Modeling of Complex Networks

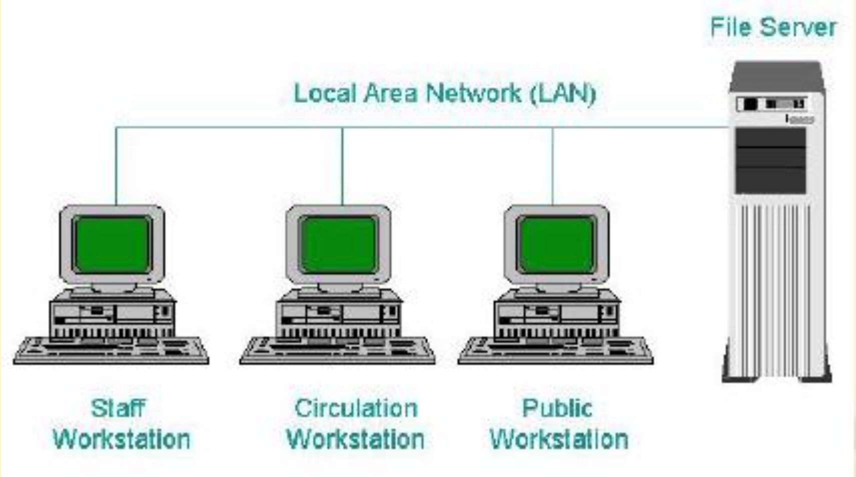
Lecture 4: Internet
--Topology and Modeling

S8101003Q-01(Sem A, Fall 2019)

Instructor: Aaron, Haijun Zhang

Network Topology Modeling

- Graph representations
- AS-level:
  - nodes are domains (AS)
  - edges are peering relationships
- Router-level:
  - nodes are routers
  - edges are one-hop IP connections
- PC-level: not manageable today
  - nodes are PCs
  - edges are optical fibers



Representative Models

- Waxman (Waxman 1988)
  Router-level model capturing locality
  --- --- ---
- Transit-Stub (Zegura 1997), Tiers (Doar 1997)
  Router level model capturing hierarchy
  --- --- ---
- Inet (Jin 2000)
  AS-level model based on degree sequence
- BRITE (Medina 2000)
  AS-level model based on evolution
- BA-Model (Barabasi-Albert 1999-2000)
  AS-level model based on degree sequence and evolution
  - ➢ HOT (CalTech 2004-2005)
    Heuristic Optimized Tradeoffs
  - ➢ MLW (Fan-Chen, 2007-2010)
    Multi-Local-Worlds

Router-Level Internet Topology

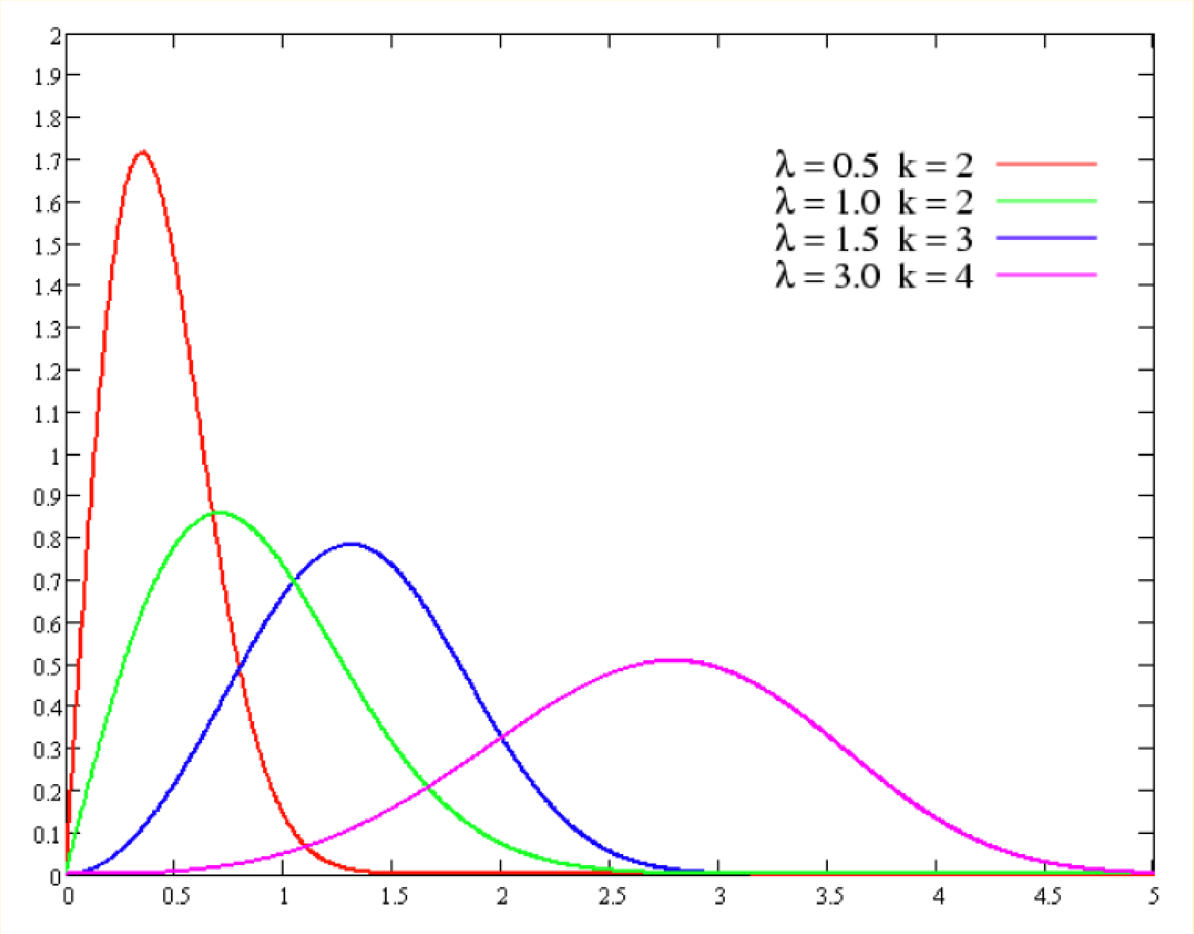
- A common software tool to represent the router-level Internet topology by a graph is the *traceroute* (Unix traceroute or Windows NT tracert.exe, free download), or its IPv6 version *traceroute6(8)*
- The *traceroute* uses hop-limited probe, which consists of a hop-limited IP (Internet Protocol) packet and the corresponding ICMP (Internet Control Message Protocol) response, to probe every possible IP address and record every reached router and the corresponding edges.

Router-Level Internet Topology

- Some analysis on the real data collected during October-November of 1999 shows that in the router-level of the Internet topology:
  - Basically, does not have hierarchical structure
  - power-law node-distribution is not prominent but Weibull distribution seems better, yet the latter can only reflect Transit but not Stub subnets
- Some analytical results on the real data collected during December 2001 -- January 2002 show that the Weibull distribution can better fit the complementary cumulative distribution function of router out-degree than the Pareto and power-law distributions

Weibull Distribution

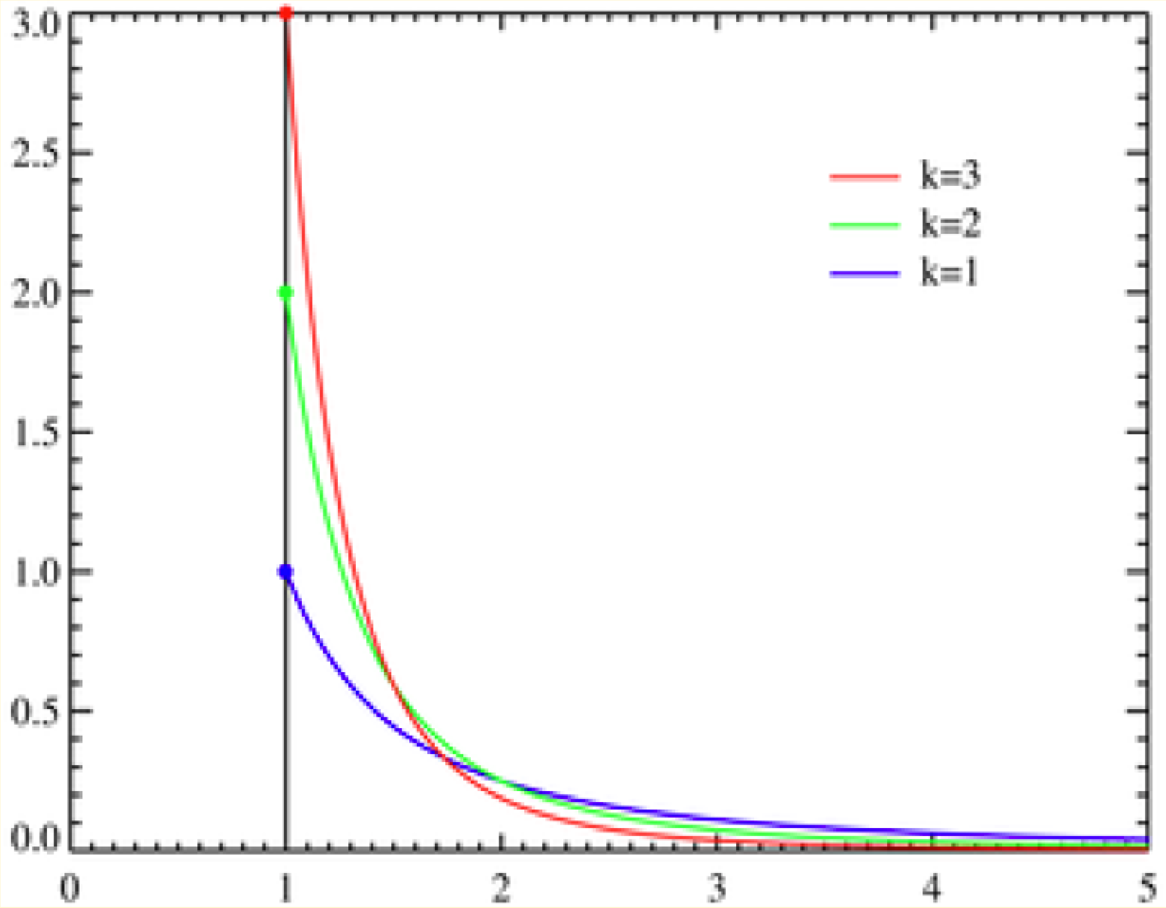
$$f_k(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$$



