for June 30, 2012. (2) CLICK on each world region name for detailed regional usage information. (3) Demographic (Population) numbers are based on data from the US Census Bureau and local census agencies. (4) Internet usage information comes from data published by Nielsen Online, by the International Telecommunications Union, by GfK, local ICT Regulators and other reliable sources. (5) For definitions, disclaimers, navigation help and methodology, please refer to the Site Surfing Guide. (6) Information in this site may be cited, giving the due credit to www.internetworldstats.com.

Recent Internet Statistics

www.internetworldstats.com



Recent Internet Statistics

World Population: 7,017,846,922

Number of Users: 2,405,518,376

- Internet backbone traffic grew from one Tbit/sec in 1990 to 3,000 Tbit/sec in 1997
- Number of Internet hosts more than doubled every year between 1980-2000

China (in millions) CNNIC

Internet users: 538

IPv4 Addresses: 330

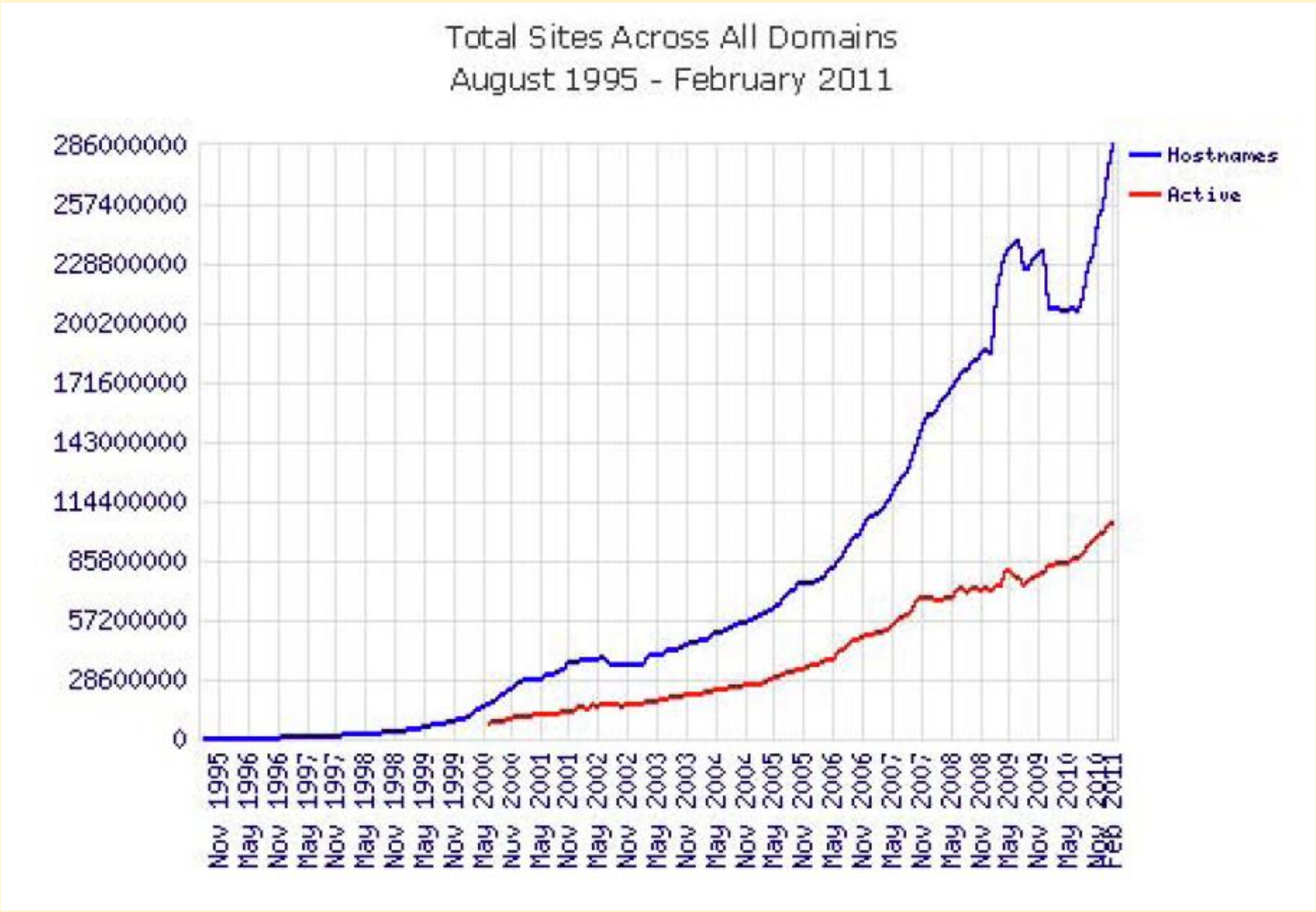
Domain Names: 873

WWW Websites: 250

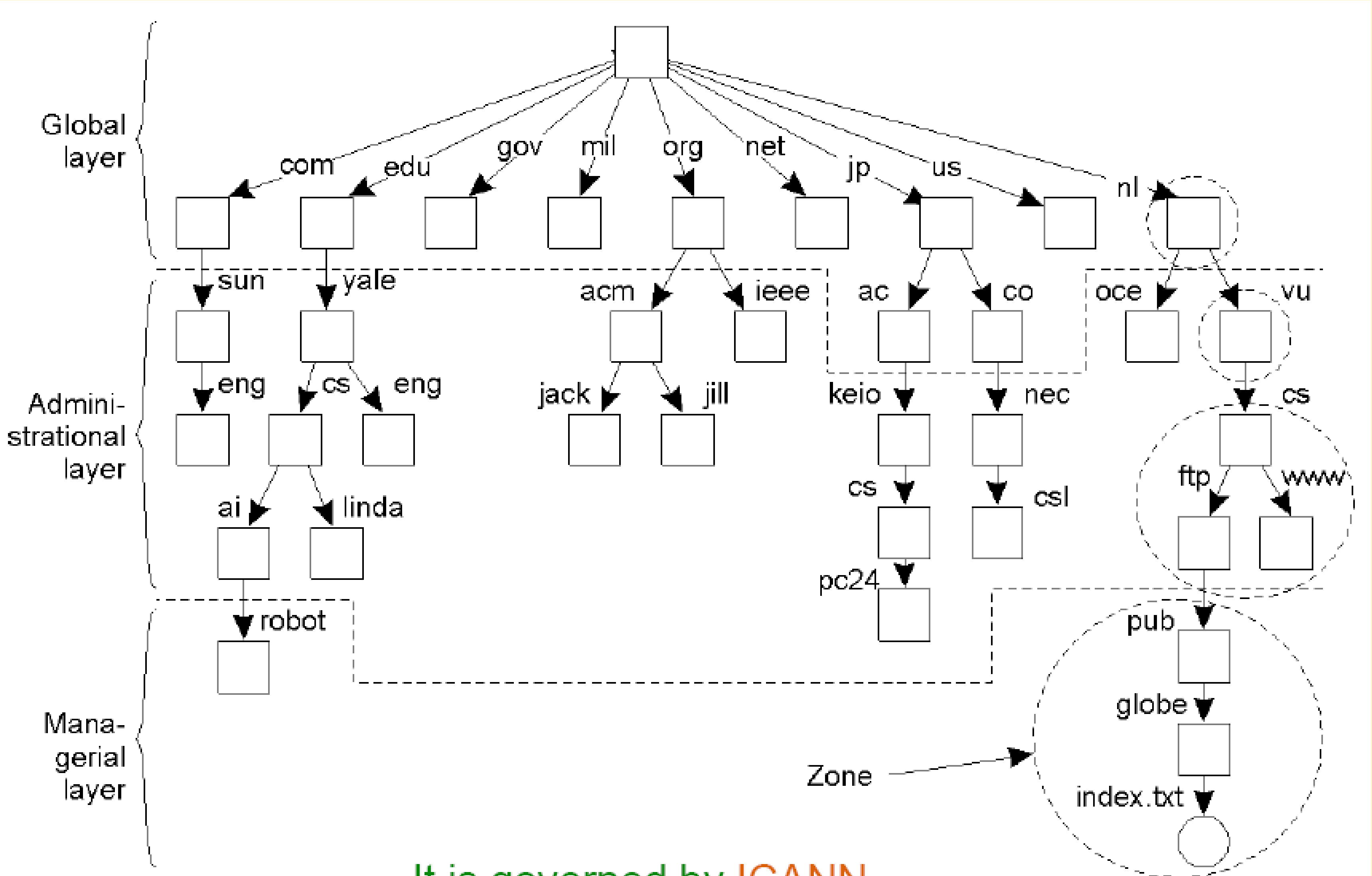
International Bandwidth: 1,548,811 Mbps

Recent Internet Statistics

Netcraft - an Internet services company based in Bath, England, providing web server and web hosting market-share analysis, including web server and operating system detection.



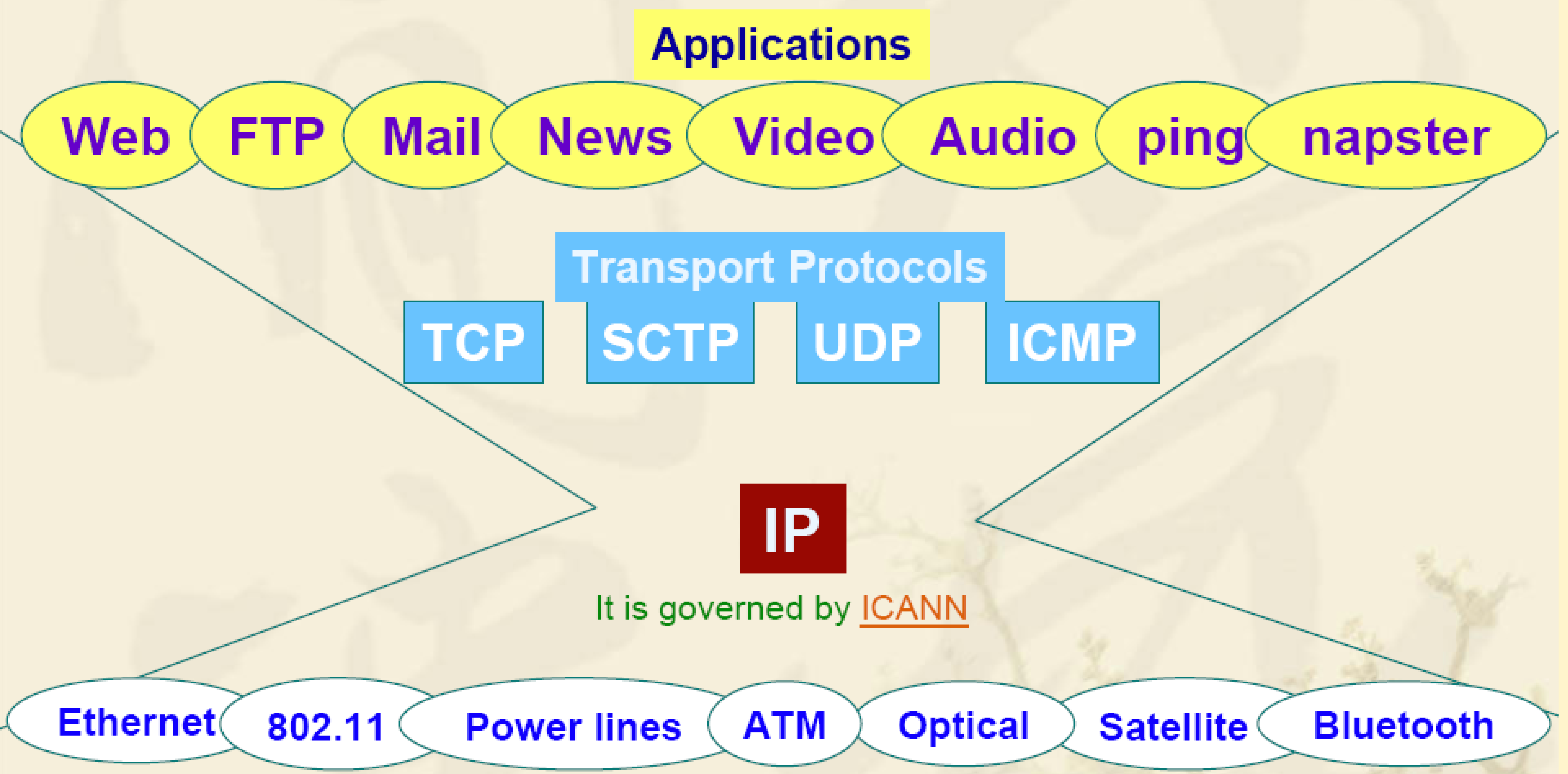
Internet



It is governed by ICANN

http://www.icann.org/

Internet

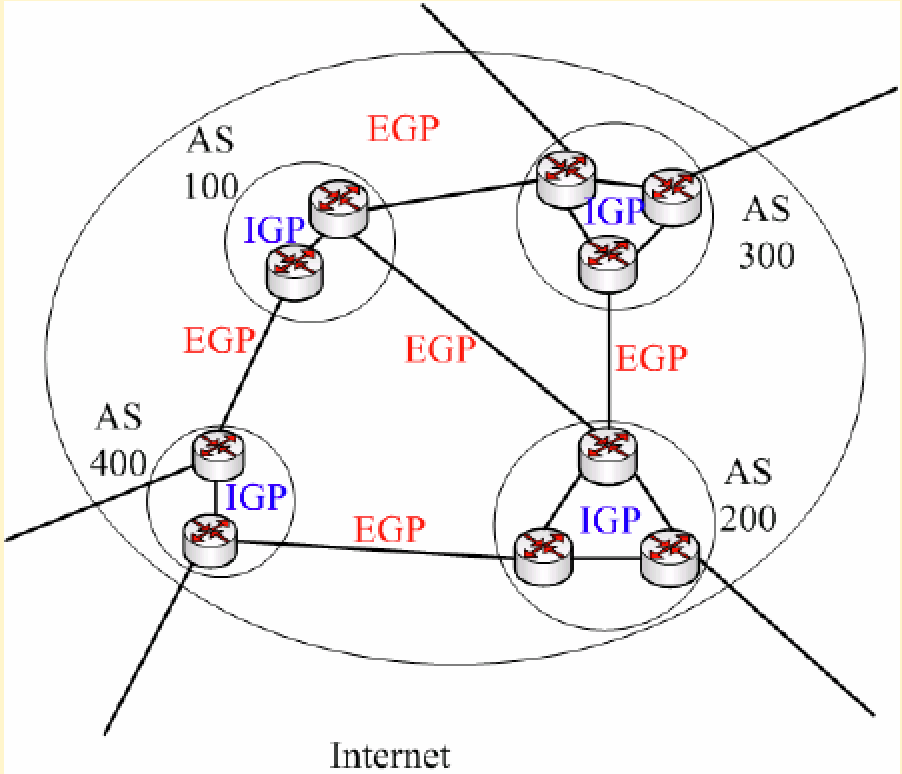


Link Technologies

Hari Balakrishnan

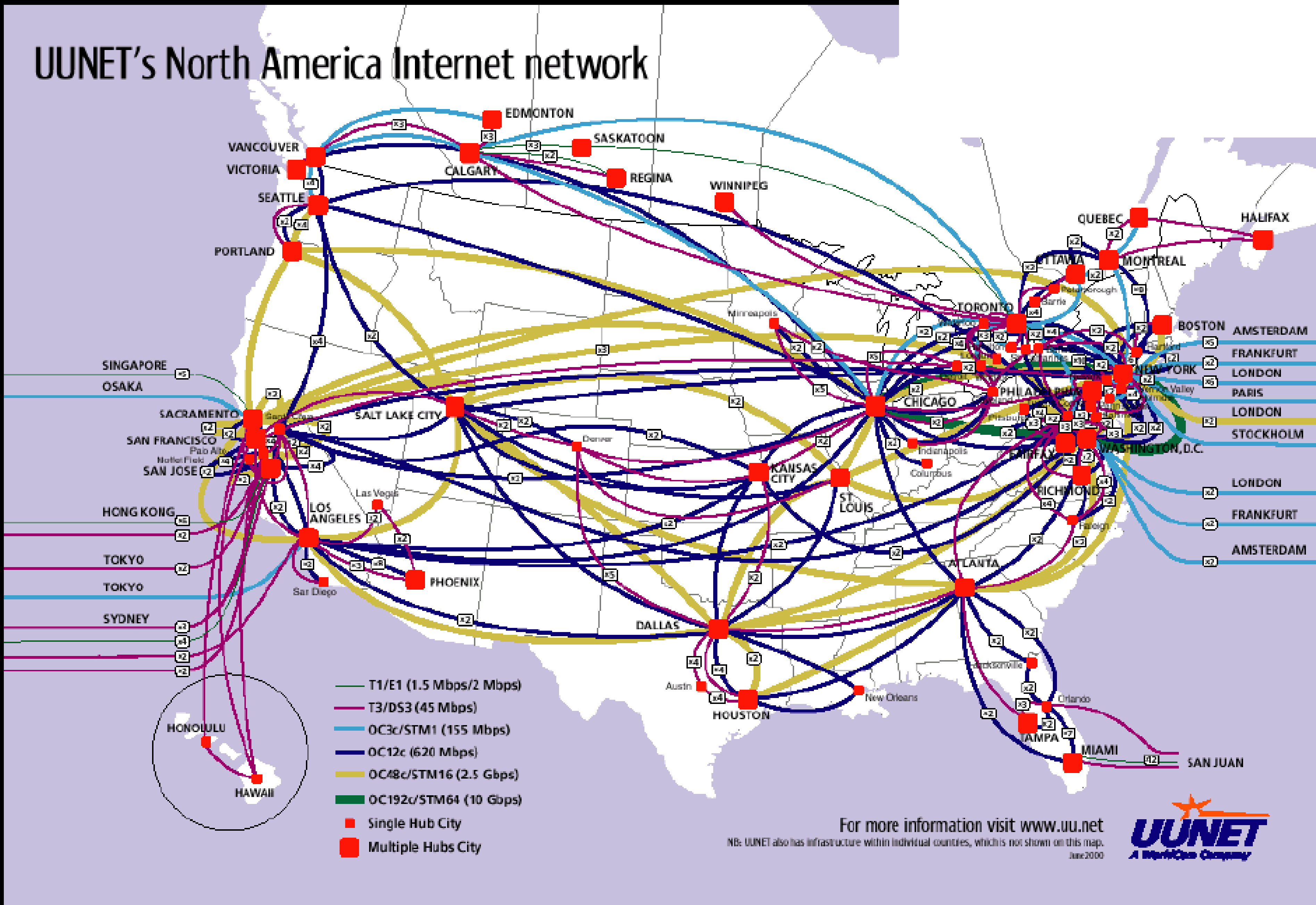
Internet

Backbone (2001):
27.9% -UUNET/WorldCom/MCI
10.0% -AT&T
6.5% -Sprint
6.3% - Genuity
4.1% - PSINet
3.5% - Cable & Wireless
2.8% - XO Communications
2.6% -Verio
1.5% - Qwest
1.3% - Global Crossing

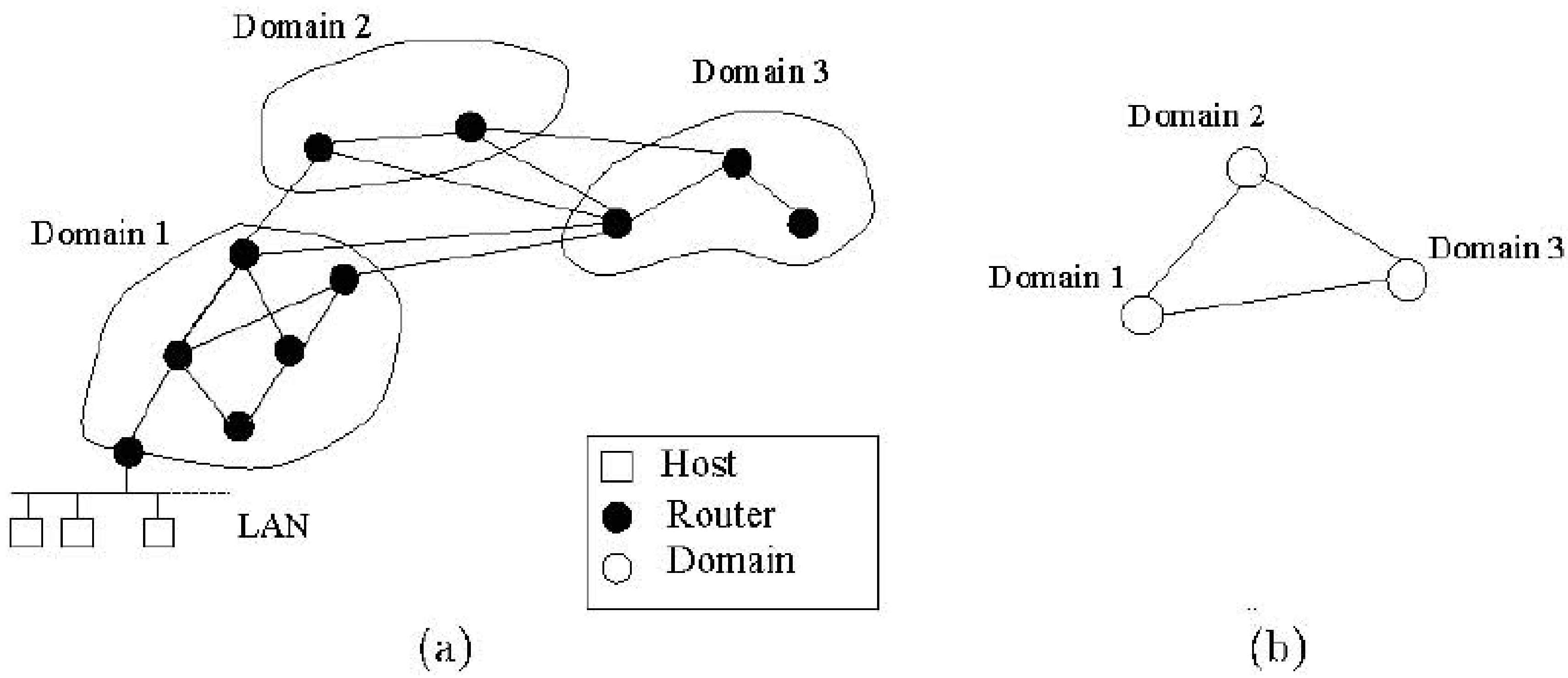


EGP – Exterior Gateway Protocol
IGP – Interior Gateway Protocol

UUNET was founded in 1987: one of the largest Internet service providers and one of the nine Tier-1 networks

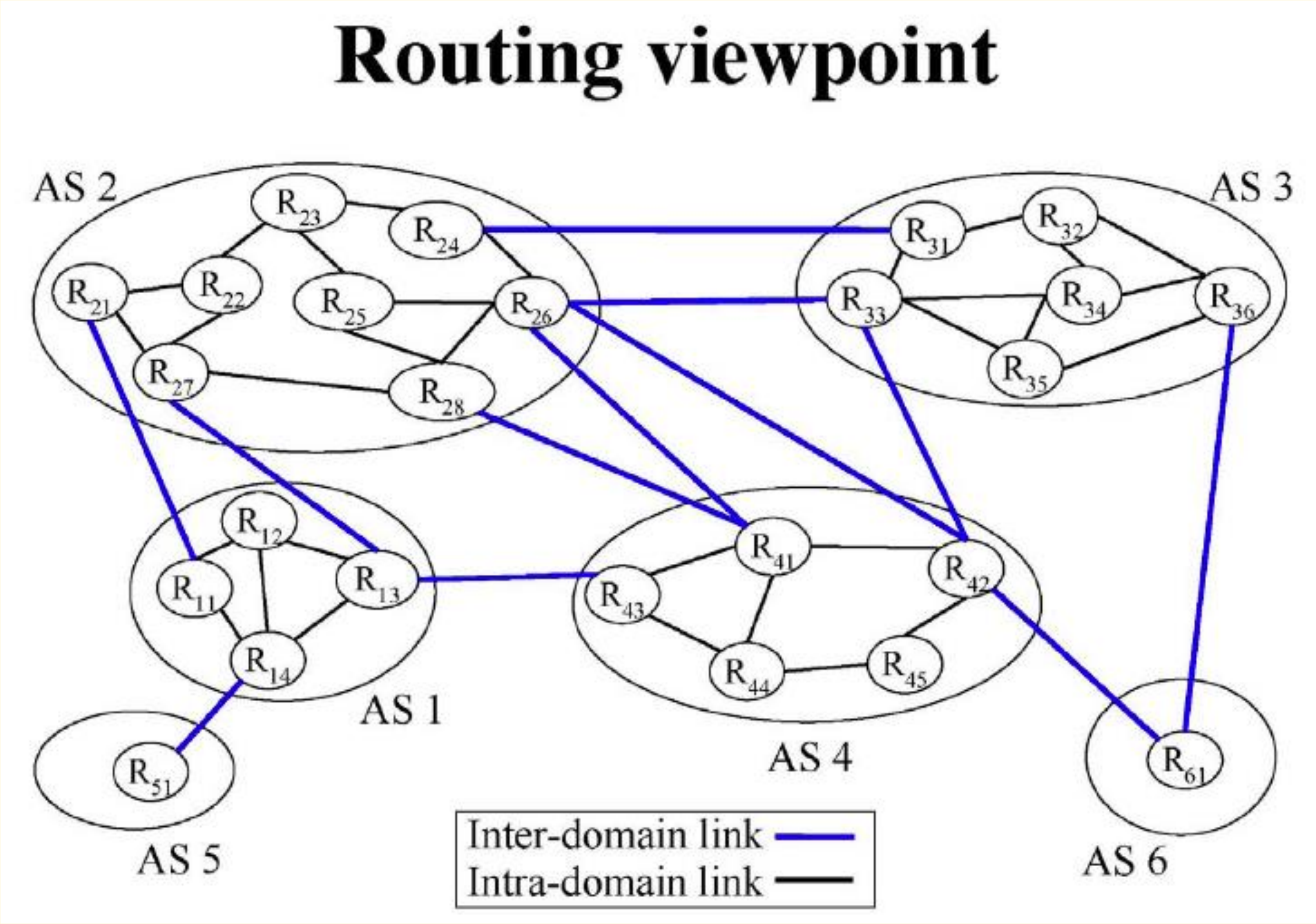


Internet at the AS Level



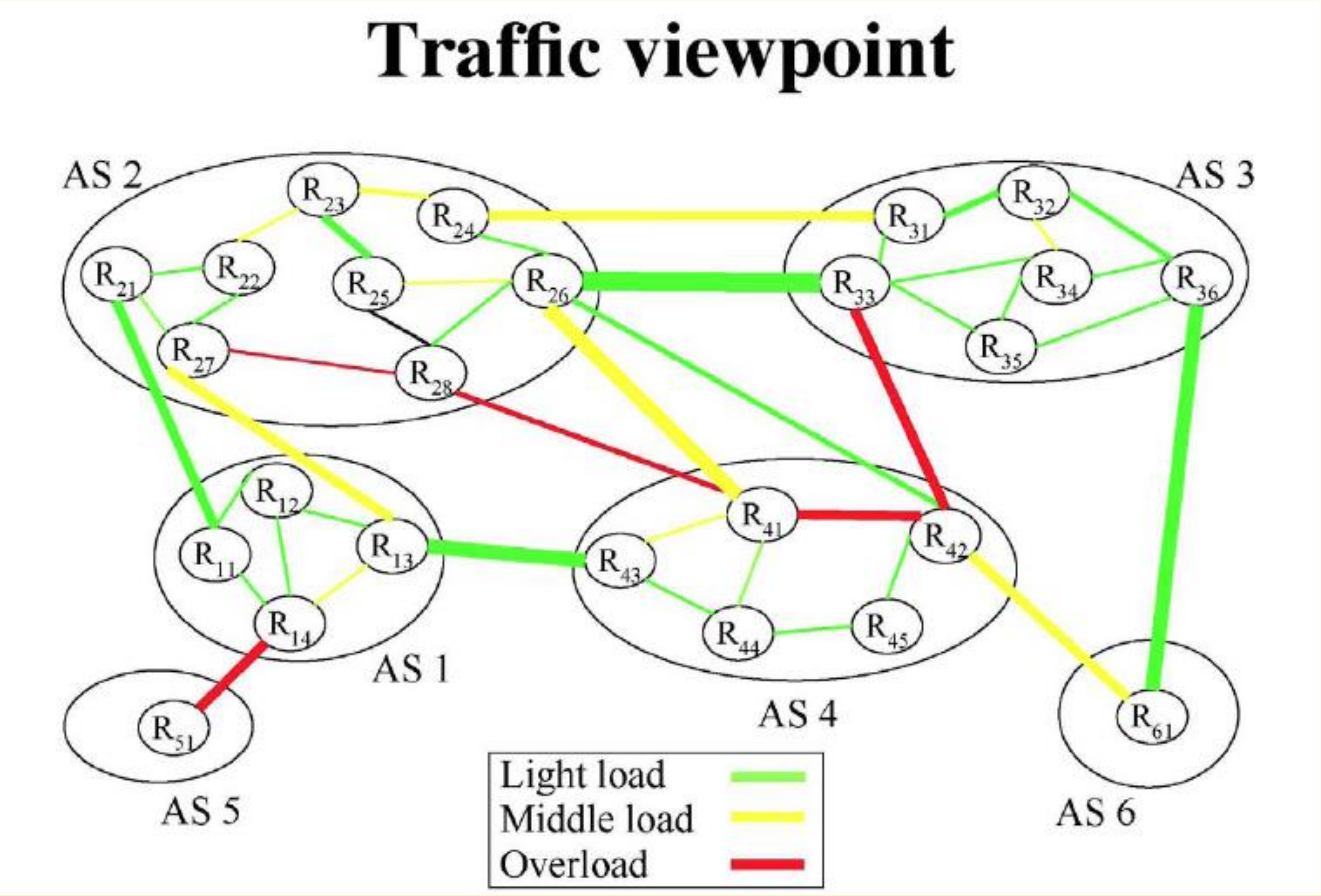
Internet at the **Autonomous Systems** (**AS**) Level
(Faloutsos brothers, 1999)

Internet at the AS Level



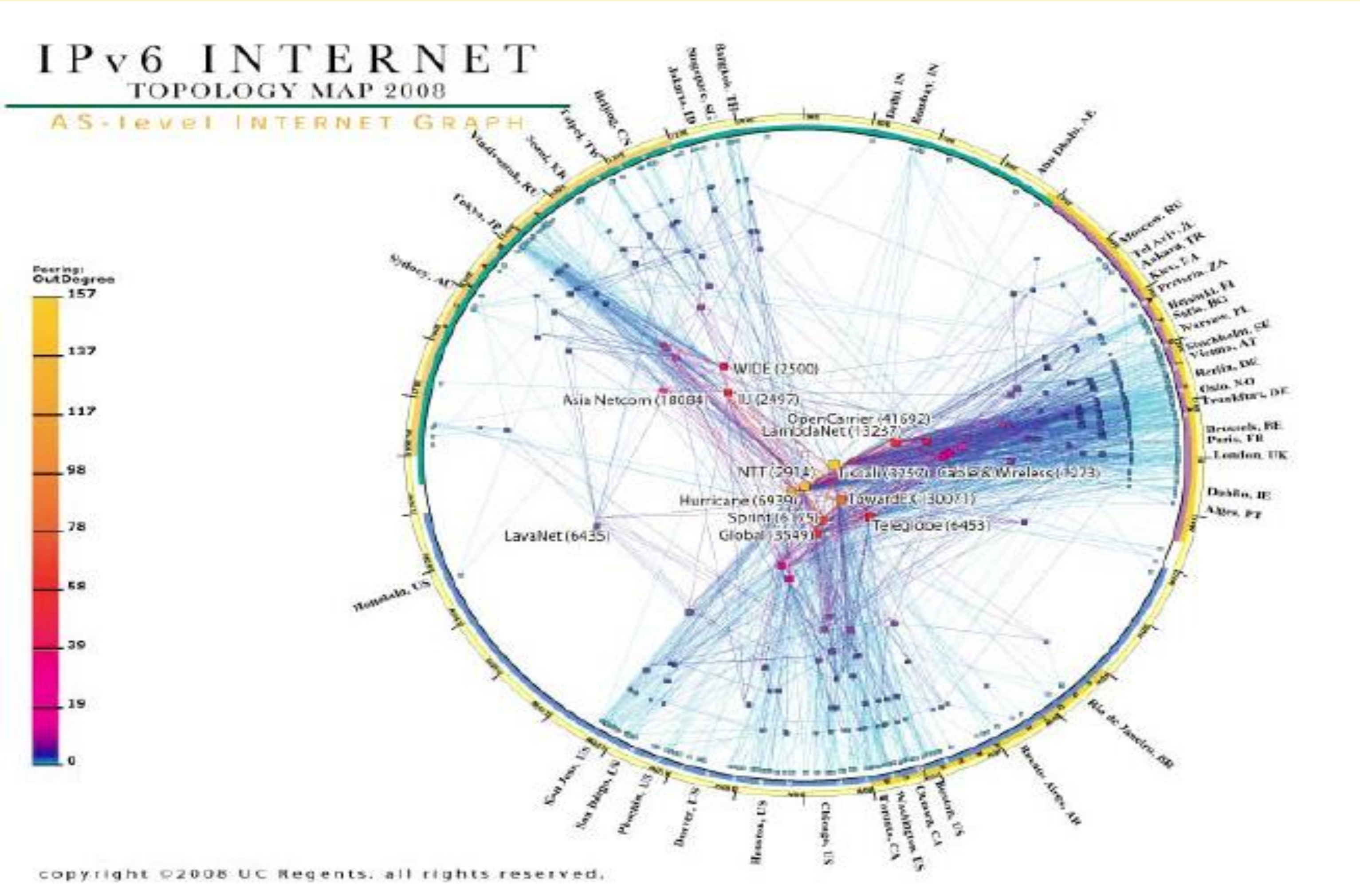
(Steve Uhlig)