x

k and λ are constant parameters

Pareto Distribution



$$f_k(x) = k\frac{\lambda^k}{x^{k+1}}, \quad x \geq \lambda > 0$$

k and λ are constant parameters

x

First Generation of Internet Topology Models

1980s

Waxman Model

Waxman modeling algorithm:

■ Start with N nodes, randomly placed on a lattice, one in each small square.

■ Each step, for every pair of two nodes, u and v, and then connect them by an edge according to the following probability (called Waxman probability):

$$P(u,v) = \alpha e^{-d(u,v)/(\beta L_{\max})}$$

where $d(u,v)$ is the distance, α is the average number of edges, L_{max} is the longest distance, β is a parameter determined by the average path length, with $0<\alpha, \beta \leq 1$

Waxman Network Model

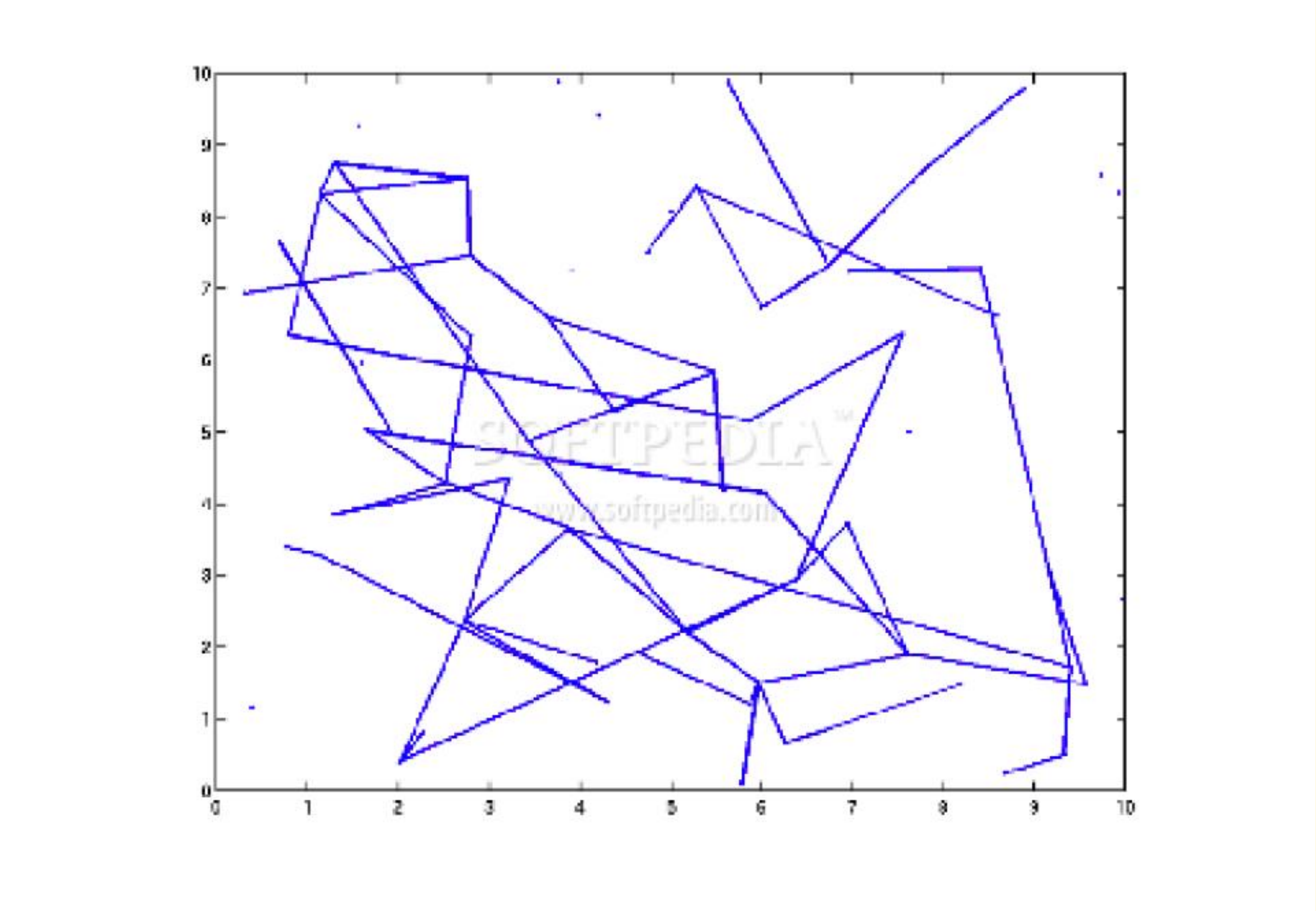
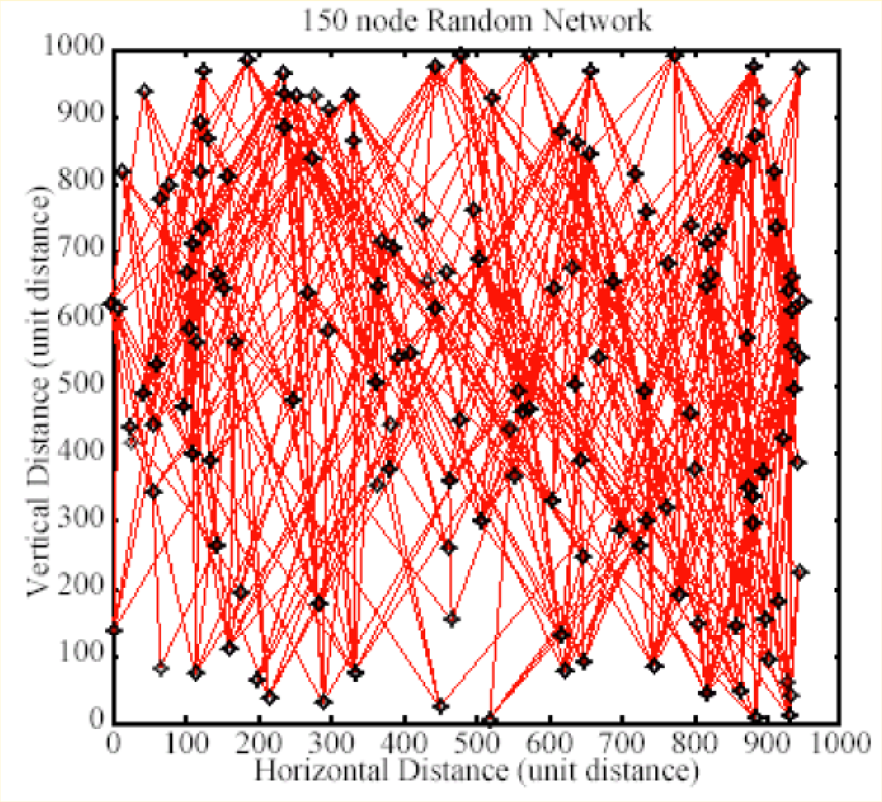