Internet at the AS Level



(Steve Uhlig)

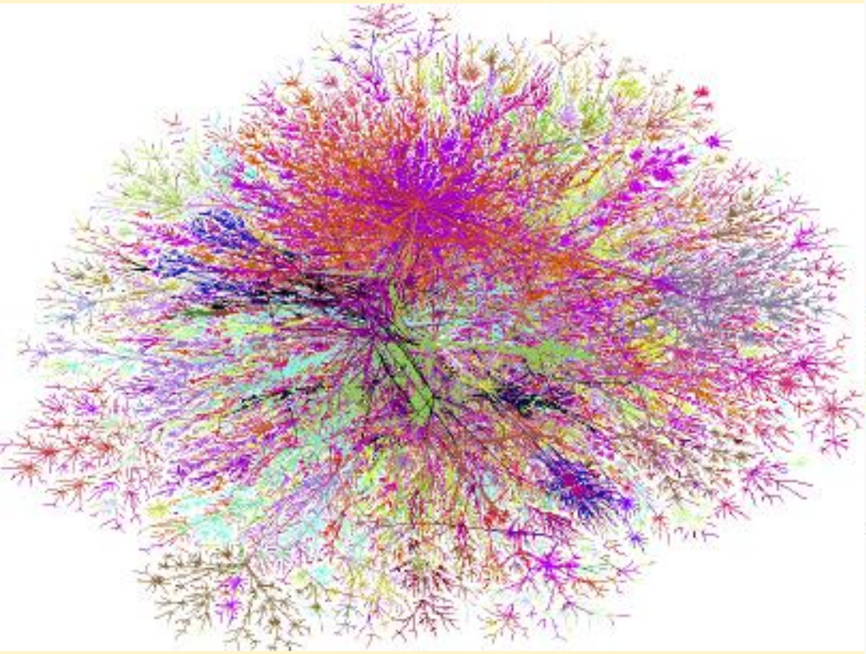
Internet Visualization



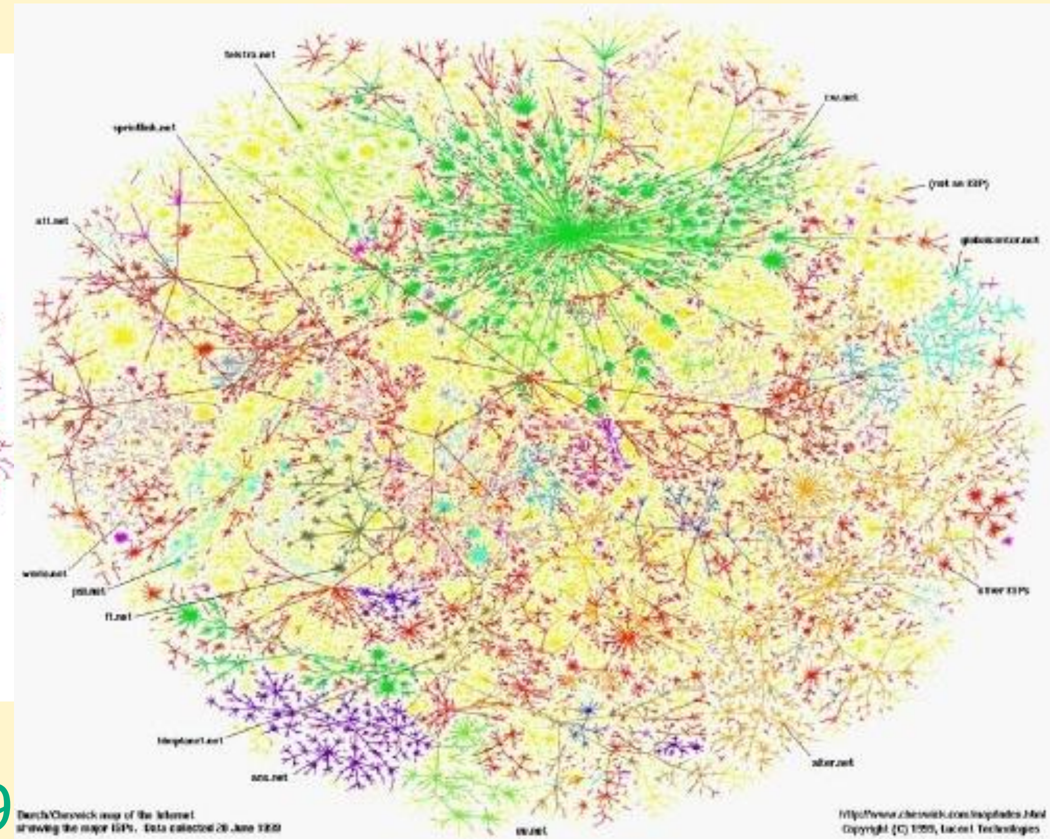
4,752 IPv6 addresses and 526 IPv6 prefixes (2008)

CAIDA Topology Mapping Analysis Team

Internet Visualization



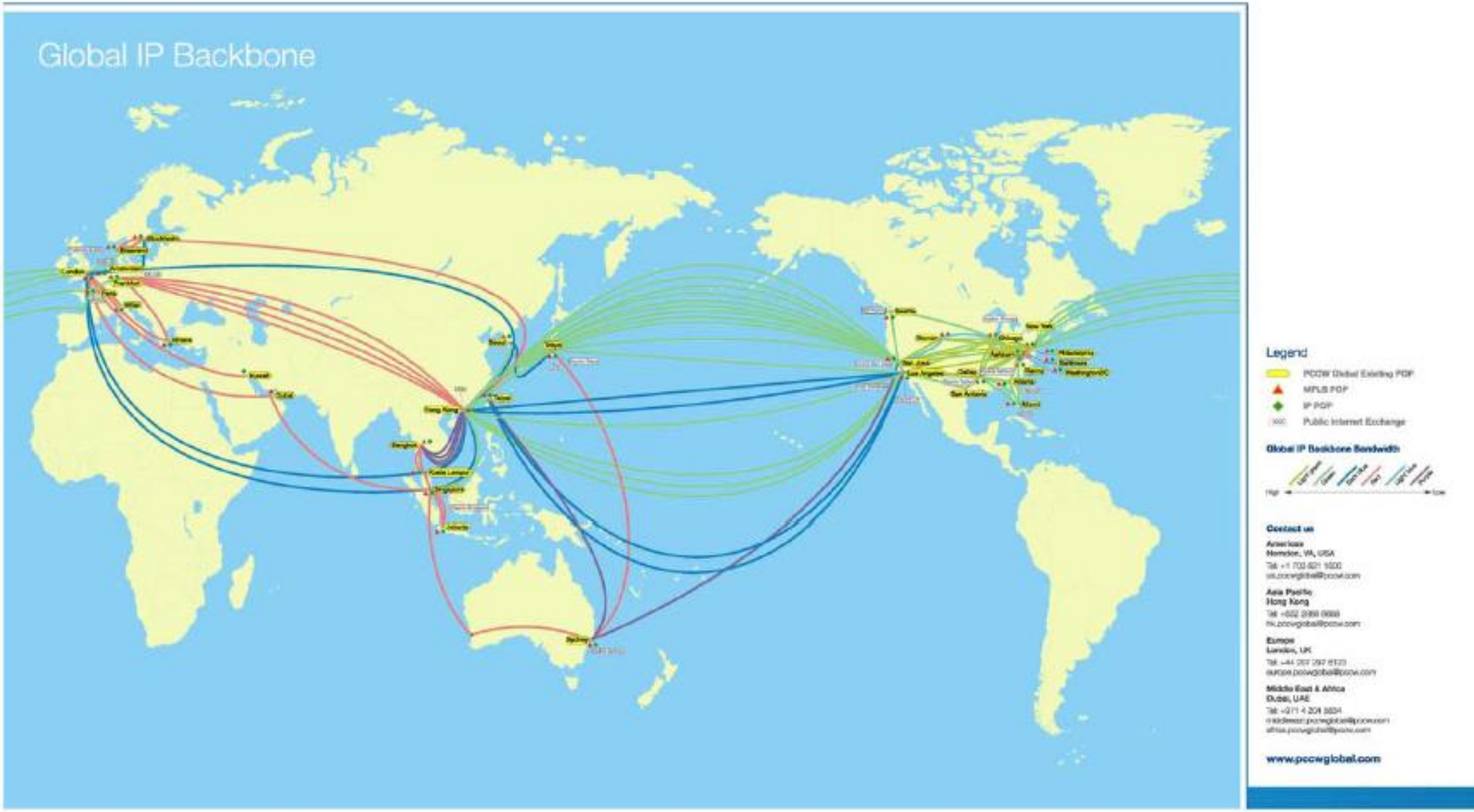
1998



1999

IP addresses at the router level of the Internet

(William R. Cheswick)



To Connect & Serve

PCCW Global

Internet Topology Generators

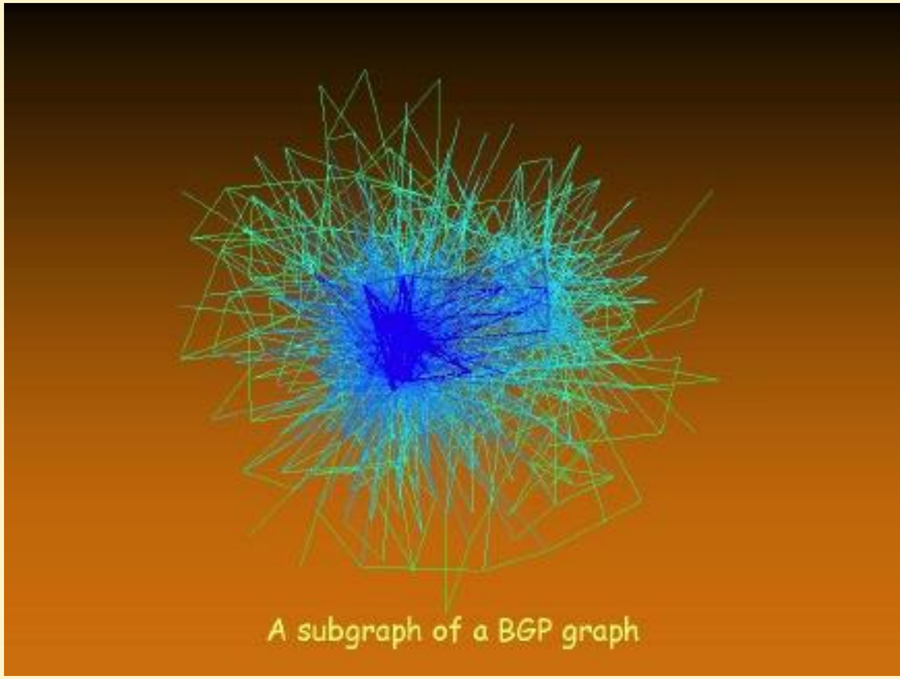
- Three stages of development:
  - First generation includes: random topology generators (in 1980s); representative - Waxman generator
  - Second generation includes: structural topology generators (in 1990s); representatives - Tiers and Transit-Stub generators (with hierarchical structures)
  - Third generation (since 2000), based on node degrees; representatives - BRITE and Inet, small-world and especially scale-free network models

  - IRL UCLA -- Internet Topology Collection

Internet Topological Properties

- Real AS-level Internet data can be obtained from the website of the Oregon Router Views Project, which was managed by the National Laboratory for Applied Network Research (NLANR), expired at the end of 2006, and now managed by the Cooperative Association for Internet Data Analysis (CAIDA)

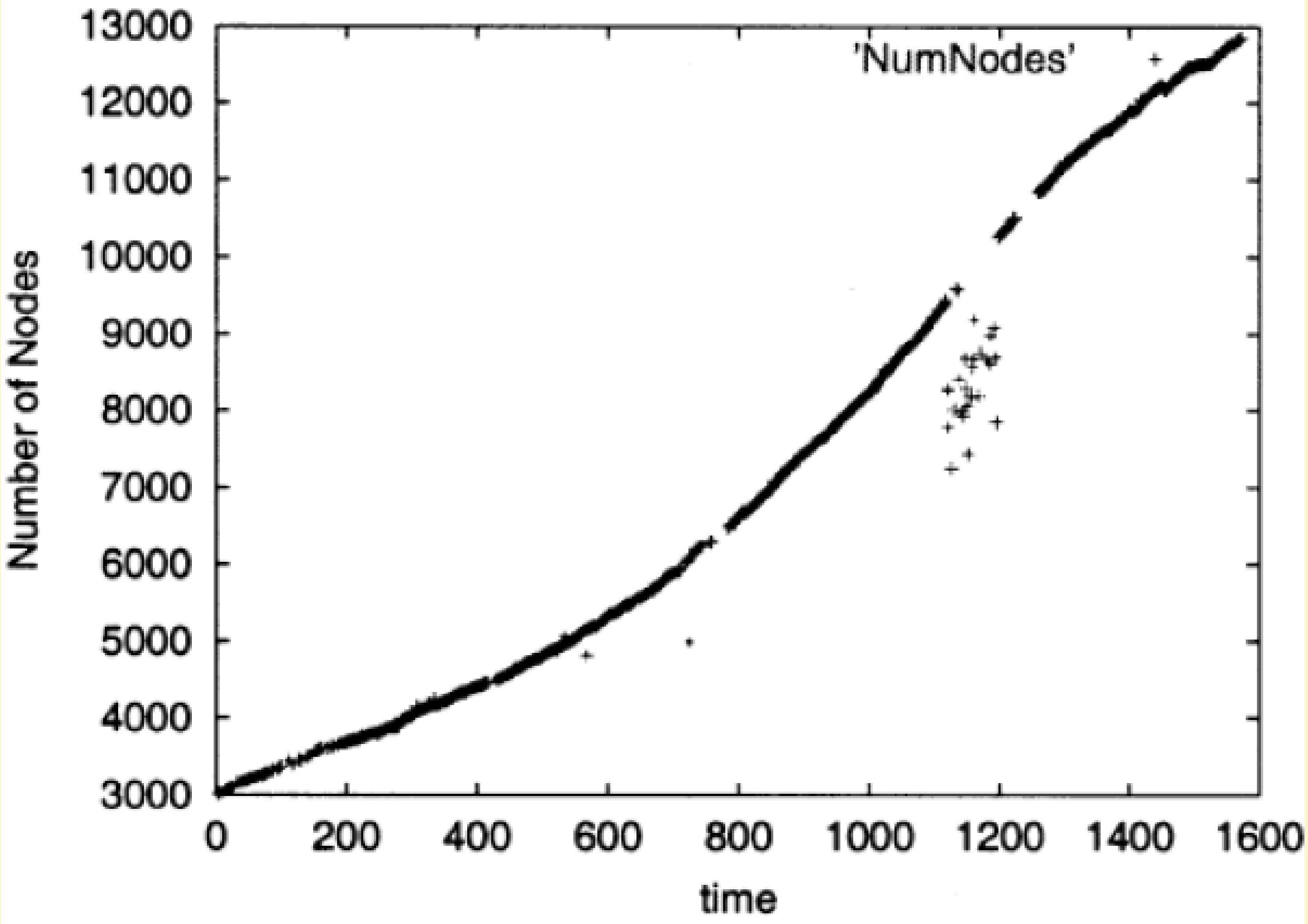
- This website is being updated within hours daily by taking snapshots from the routing tables of the Border Gateway Protocol (BGP)



Internet Topological Properties

- Other useful data about the AS-level Internet:
- Skitter provides Internet topological measures by the CAIDA, using Traceroute (a computer network tool for determining the route taken by packets across an IP network.
- Whois is a domain search tool and data base, identifying the owners and IP addresses of all domains, but it is not automatically managed.
- RIPE NCC supports the infrastructure of the Internet and provides global Internet resources and related services (IPv4, IPv6 and AS Number resources) in Europe, Middle East and parts of Central Asia.