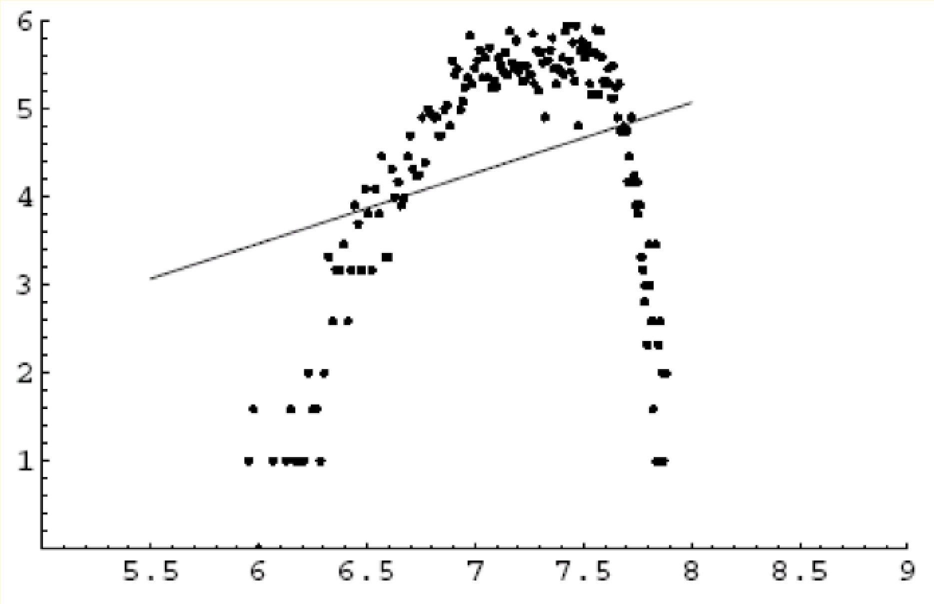
Illustration of a generated network

Waxman Model



$N = 150, \alpha=0.25, \beta=0.3$

(Waxman, 1988)



Degree distribution (~Weibull)

(Medina et al., 2000)

Second Generation of Internet Topology Models

1990s

Transit-Stub Topology

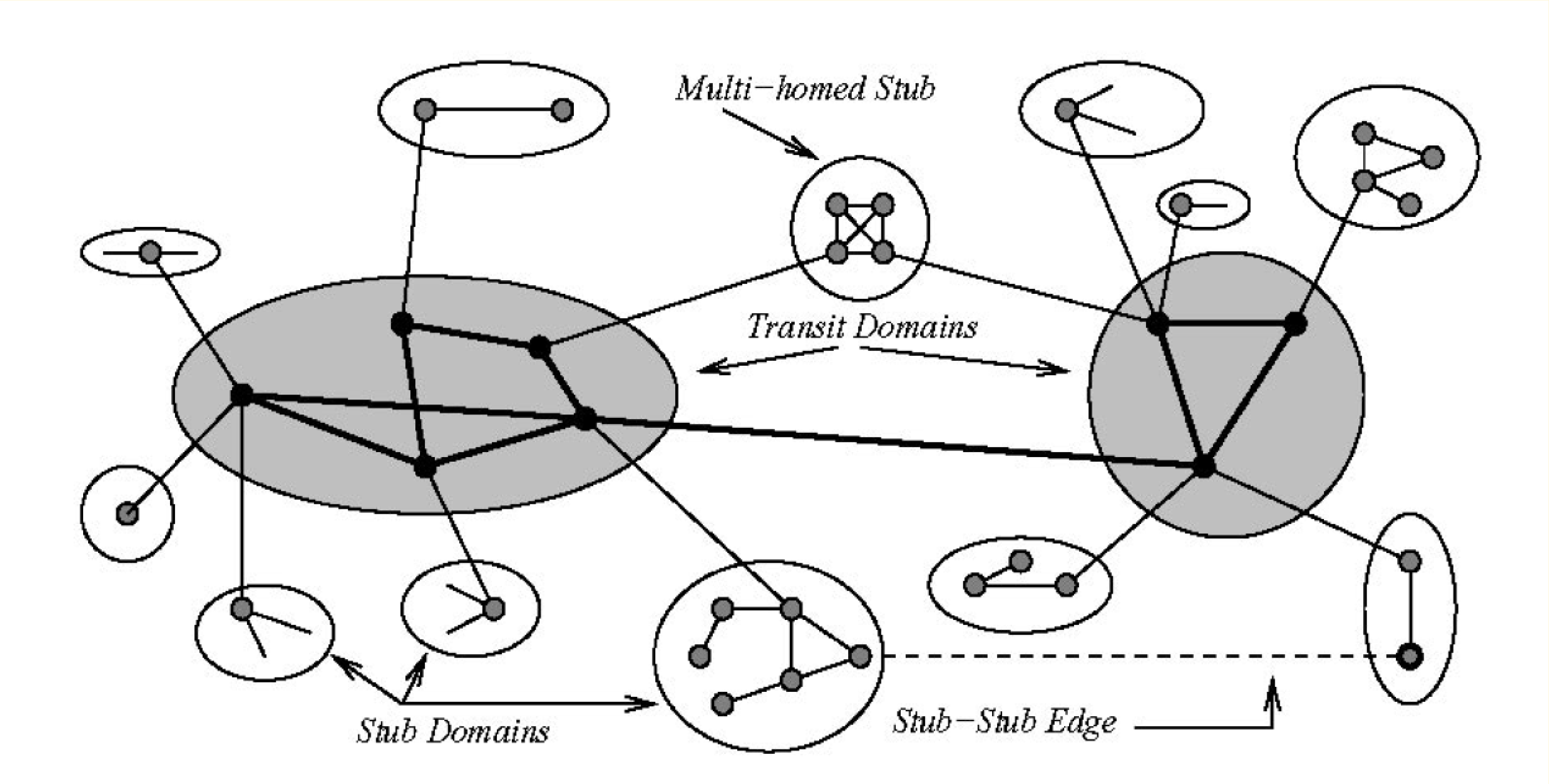
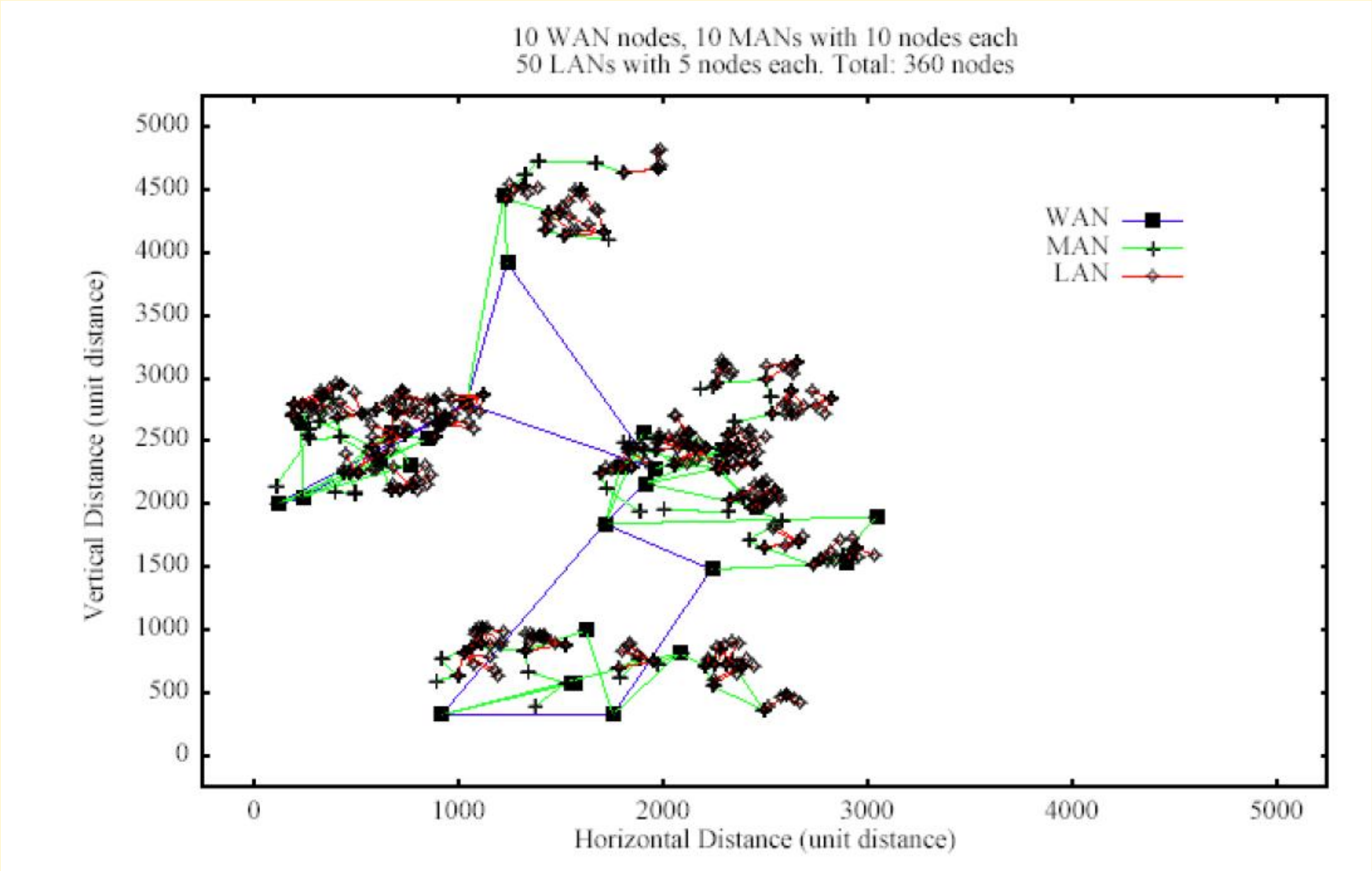


Illustration of network structure from Transit-Stub topology generator

Transit-Stub Topology Generator

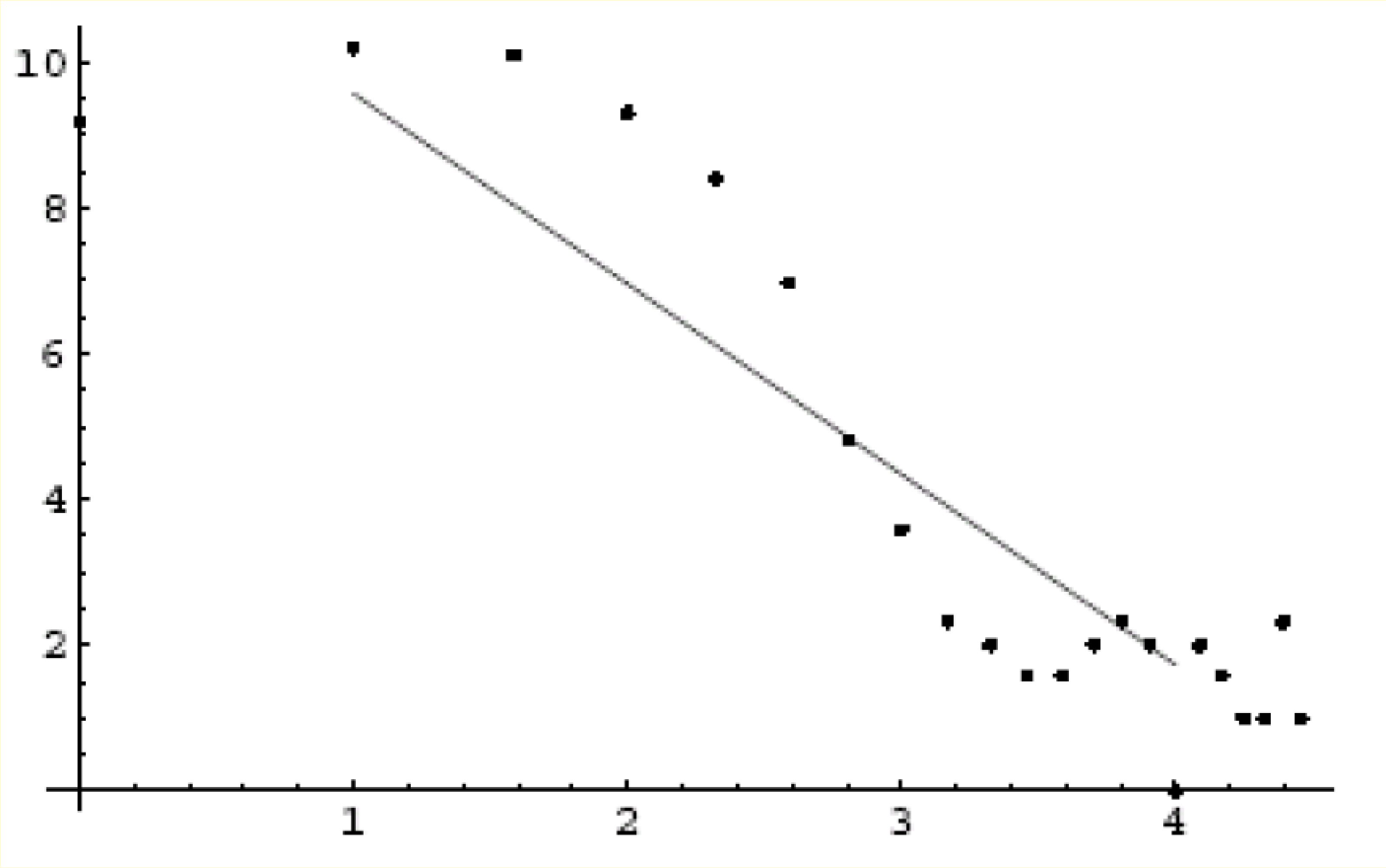
- Software
- Generate all Transit domains
  - Use a random-graph generation method (e.g., the Waxman algorithm), where each node represents a Transit domain.
  - Generate nodes in each Transit domain by adding some nodes around the Transit point, and then connect these nodes with edges at random.
- Generate Stubs for each Transit:
  - This is similar to the above Transit-domain generation, but at a lower level.
  - Connect every Stub domain to a Transit domain: Randomly select one node from a Stub domain and then connect this node to the Transit domain by an edge.
- Generate LANs for each Stub:
  - This is similar to the above Transit-Stub generation, but at the lowest level.
  - They all have star-shaped structures.
  - Connect each LAN to a Stub domain.

Transit-Stub Topology Generator



A typical Transit-Stub topology (Calvert, 1997)

Transit-Stub Topology Generator



Out-degree distribution of a Transit-Stub network with 6660 nodes
(Medina et al., 2000)

Third Generation of Internet Topology Models

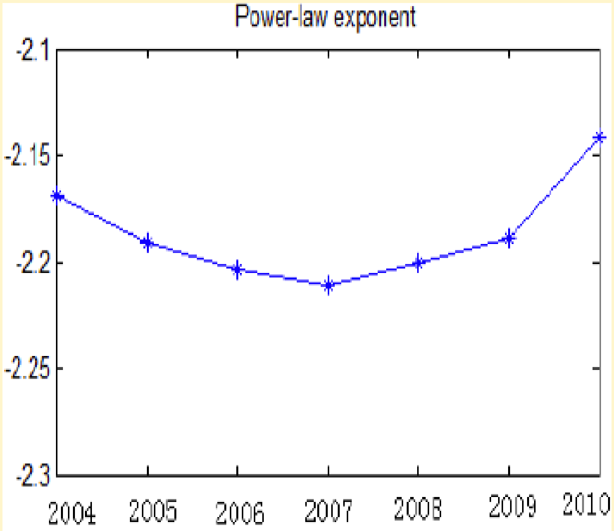
2000s

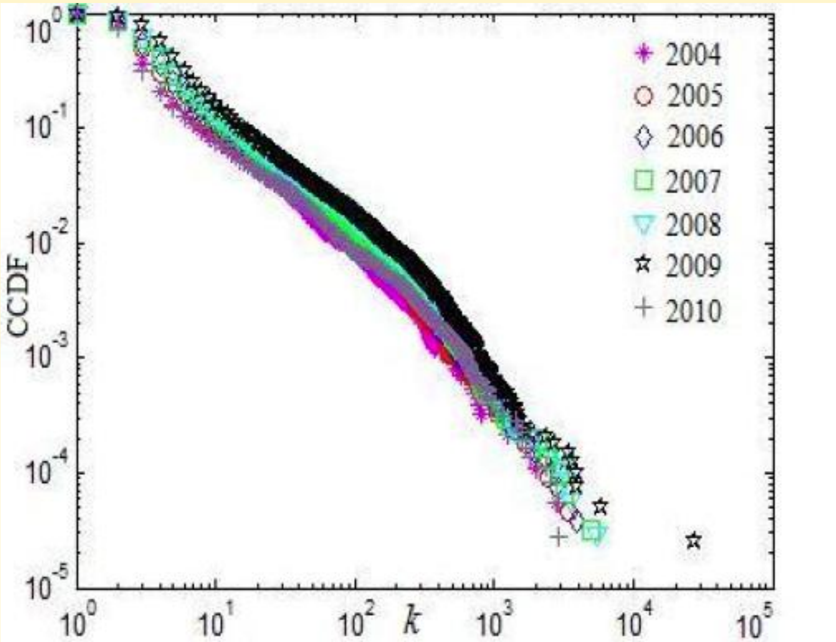
Inet

- Router-level model and AS-level model
- Input:
  - Total number of nodes
  - Percentage of degree-one nodes
- Degree sequence: power-law

$$P(k) \propto k^{-\gamma}$$

$$\gamma = 2.14 \sim 2.21$$





CAIDA

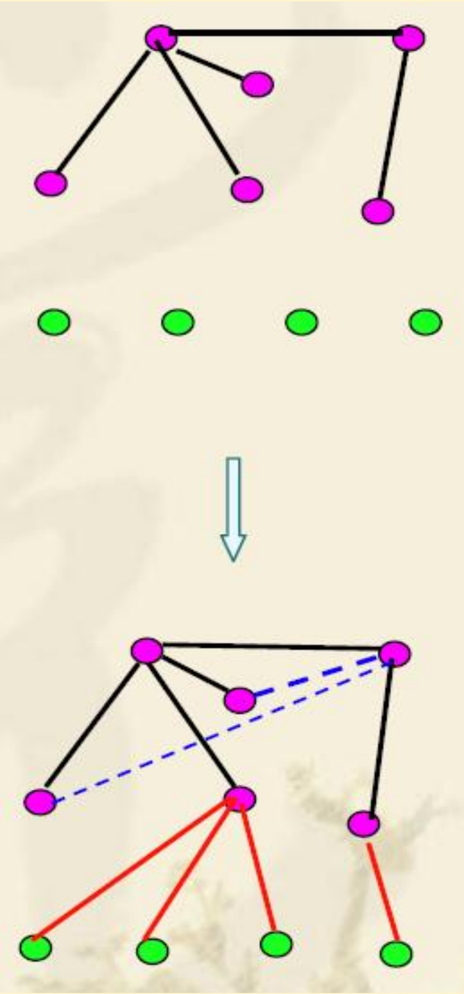
Inet

- From the real data set, let V_1 be the set of all degree-1 nodes, typically has about *30%* of the total (green nodes). Let the rest be V_2 (pink nodes).
- Generate a spanning tree consisting of nodes from V_2