Real AS-Level Internet Data



Numbers of Internet AS (Nov. 1997 – Feb. 2002)

(Siganos et al., 2003)

Power-Law Node-Degree Distributions

Faloutsos M, Faloutsos P, Faloutsos C.
On power-law relationships of the Internet topology.
ACM SIGCOMM Computer Communication Review, 1999, 29(4): 251-262

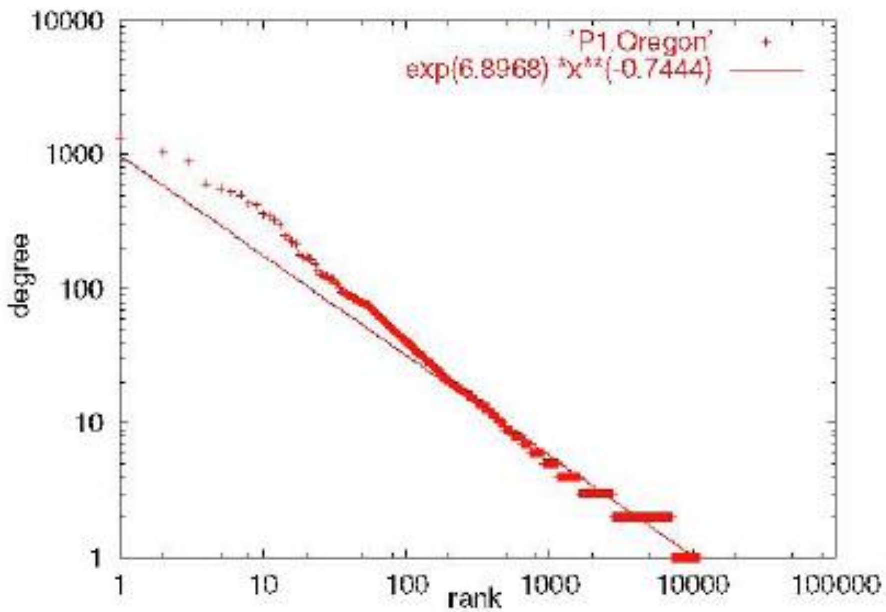
UC

UT

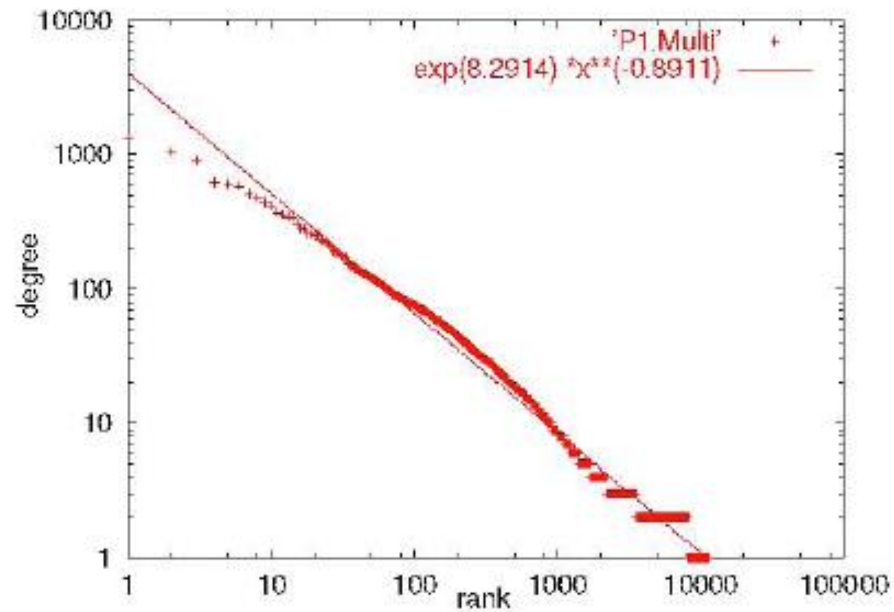
CMU

The Faloutsos Brothers

Power-Law Node-Degree Distribution I



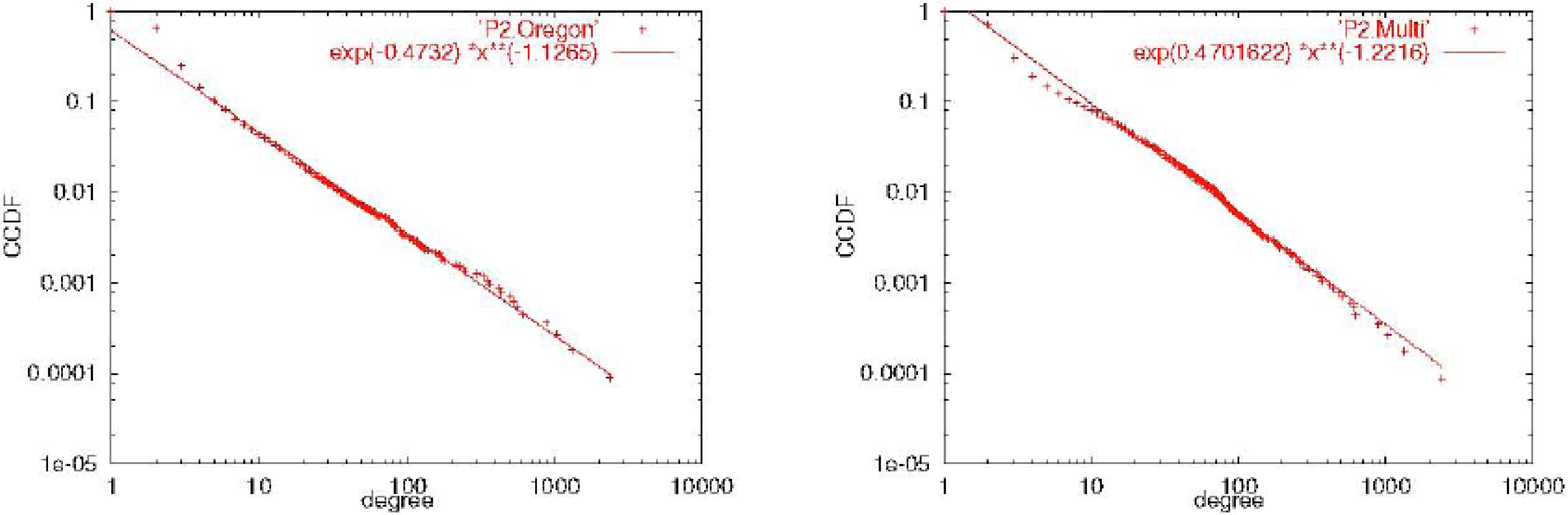
(a) Oregon



(b) Multi

Power law I: $d_v \sim r_v^R$, where d_v is the degree of node v , r_v is the index of node v in the degree-decreasing ordering of all nodes, and $R<0$ is a rank constant exponent.

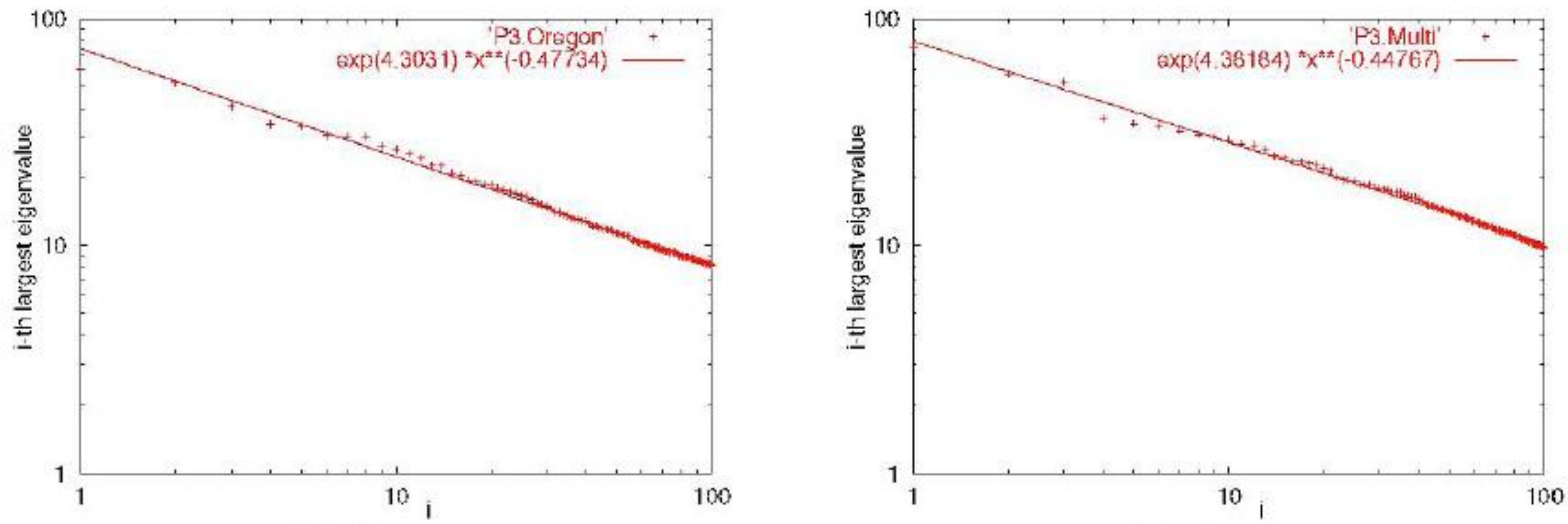
Power-Law Node-Degree Distribution II



(a) Oregon (b) Multi

Power law II: $D_d \sim d^D$, where D_d is the percentage of node with degrees larger than d , and D is a degree constant exponent, satisfying $D=1/R<0$

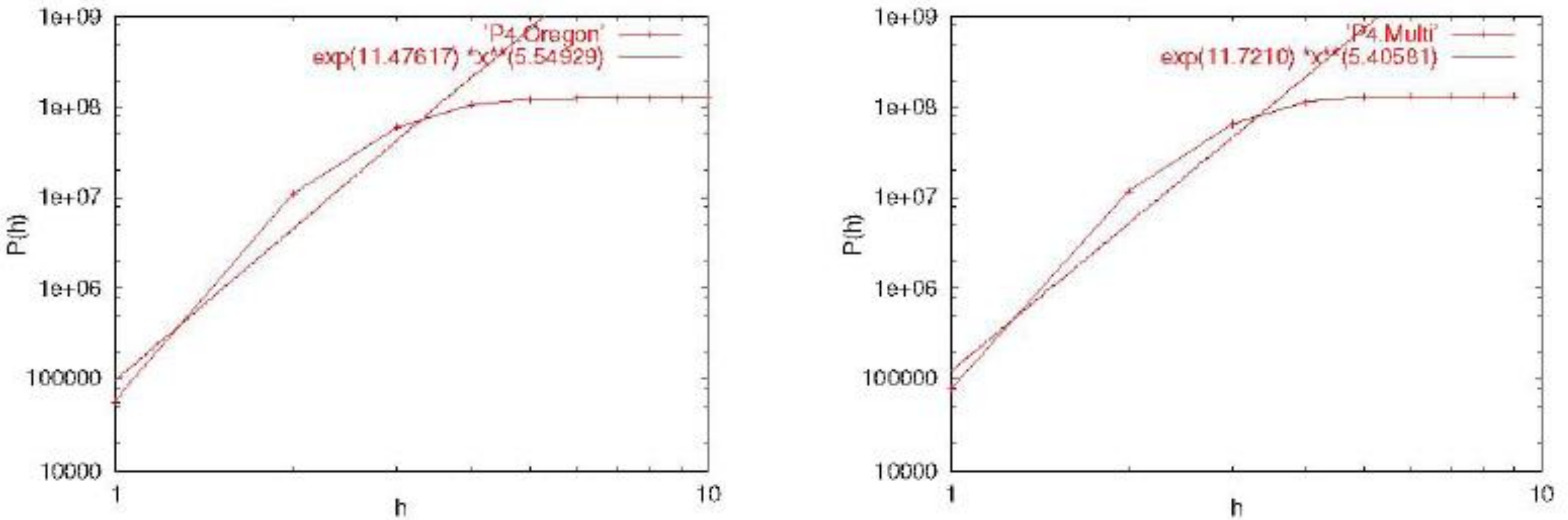
Power-Law Node-Degree Distribution III



(a) Oregon (b) Multi

Power law III: $\lambda_i \sim i^{\varepsilon}$, where λ_i is the i-th eigenvalue in decreasing order of the network connectivity matrix, and ε is the characteristic constant exponent, satisfying $\varepsilon \approx D/2 < 0$

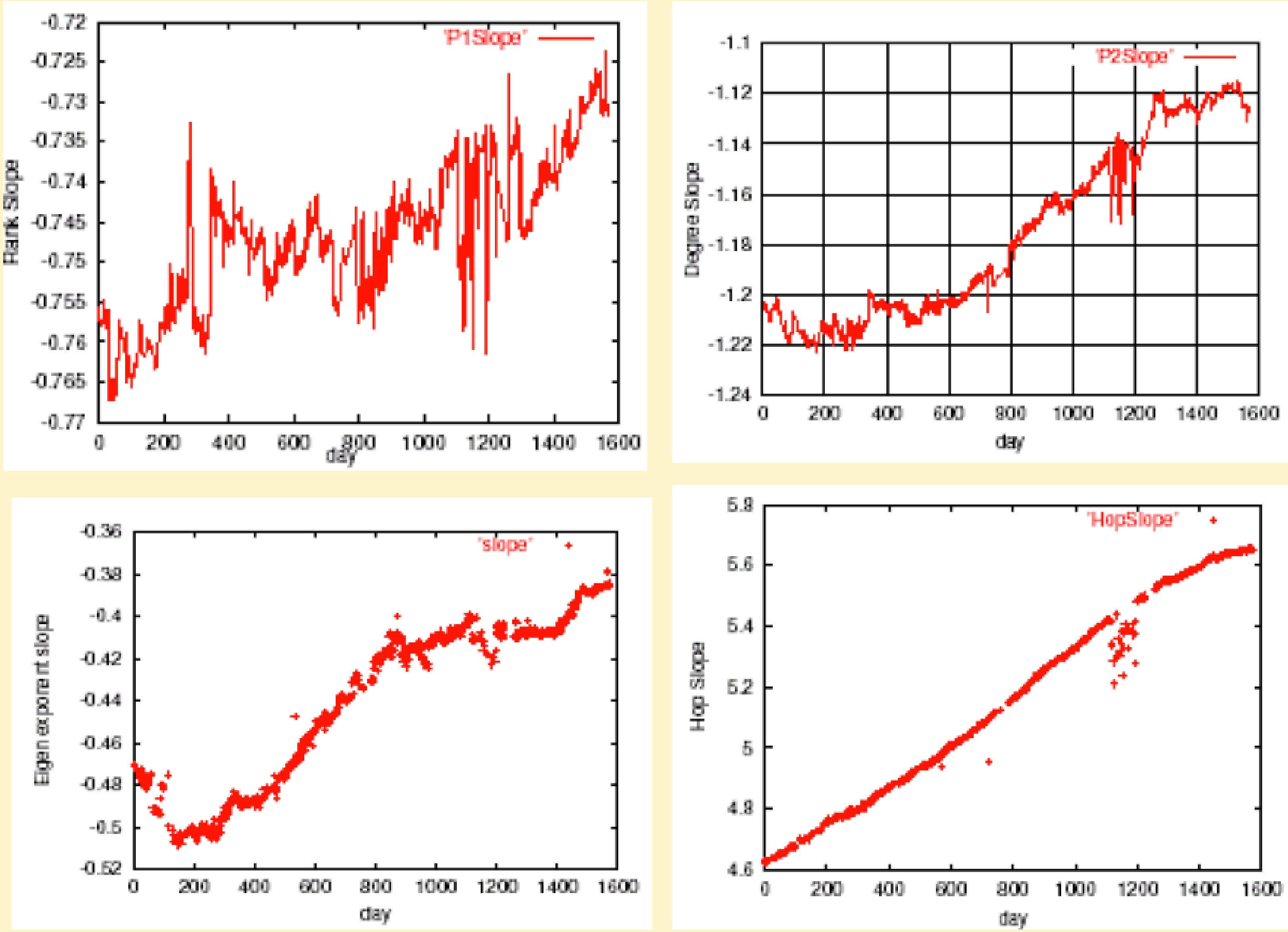
Power-Law Node-Degree Distribution IV



(a) Oregon (b) Multi

Power law IV: $P(h) \sim h^H$, where $P(h)$ is the number of node pairs of distance not larger than h, and H is a hop constant exponent.