To generate the spanning tree in a network G, start from empty initial conditions, and then a node i is connected to a node j , both in V_2, according to the following (preferential attachment) probability:

$$\Pi(i, j) = \frac{w_i^j}{\sum_{k \in G} w_i^k} \qquad w_i^j \text{ --weigt (reverse distance) from } i \text{ to } j$$

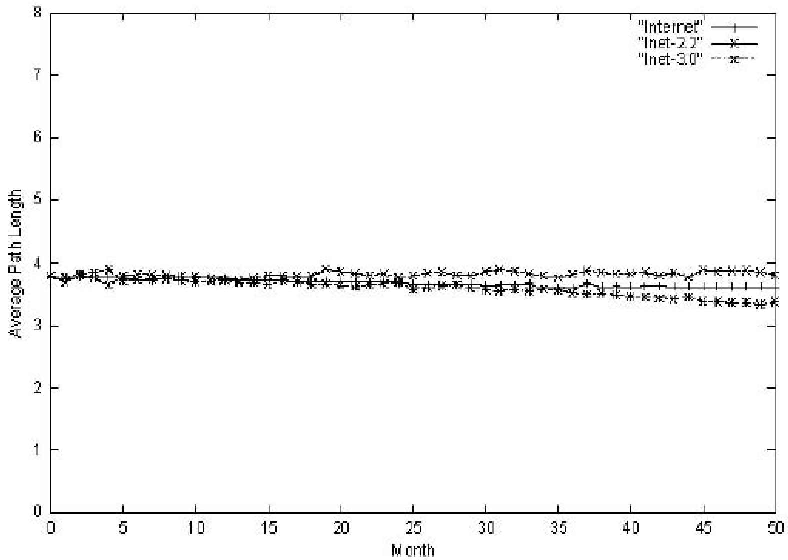
- Connect the degree-*1* nodes from V_1 to the spanning tree, according to the above same probability.
- Connect high-degree nodes to those available nodes without connections to V_1, also according to the above same probability (blue dashed lines).

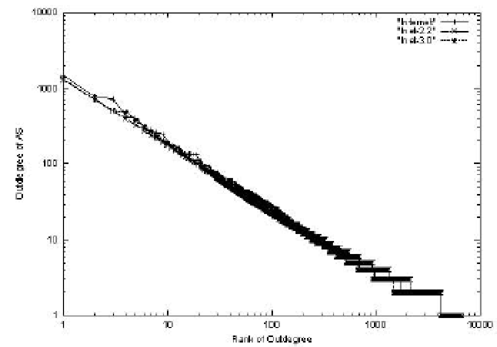


(Jin 2000)

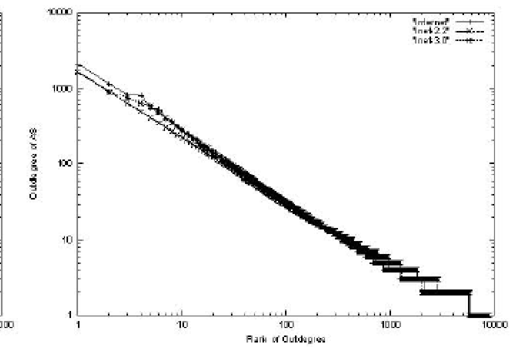
Inet

- University of Michigan (2002)
- Inet3.0: Program
- Simulations:
- 6700 nodes (Feb. 2000), 8880 nodes (Feb. 2001) and 12700 nodes (Feb. 2002)

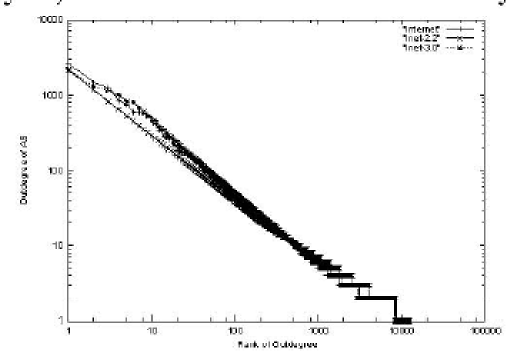




a. February 01, 2000.



b. February 01, 2001.



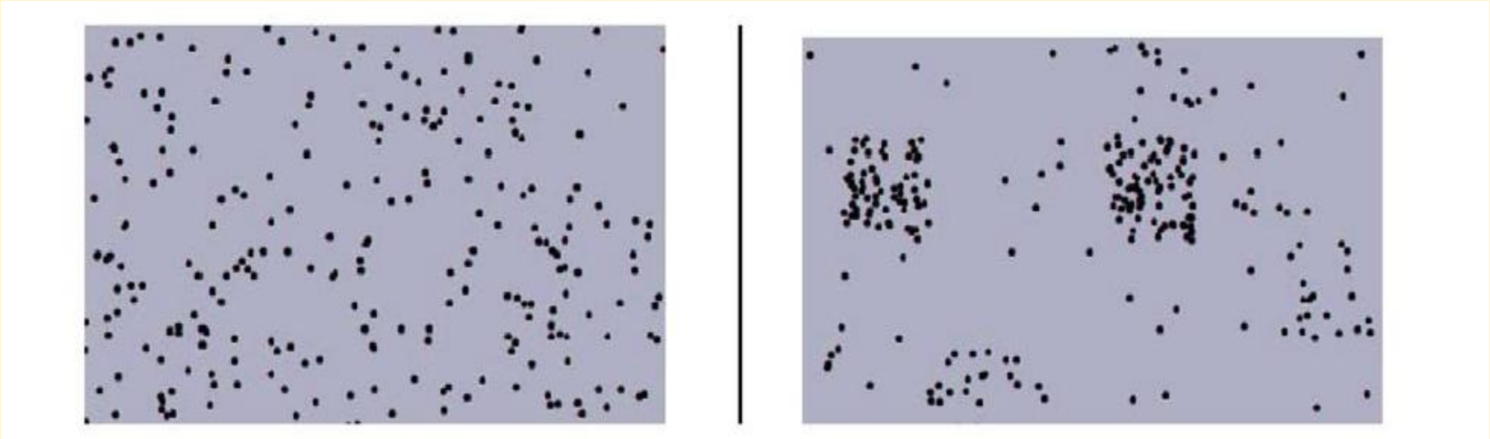
a. February 01, 2002.

Average path lengths

Out-degree power-law distributions

BRITE

- (Boston university Representative Internet Topology gEnerator)
- Software
- Framework:
  - Set a lattice on the plane, divide the lattice into some large squares, and then further divide all large squares into small squares.
  - According to a certain (e.g., uniform or Pareto) distribution, determine how many nodes will be assigned into each large square.
  - Then, in each large square, randomly pick a small square and assign at most one future node to it (next page shows how to add future nodes).



Average nodes: (a) Uniform distribution (b) Pareto distribution

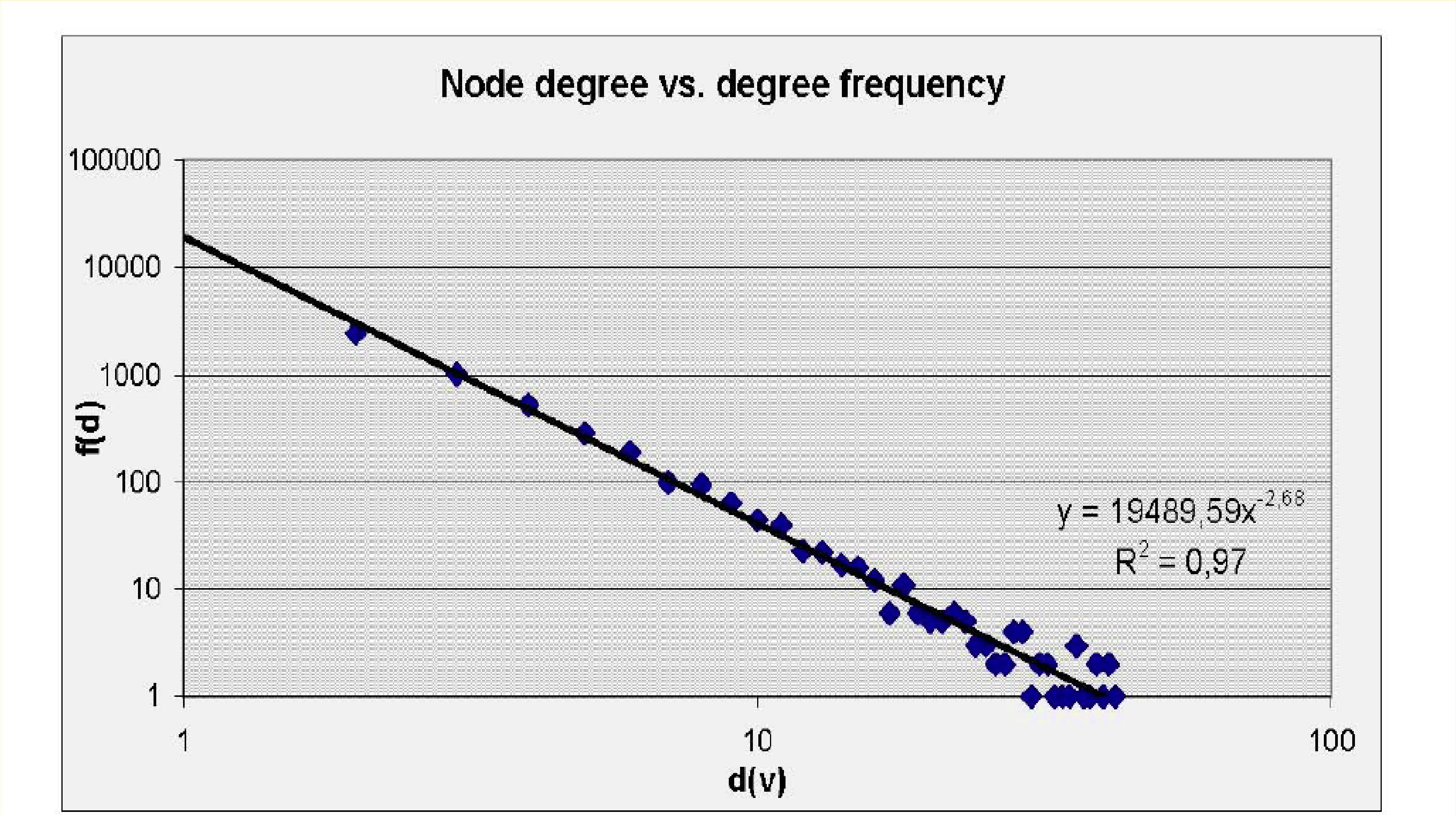
Now, start to add nodes:

- Initially, generate a random graph with m_0 nodes
- Then, add more nodes to the graph gradually.
- The way to connect nodes is determined by two parameters: Incremental Growth (IG) and Preferential Connectivity (PC):
  - if $IG = 0$ then put m nodes onto the plane simultaneously, and randomly pick one node among them and then connect it to the other nodes;
  - if $IG = 1$ then put one node onto the plane each time, and connect this new node to m existing nodes in the network.
- The way to establish connections is based on the PC parameter value:
  - if $PC = 0$ then follow the Waxman probability to connect the new node to the existing nodes;
  - if $PC = 1$ then follow the BA linear preferential attachment probability;
  - if $PC = 2$ then use the following weighted preferential attachment probability:

$$\Pi(k_i) = \frac{w_i k_i}{\sum_{j \in C} w_j k_j}$$

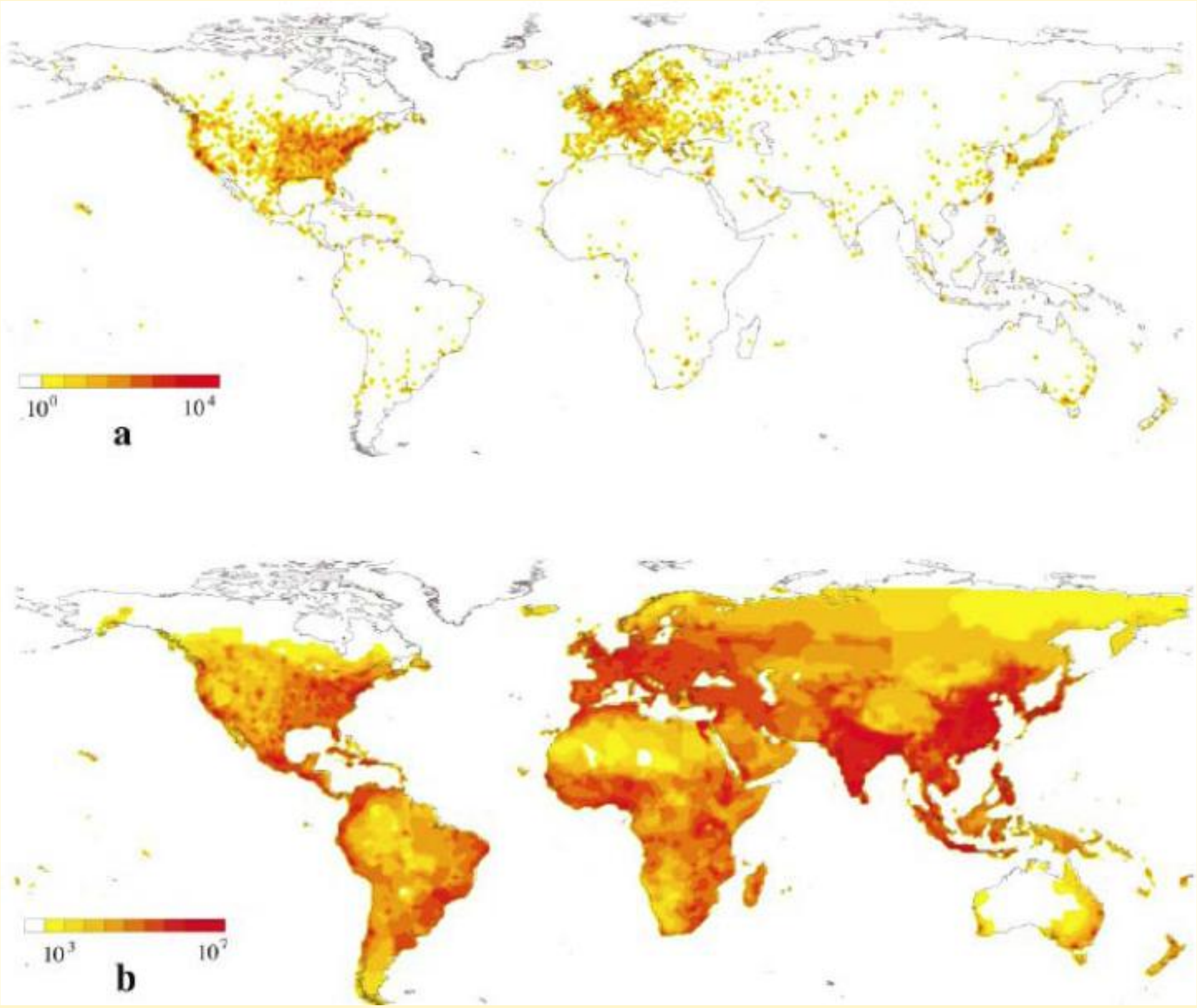
where k_i is the degree of node i, w_i is the Waxman probability, and C is the set of all m nodes being connected to node i.