Evolution of the four power-law distributions

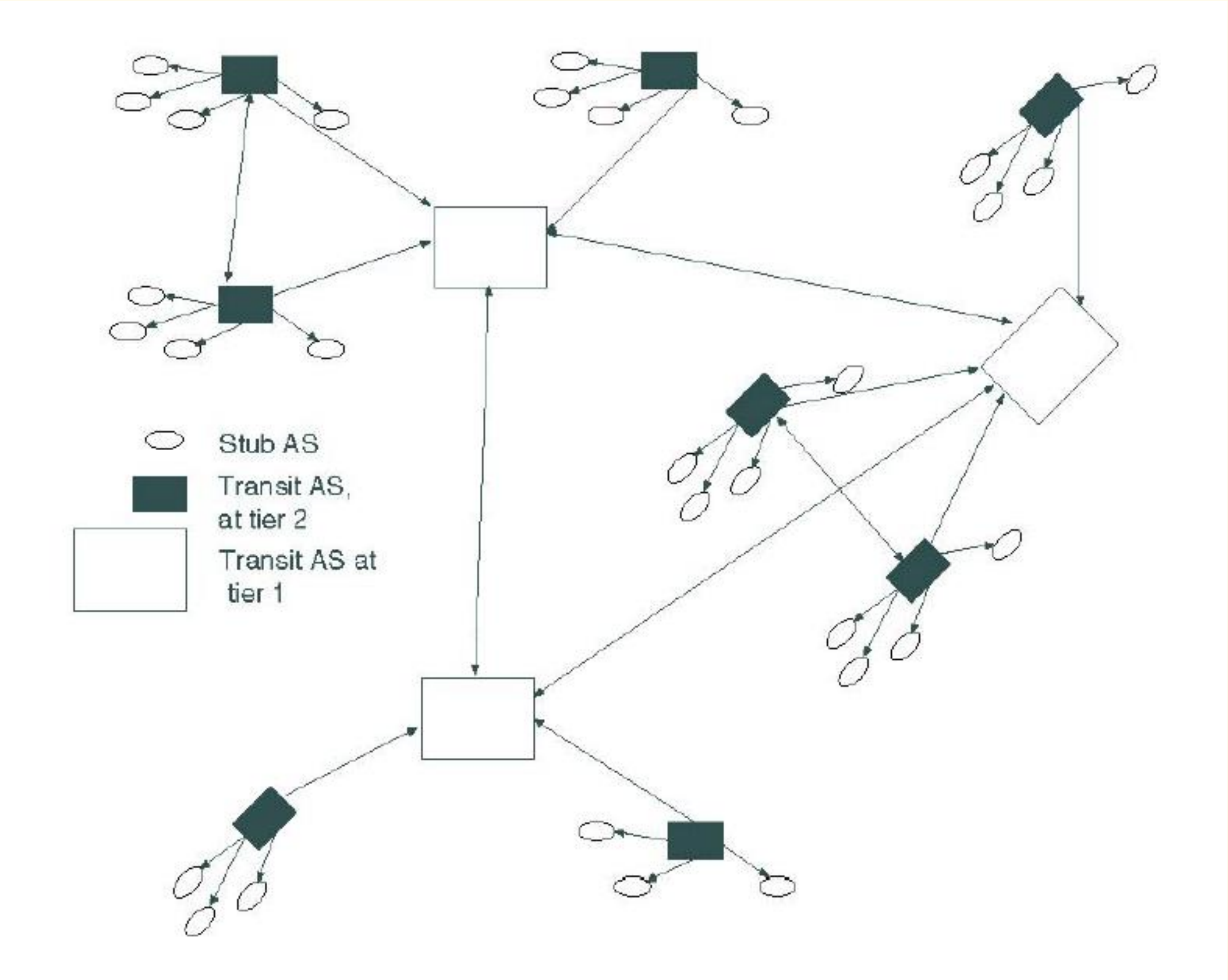


The evolution of the exponents R, D, ε and H, respectively, during Nov. 1997 to Feb. 2002.

Internet Hierarchical Structures

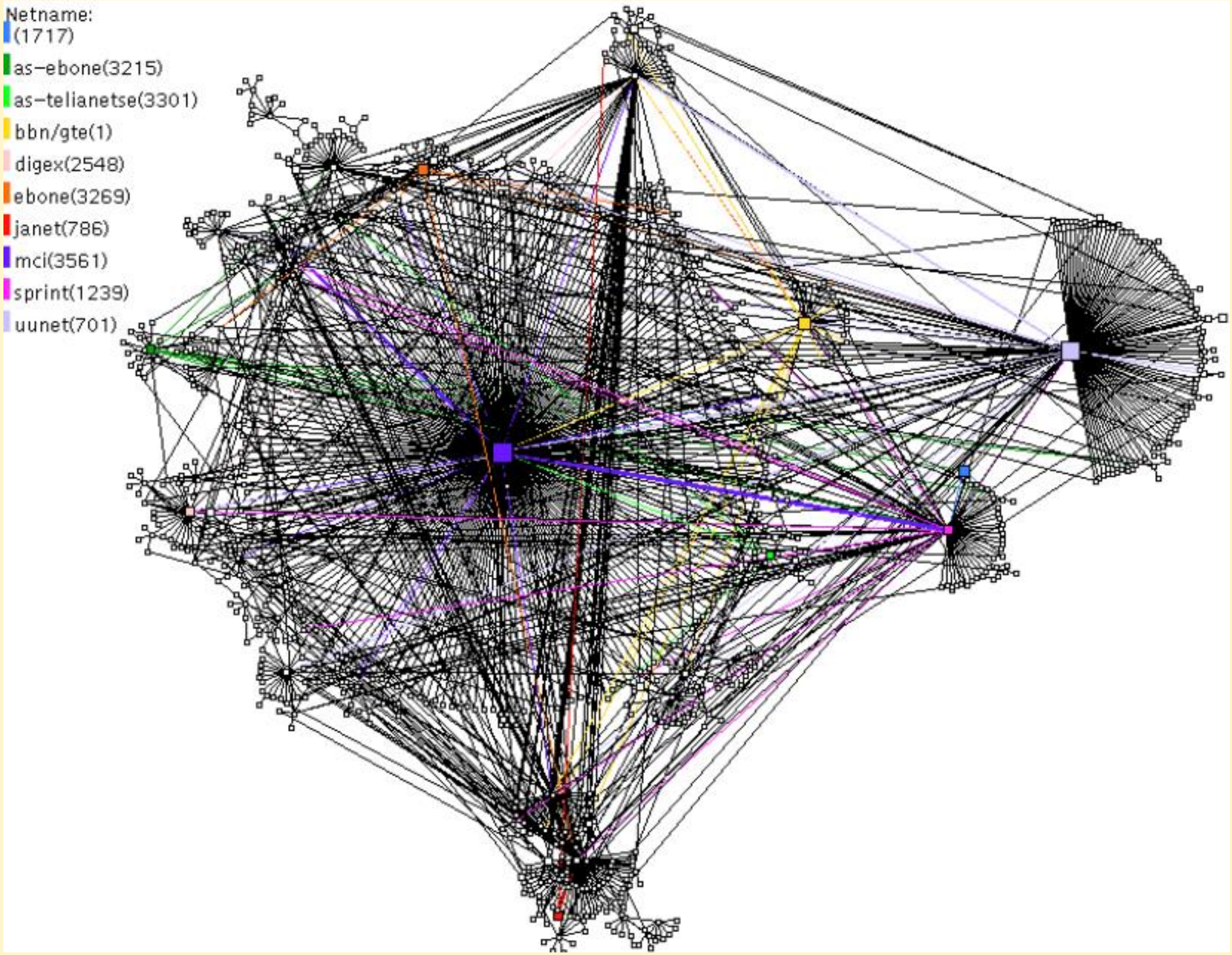
- Internet consists of a large number of interconnected AS
- AS may be considered as a *Stub domain* or a *Transit domain*
- Transit domain can be a Metropolitan Area Network (MAN) or a Wide Area Network (WAN), typically a regional or even a national Internet Service Provider (ISP)
- Stub domain consists of campus networks or some other interconnected Local Area Networks (LAN)
- Transit domain is used to link many nearby Stub-domains together
- Stub domain usually only processes the information starting and ending inside the domain, while a Transit domain has no such restriction

Internet Hierarchical Structures



AS-level Internet structure (Jaiswal et al., 2004)

Internet Hierarchical Structures



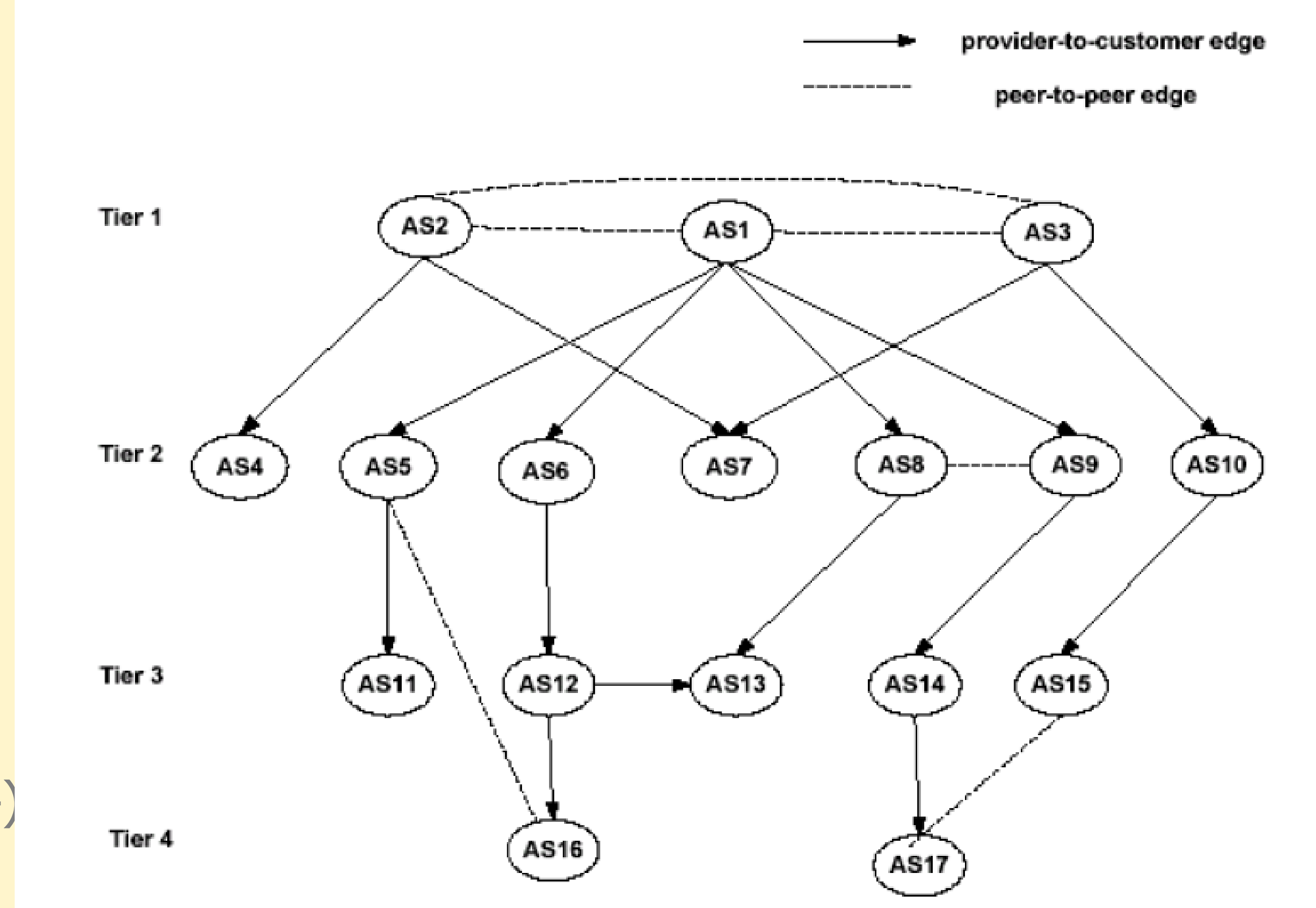
AS-level Internet on 3 December 1998 (generated by Skitter)

Internet Hierarchical Structures

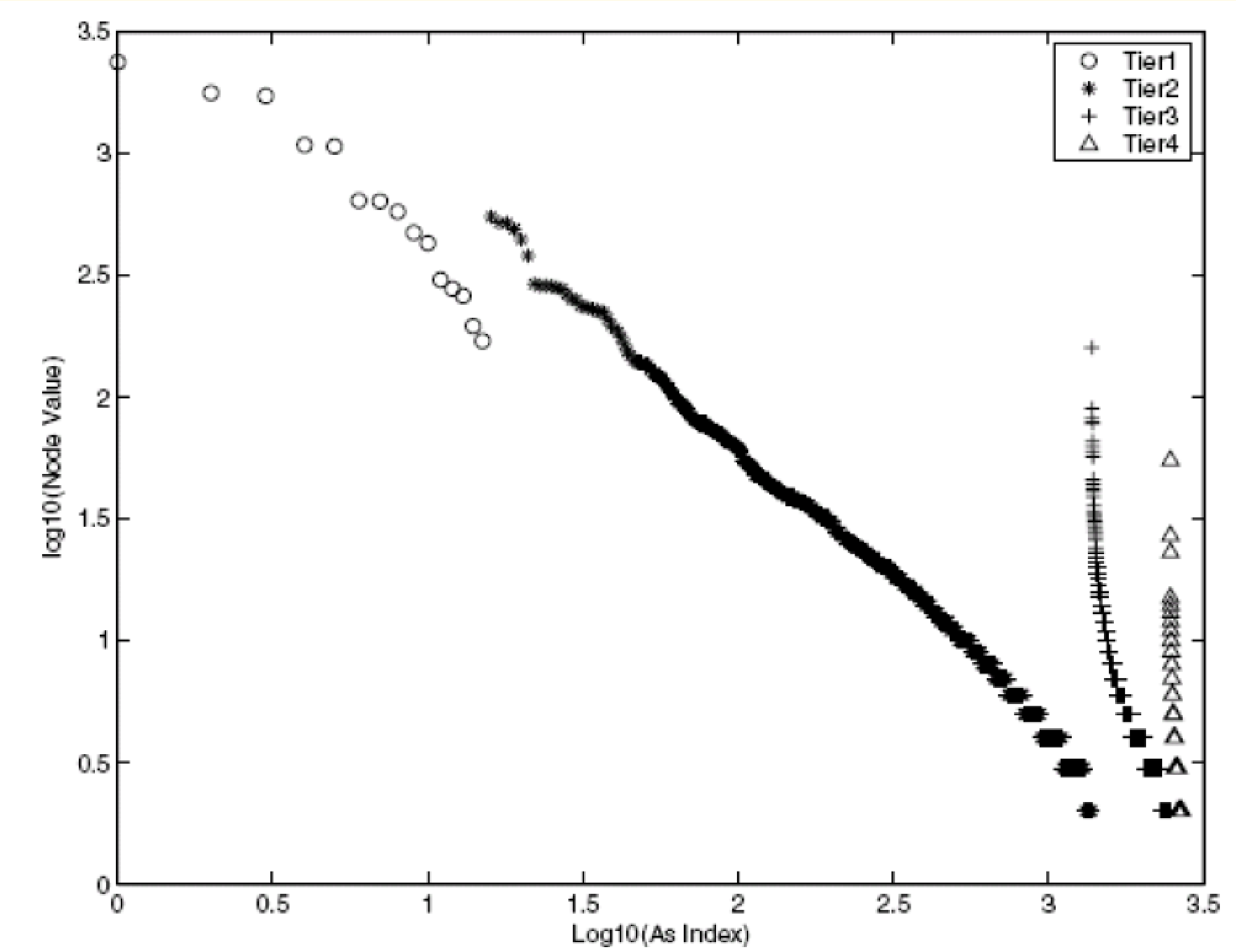
- AS on the Internet can be considered as some kind of *Tier*
- An AS at the highest Tier belongs to the Transit domain, called *Tier-1 provider*

➢ Those Transit and Stub domains at a *lower Tier* depend on the Transit nodes at a *higher Tier* to communicate with the other domains at their same level

(Cai et al., 2004)



Internet Hierarchical Structures

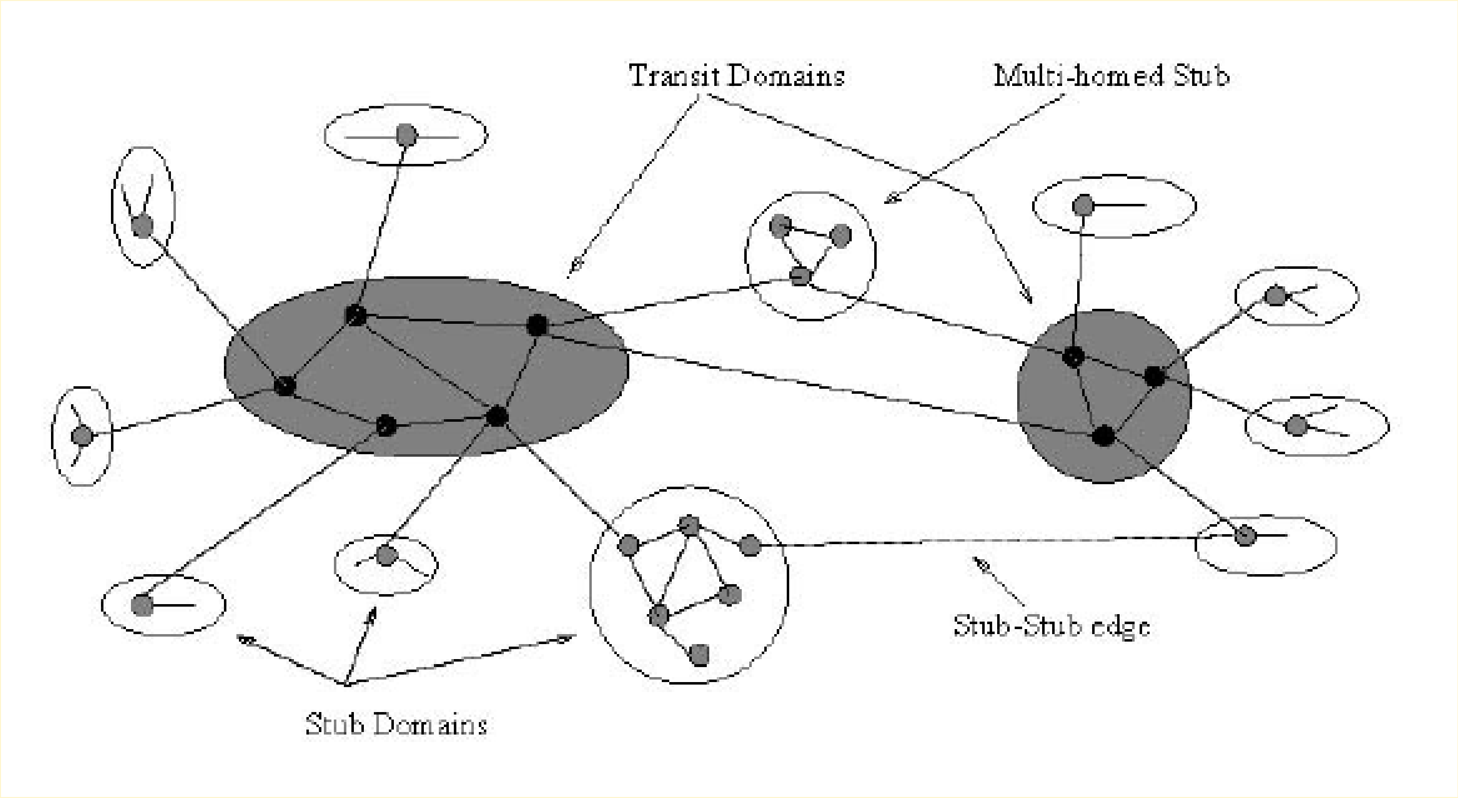


Degree distribution of the AS-level Internet at different tiers

(Jaiswal et al., 2004)

Rich-Club Phenomenon

In the Internet, a few nodes have a large number of edges, called **hubs**, and they tend to connect to each other – Rich-club phenomenon



Rich-club phenomenon (Zegura et al., 1997)

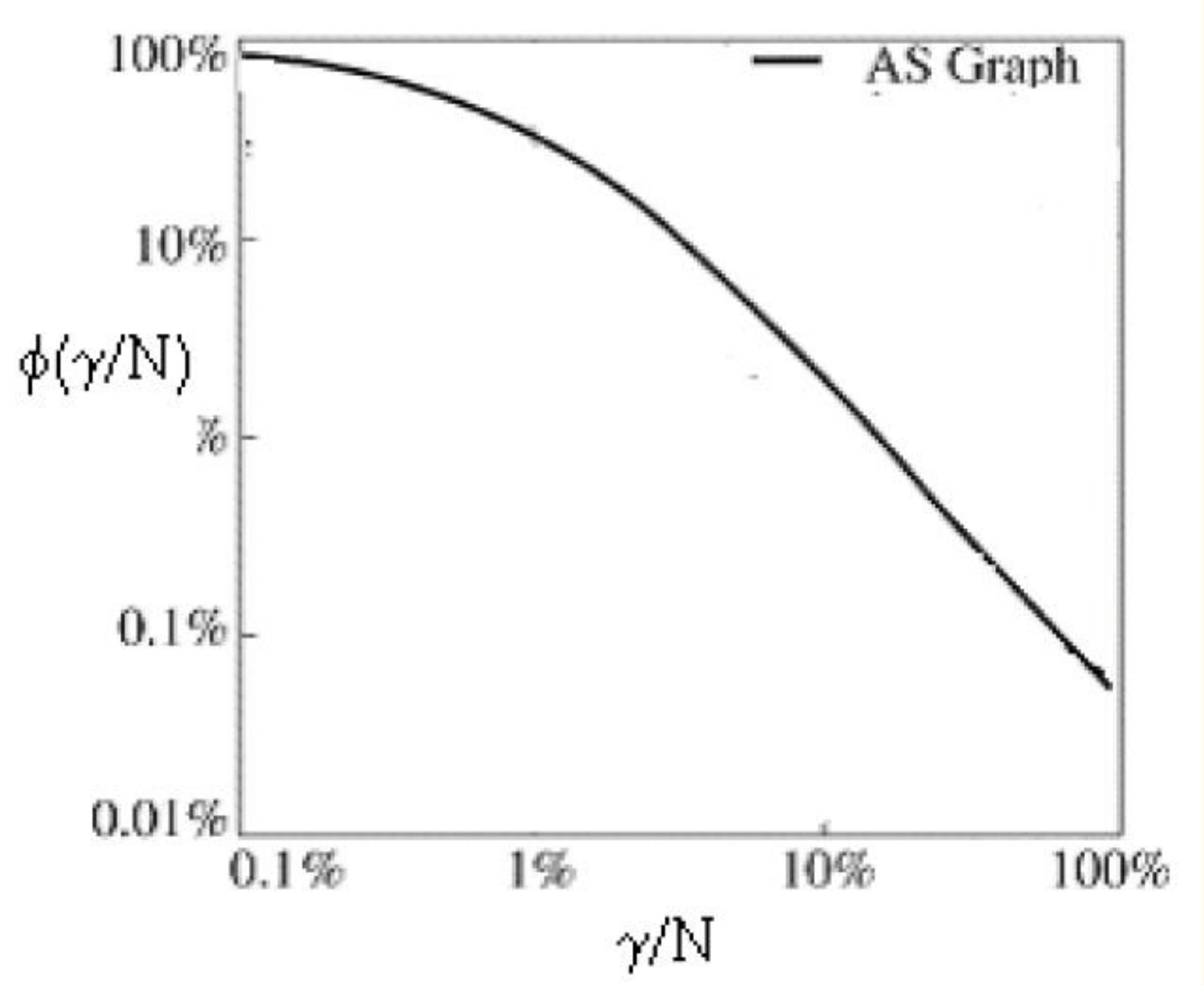
Rich-Club Phenomenon

- The rich-club phenomenon in an AS layer of size N can be described by the connectivity index $\phi(r/N)$ of its first r biggest nodes defined by the ratio of the number L of their existing edges versus the number $r(r\text{-}1)/2$ of all possible edges among them:

$$\Phi(r/N) = \frac{L}{r(r-1)/2} = \frac{2L}{r(r-1)}$$

- If $\phi(r/N) = 1$, then the first r biggest nodes compose a fully connected sub-network.
- Some real data have verified that the connectivity index follows nearly a power-law form.

Rich-Club Phenomenon



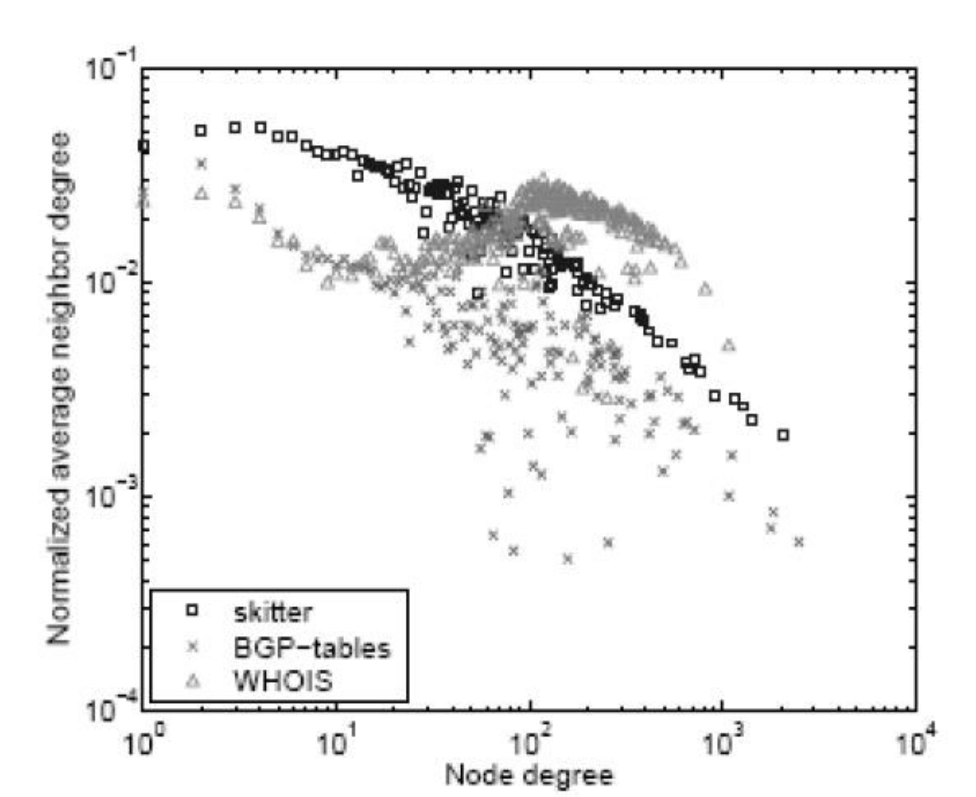
Rich-club phenomenon in the AS-level Internet (Zegura et al., 1997)

Yet, most nodes are small in the Internet

- In the Internet, hubs are well interconnected (rich clubs).
- Most neighbors of a hub typically have small degrees.

- Analysis on the Internet data of April 2002 shows that nodes with degrees of 1, 2, and 3 were 26%, 38% and 14%, respectively, which sums up to about 80% of the whole network.

(Zhou and Mondragon, 2004)



(Mahadevan et al., 2005)

(Dis)assortative Properties

Definition. The *Assortativity Coefficient* of a network is defined by

$$r = \frac{M^{-1}\sum_i j_i k_i - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}$$

where k_i and j_i are the degrees of the end nodes of edge i, and M is the total number of edges in the network.

If $r>0$ then the network is *assortative* (big-big nodes);

if $r<0$ then the network is *disassortative* (big-small nodes).

Analysis on real data using BGP (Border Gateway Protocol), Skitter and Whois about the Internet at some AS levels shows that their assortativity coefficients of the Internet are – 0.19, – 0.24 and –0.04, respectively, implying that Internet is overall <u>dis</u>assortative.

Coreness

Definition.

■ The *k-core* in a graph is defined to be the remaining sub-graph after all the nodes with degrees $\leq k\text{-}1$ have been removed successively, during which:

(i) when a node is removed, all its adjacent edges will also be removed;

(ii) after a node of degree $\leq k\text{-}1$ is removed, in the remaining graph all the remaining nodes with a new degree $\leq k\text{-}1$ also need to be removed.

- If a node belongs to a *k-core* of a graph, but it will be removed from the *(k+1)-core*, then this node is said to have coreness (core value) k.

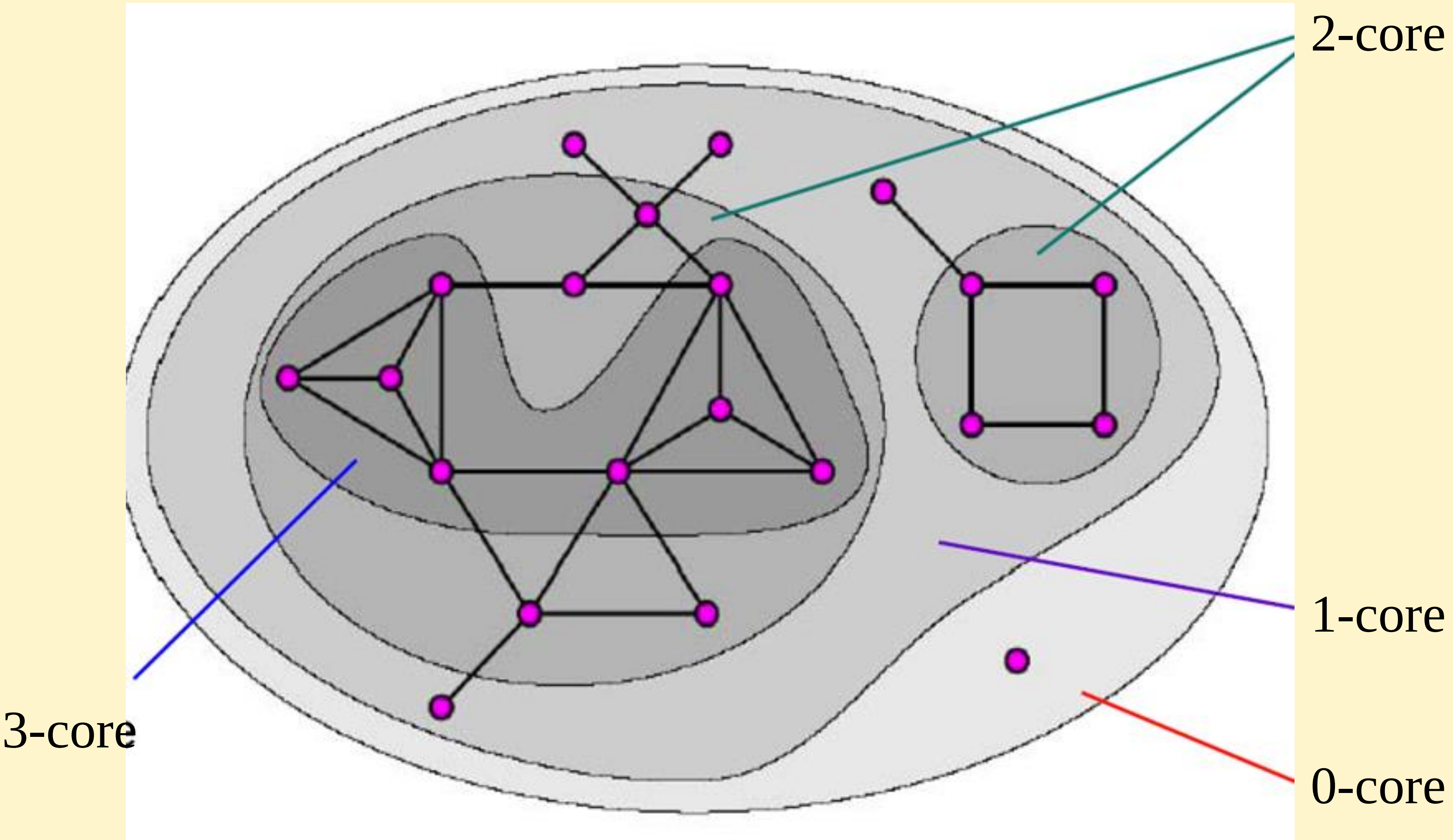
- The largest coreness in a graph is called the coreness of the graph.