BRITE



Node-degree distribution - 5000 router nodes (Di Fatta et al., 2001)

Geographic Layout of the Internet



(a) Router density   (b) Human population density

(Yook et al., 2002)

Geographic Layout of the Internet

Correlation between router interfaces and human population

	Population (Millions)	Interface	People per interface
Australia	18	18,277	975
Japan	136	37,649	3,631
Mexico	154	4,361	35,534
USA	299	282,048	1,061
South America	341	10,131	33,752
W. Europe	366	95,993	3,817
Africa	837	8,379	100,011

Data source

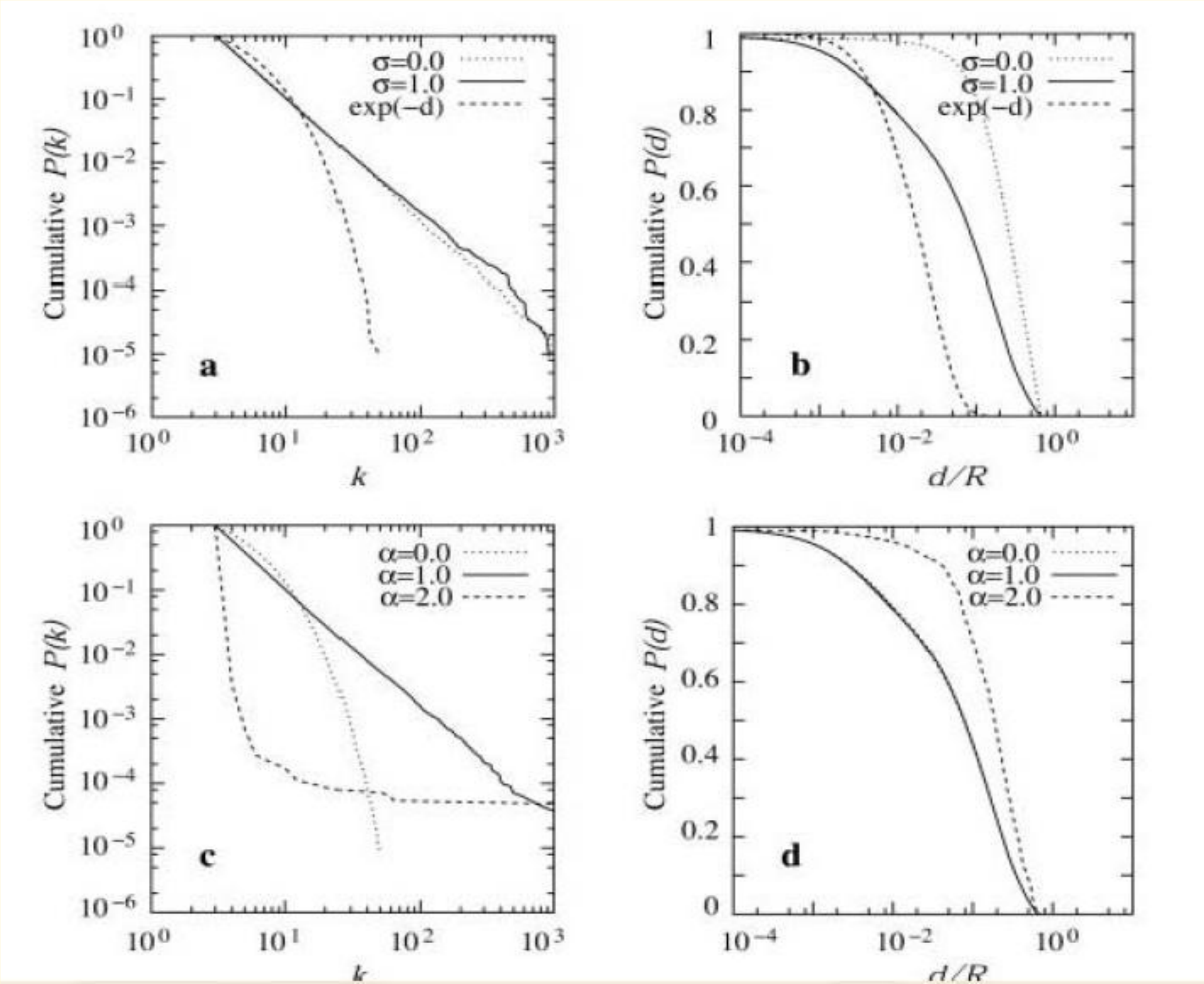
GeoBA model (Yook et al., 2002)

- Starting with a lattice consisting of many small squares
- Assign to each square a user population density $\rho(x,y)$
- At each step, add a node i into a square centered at (x,y) in such a way that the probability of adding node i to this square is proportional to its user population density
- This new node will bring in m new edges, and each edge connects to an existing node j of degree k_j, with geographic distance d_{ij} to node i, according to the probability (nonlinear preferential attachment)

$$\Pi(k_j, d_{ij}) \sim \frac{k_j^{\alpha}}{d_{ij}^{\sigma}}$$

where α and σ are constant parameters.

GeoBA model: Simulation Results



α =1 gives power-law

Limitation
of most scale-free Internet models

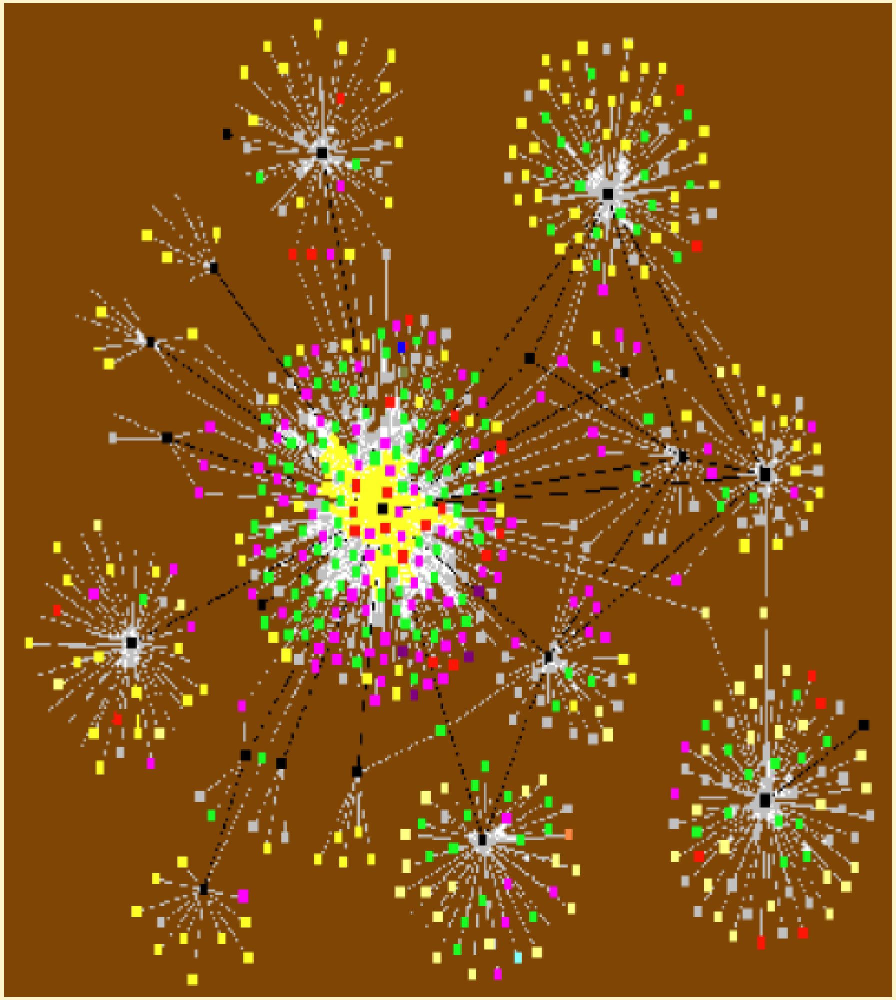
- Preferential Attachment:

$$\Pi_i = \frac{k_i}{\sum_j k_j} \qquad \text{(or, its variants)}$$

- They all use global preferential attachment –

  - Every newly added node requires the connectivity information of all nodes in the network
  - A real network only uses local preferential attachment with information of only some nodes in the network

Question: How to describe a topology of the AS-level Internet with localization property?



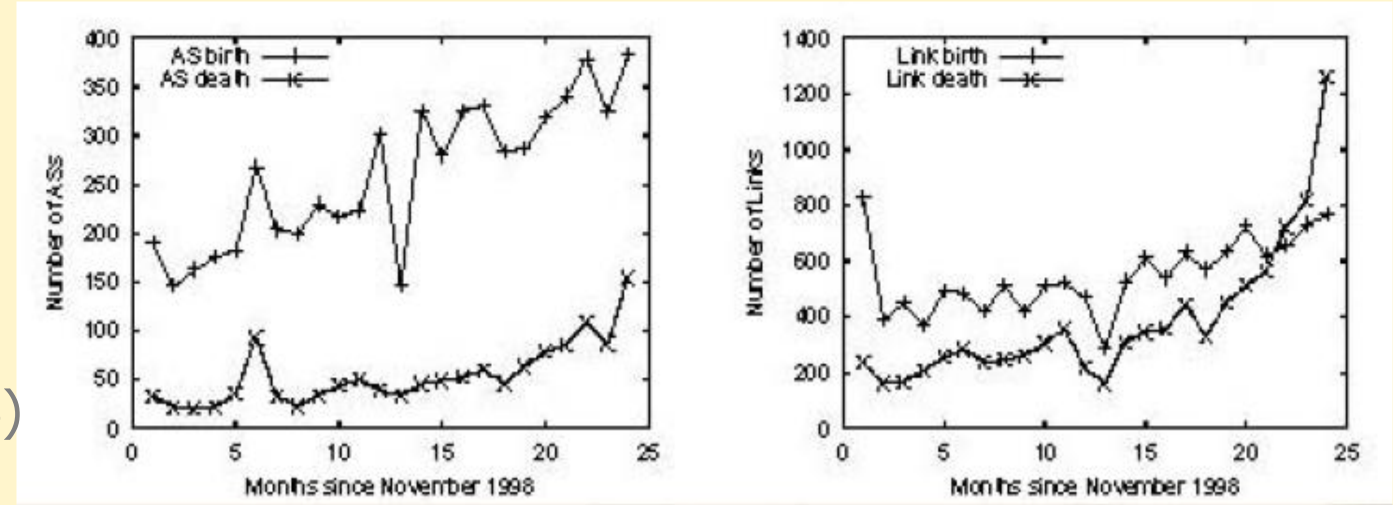
- The Internet consists of several sub-networks: each sub-network is called a "local-world"
- The newly added node only needs connectivity information of those nodes in a local-world
- The connections among different local-worlds are sparse
- The connections of nodes within the same local-world are dense