Examples:

An isolated node has coreness $k=0$

A fully-connected network of size N has coreness $k = N - 1$

Coreness



Coreness

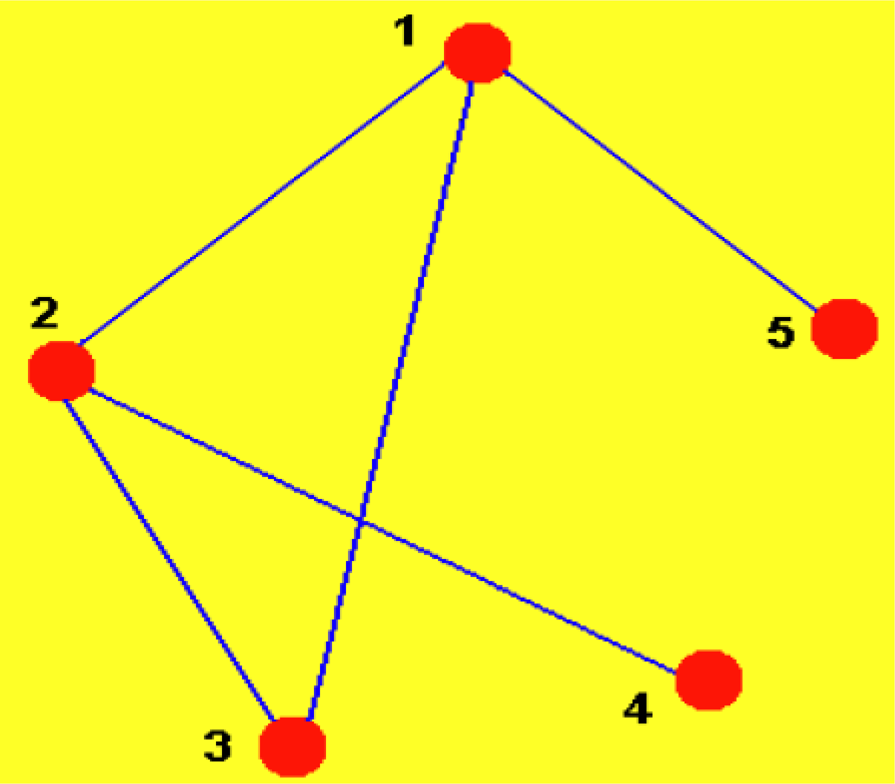
Star-shaped networks

- the 1-core of the network is the network itself
- all nodes, including the central node, have coreness 1
- the coreness of the network is 1

Ring networks

- the 1-core of the network is the network itself
- the 2-core of the network is the network itself
- all nodes have coreness 2
- the coreness of the network is 2

Coreness

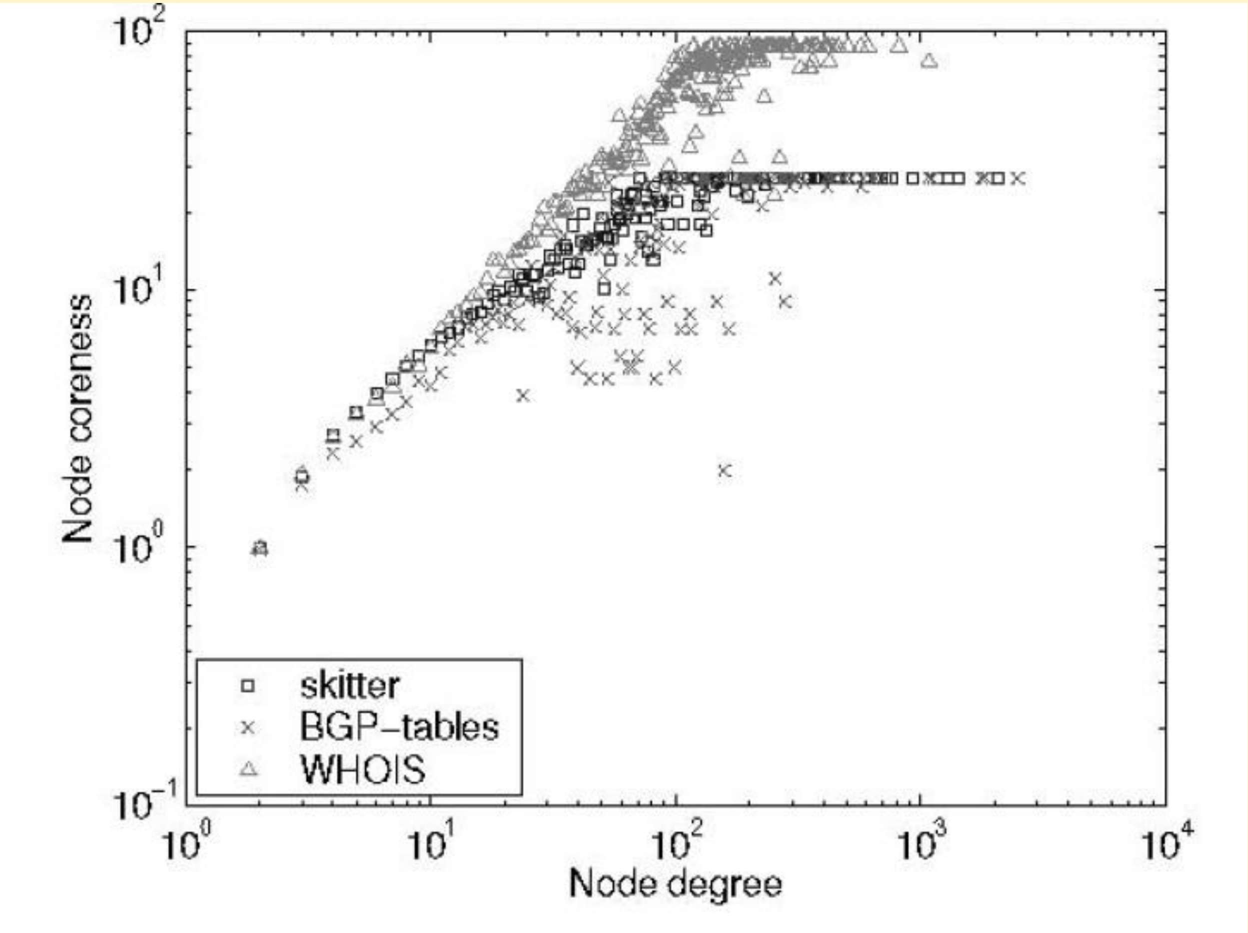


Meaning here:
Preserving 2-core
= Cutting nodes of coreness 1
= Removing tree-like subnets

Examples:

- 1-core is the network itself
- 2-core is Triangle 1-2-3
- Node 1, Node 2, Node 3 have coreness 2
- Node 4 and Node 5 have coreness 1
- Coreness of the graph is 2

Coreness



Internet data from Skitter, BGP and Whois: (Mahadevan et al., 2005)

- ➢ when the node degree is relatively small, they have a power-law relation, with exponents 0.58, 0.68 and 1.07, respectively;
- ➢ when the node-degree is larger than 100, their coreness values become saturated.

Why we define Coreness ?

- The main purpose of introducing the concept of coreness is to reflect the fact that a higher core is more important than a lower core, and a higher-coreness node is more important than a lower-coreness node, in a network. This can reveal the hierarchical structure of a network, where higher cores and higher-coreness nodes belong to higher-levels of the hierarchical network. Clearly, the star-shaped and ring-shaped networks do not have prominent hierarchical structures.

- The main implication of the concept of coreness is that a network with a higher coreness will have better robustness against intentional attacks. Apparently, both star-shaped and ring-shaped networks are fragile to intentional attacks.

Betweenness

Definition 1 (Node-betweennes) In a network of size N, the *node-betweenness* of node i is defined by

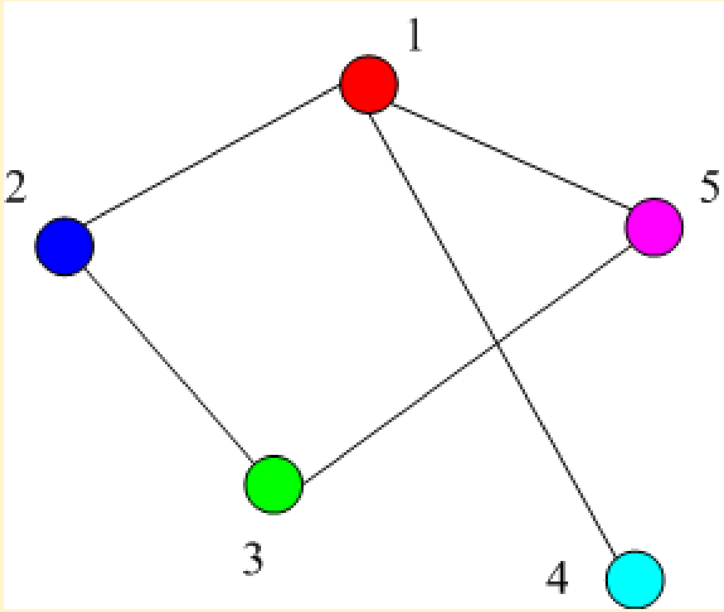
$$B(i) = \sum_{i \neq j \neq l} \frac{L_{jl}(i)}{L_{jl}}$$

where L_{jl} is the number of all existing shortest paths from node j to node l, and $L_{jl}(i)$ is the number of all shortest paths from node j to node l that pass through node i.

(since $i \neq j \neq l$ the node i itself is excluded)

Normalization: divided by the total number of nodes not including i: $(N\text{-}1)(N\text{-}2)/2$

Node-Betweenness



Example

betweenness of node 1 is

$$B(1)=\frac{(5,1,4)}{(5,1,4)}+\frac{0}{(5,3)}+\frac{(5,1,2)}{(5,1,2)+(5,3,2)}+\frac{(4,1,2,3)+(4,1,5,3)}{(4,1,2,3)+(4,1,5,3)}+\frac{(4,1,2)}{(4,1,2)}+\frac{0}{(3,2)}$$

$$=\frac{1}{1}+\frac{1}{2}+\frac{2}{2}+\frac{1}{1}=\frac{7}{2}$$

Normalized: $$B(1)=\frac{7/2}{(N-1)(N-2)/2}=\frac{7}{12}$$

Betweenness

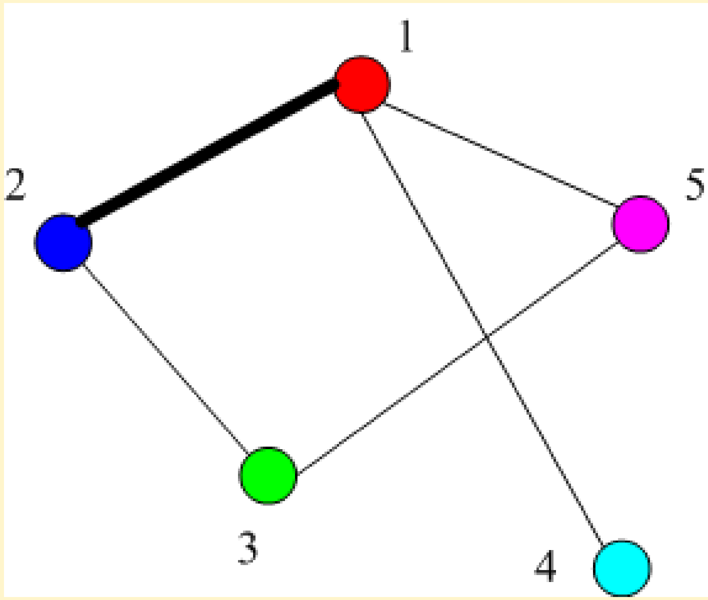
Definition 2 (Edge-betweennes)
In a network of size N, the *edge-betweenness* of edge e_{ij} is defined by

$$B(e_{ij}) = \sum_{(l,q)\neq(i,j)} \frac{\tilde{L}_{lq}(e_{ij})}{\tilde{L}_{lq}}$$

where $\tilde{L}_{lq}$ is the number of all existing shortest paths from node l to node q, and $\tilde{L}_{lq}(e_{ij})$ is the number of all shortest paths from node l to node q that pass through edge e_{ij}.

Normalization: divided by the total number of edges not including e_{ij} namely, divided by $N(N-1)/2 - 1$

Edge-Betweenness



Example

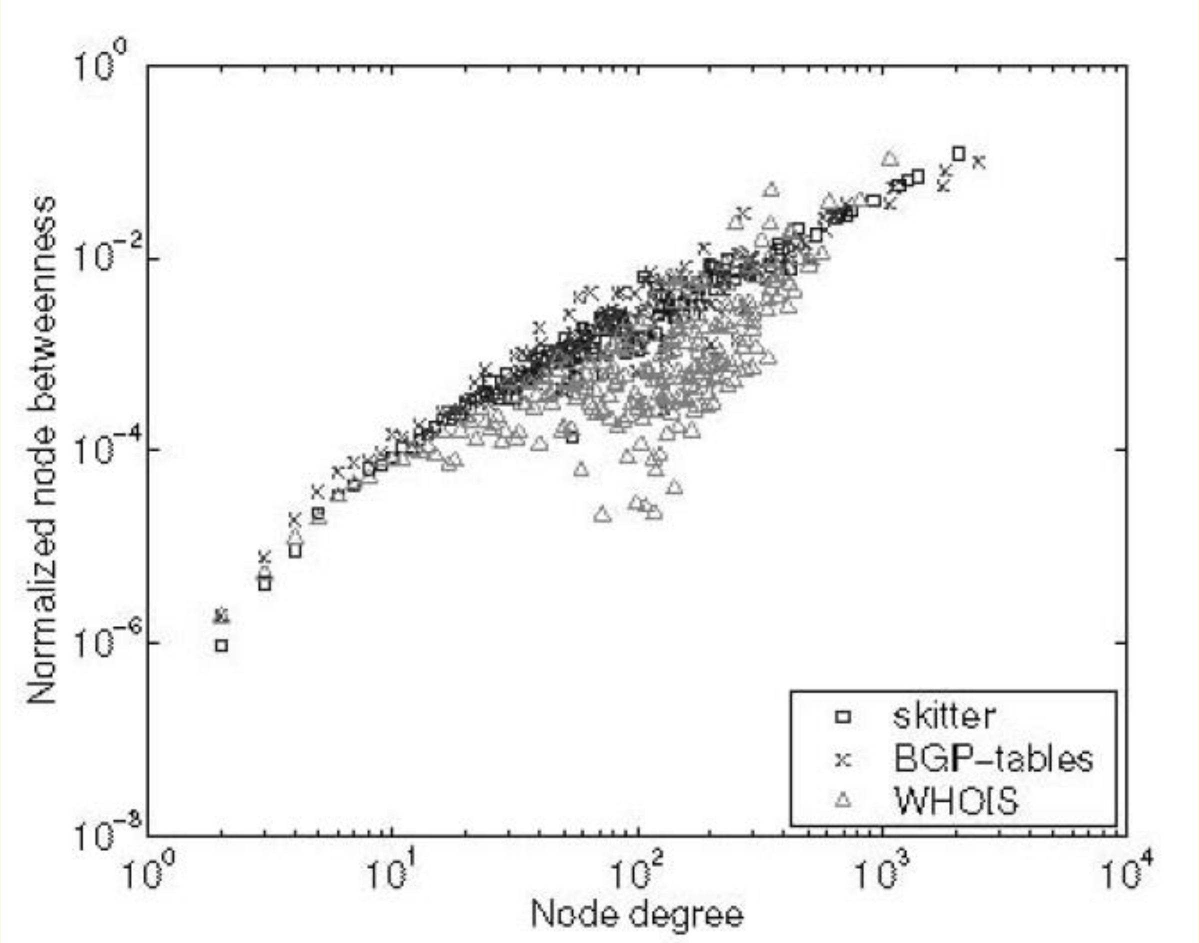
betweenness of edge e_{12} is

$$B(e_{12}) = \frac{0}{(5,1,4)} + \frac{0}{(5,3)} + \frac{(5,1,2)}{(5,1,2)+(5,3,2)} + \frac{(4,1,2,3)}{(4,1,2,3)+(4,1,5,3)} + \frac{(4,1,2)}{(4,1,2)} + \frac{0}{(3,2)} + \frac{(3,2,1)}{(3,2,1)+(3,5,1)}$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{1}{1} + \frac{1}{2} = \frac{5}{2}$$

Normalized: $B(e_{12}) = \frac{5/2}{N(N-1)/2-1} = \frac{5}{18}$

Betweenness

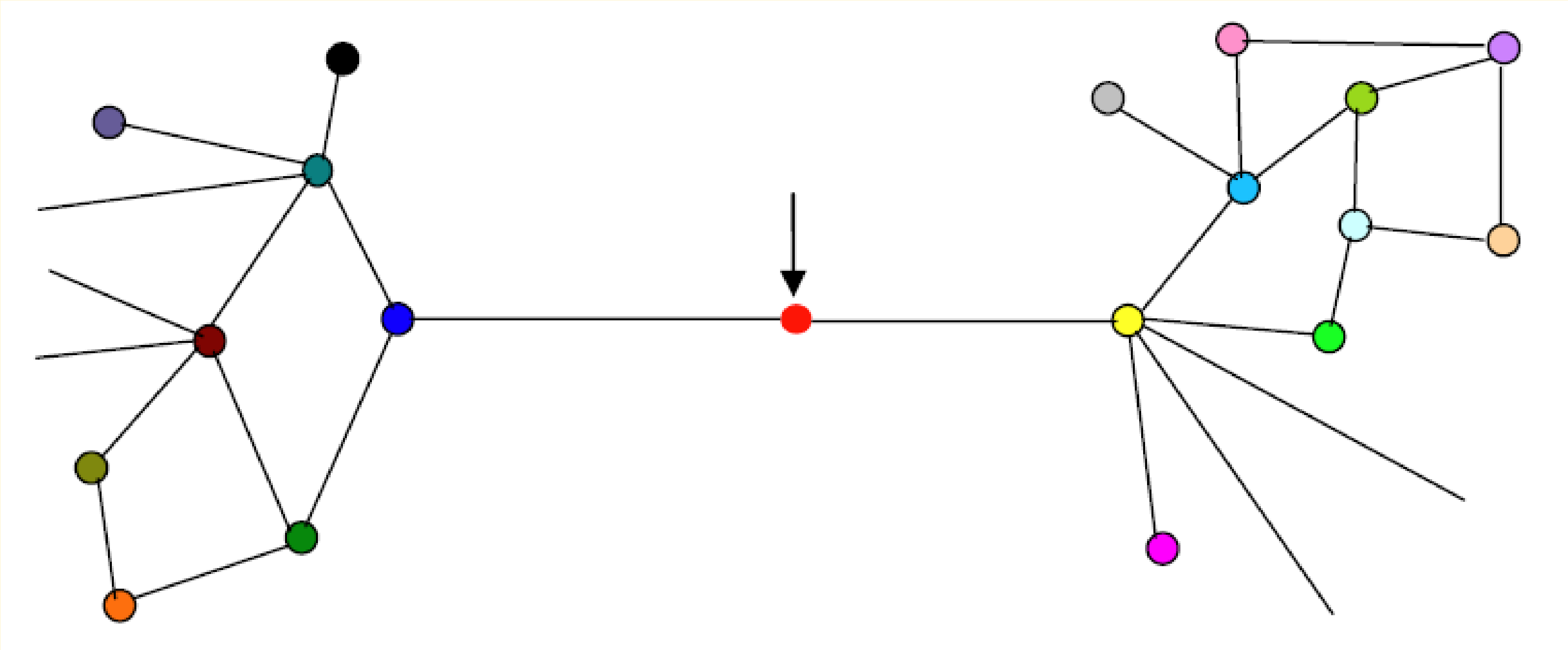


Internet data from Skitter, BGP and Whois: (Mahadevan et al., 2005)

➢ For those from Skitter and BGP, their relations follow prominent power-law distributions with components 1.35 and 1.17, respectively

Why we define Betweenness ?

➢ The importance of the node- and edge-betweenness centralities can be easily understood from the following figure, where the bridging node has very small degree yet very large node- and edge-betweeness, while generally big nodes have large betweenness.



The importance of a node with a large betweenness value

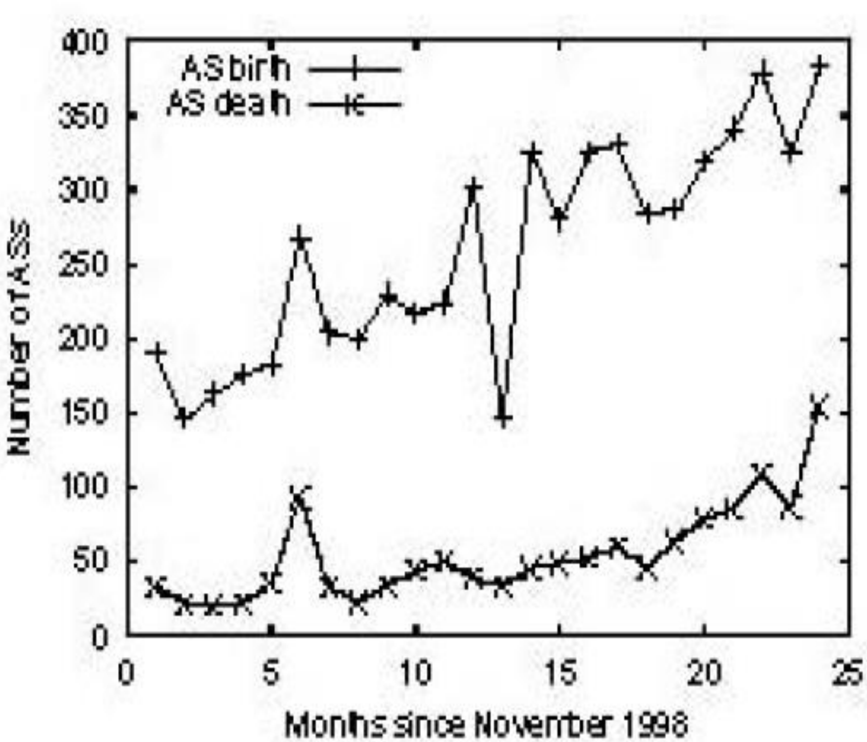
Growth of the Internet

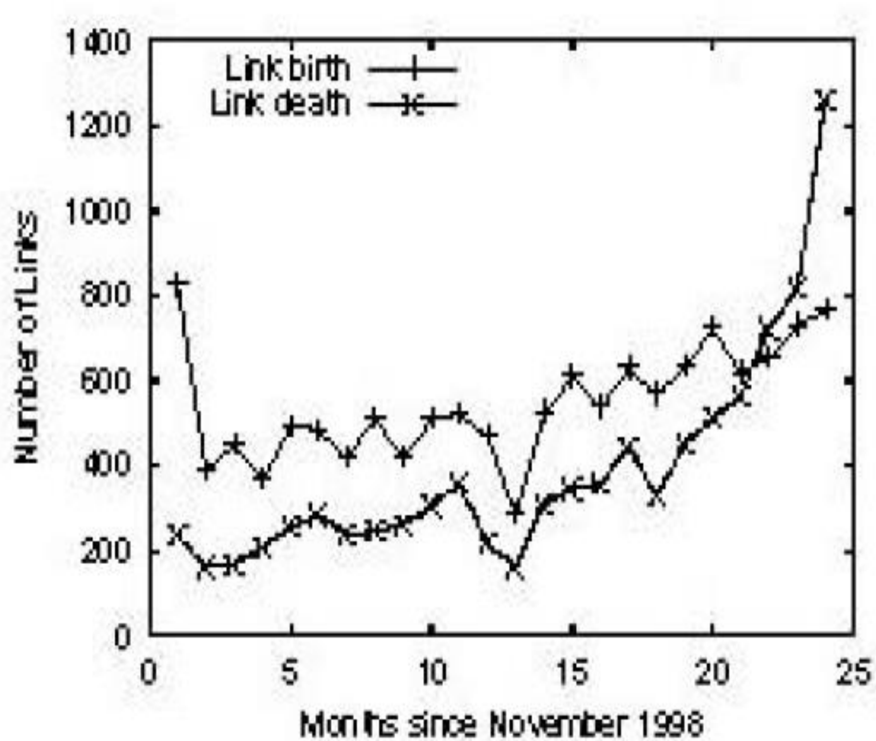
- The Internet is actually a dynamically evolving complex network, which is rapidly and continuously growing and restructuring

- Some historical data of the Internet will be examined, which were provided by

  - the SCAN Project with the software named Mercator
  - the Oregon router services (at the AS level)
  - the Topology Project of the Computer Science Department of the Michigan University (at the Extended AS level)

Growth of the Internet

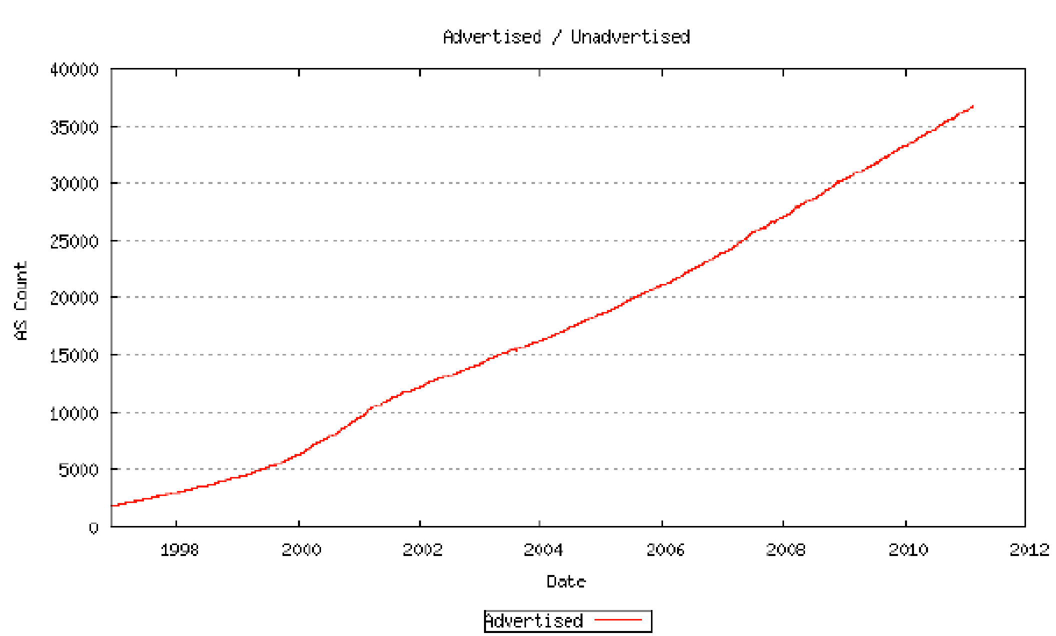




Nodes (Qin et al., 2002) Edges

- Monthly numbers of birth and death of the Internet at the AS level
- Real data collected from Oregon servers, the Looking Glass site, and the Internet Routing Registry (IRR)

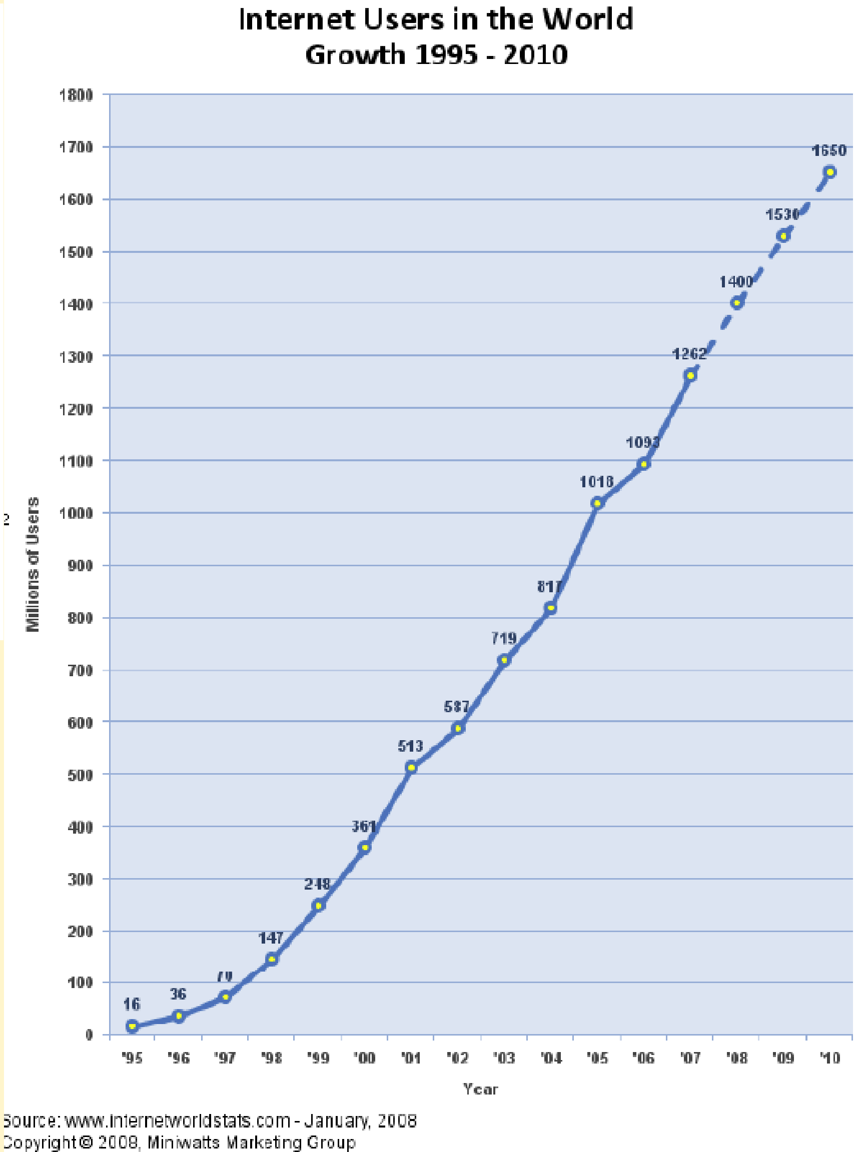
Growth of the Internet



AS Numbers

http://www.potaroo.net/tools/asn32/

Users Numbers →



Features of the Internet

Table 4-1 Some average measures of the Internet at the IR, AS, and EAS levels

Level	N	E	$\langle k \rangle$	C	L	B
IR	228,263	320,149	2.8	0.03	9.5	5.3
AS	11,174	23,409	4.2	0.30	3.6	2.3
EAS	11,461	32,730	5.7	0.35	3.6	2.3

IR -- Internet Router
AS -- Autonomous System
EAS -- Extended Autonomous System

N -- number of nodes
$\langle k \rangle$ -- average node degree
L -- average shortest path lengths

E -- number of edges
C -- average clustering coefficient
B -- average node-betweenness

(Pastor-Satorras and Vespignani, 2004)

Node Evolution of the Internet

Table 4-2 Total numbers of additions and deletions of the Internet at the AS level

Degree	Number of Additions	Number of Deletions
1	5,591	1,184
2	816	204
3	23	22
4	4	6
5	1	4
6	1	1
7	1	1
9	0	1
10	1	0
11	1	0
12	0	1
14	1	0
48	0	1

(Qin et al., 2002)

edge Evolution of the Internet

Table 4-3 Monthly rate of new edges connecting old nodes to new and old ones

Year	1998	1999
$E_{n,o}$	170	231
$E_{o,o}$	350	450
$E_{n,o} / E_{o,o}$	0.48	0.53

$E_{n,o}$ -- numbers of new edge additions between incoming nodes and existing nodes

$E_{o,o}$ -- numbers of new edge additions between two existing nodes

(Vazquez et al., 2002)