Multi-Local-World (MLW) Model

This model includes 5 events:

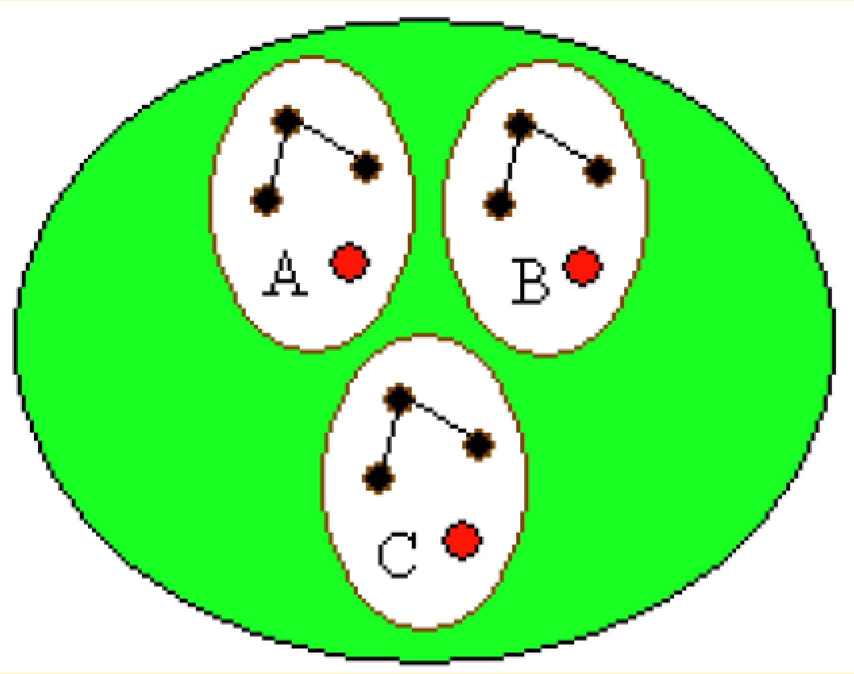
- Addition of local-worlds
- Addition of new nodes to local-worlds
- Addition of edges of new nodes to local-worlds
- Deletion of edges within a local-world
- Addition of edges among local-worlds

Oregon data (1998)



MLW Model

Start with m isolated local-worlds, with m_0 nodes and e_0 edges in each local-world



Example:

Start with $m=3$ local-worlds (A, B, C), with $m_0=3$ nodes (black circles) and $e_0=2$ edges in each local-world

Each local-world has a unique identifier (red circle)

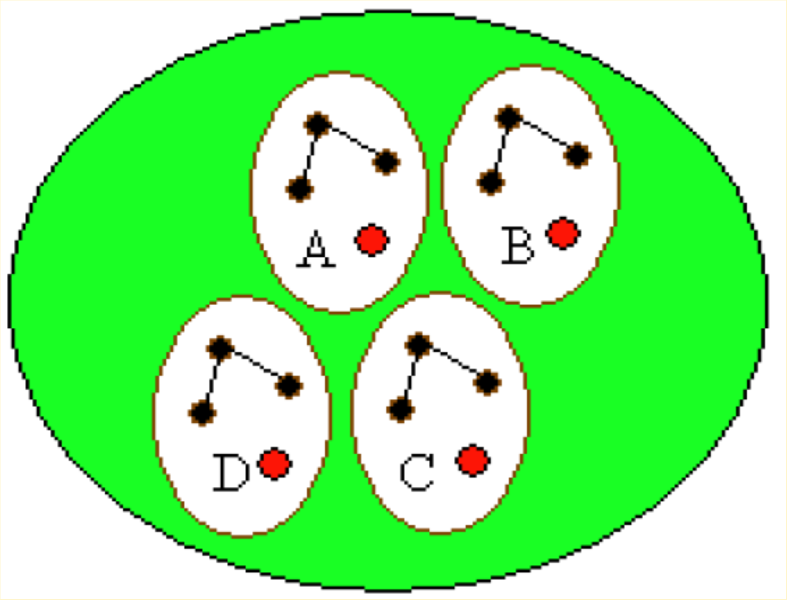
MLW Model

At each step, perform one of the five operations:

(i) With probability *p* a new local-world is created, which contains m_0 nodes and e_0 edges. Meanwhile, a unique identifier is generated for this new local-world.

Local world D is created with probability *p*

(with $m_0=3$ nodes (black circles) and $e_0=2$ edges)



MLW Model

(ii) With probability q a new node is added to an existing local-world, which has m_1 edges with the nodes within the same local-world:

First, a local-world Ω is selected at random.

Then, a node to which the new node connects in the local-world Ω is chosen with probability

$$\Pi(k_i) = \frac{k_i + \alpha}{\sum_{j \in \Omega}(k_j + \alpha)} \qquad (1)$$

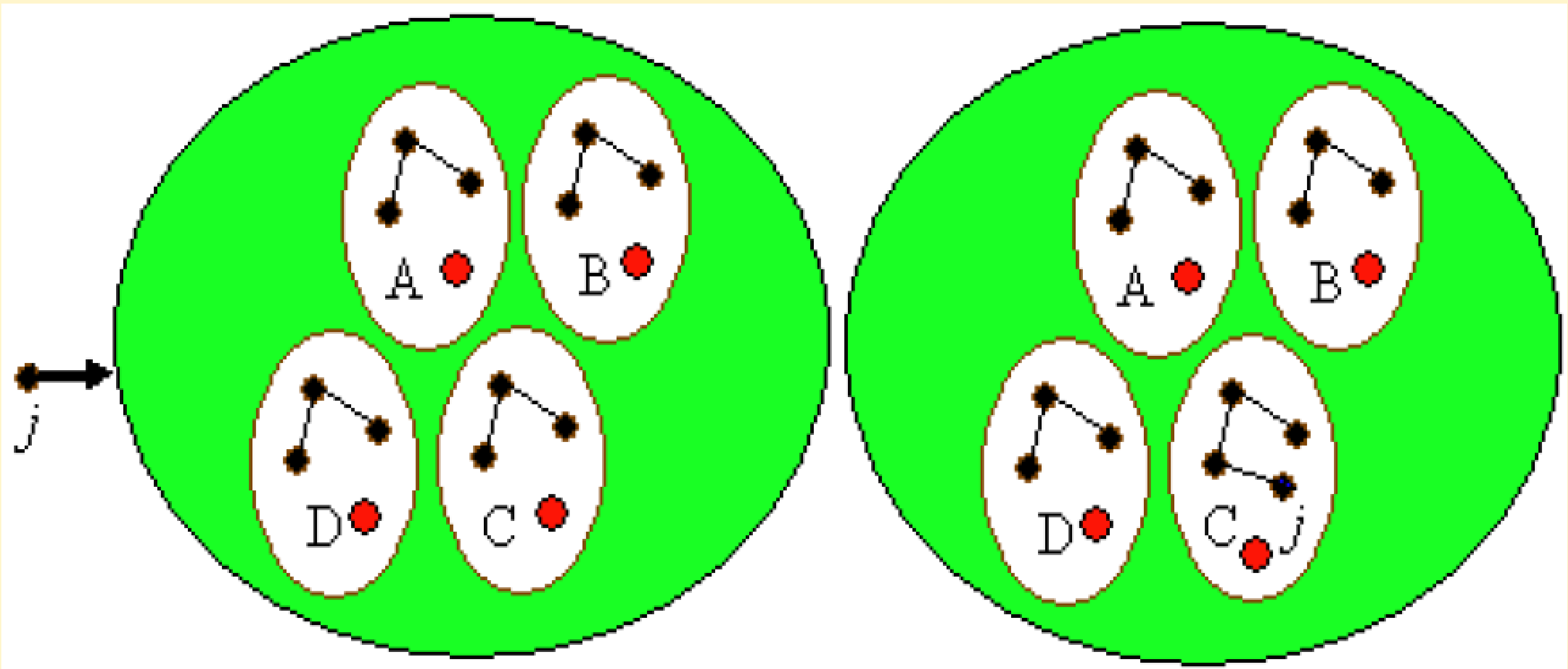
MLW Model

(Step (ii) continued)

In (1), Ω is the Ω-th local-world in which node i locates, and the parameter $\alpha>0$ represents the "attractiveness" of node i which is used to govern the probability for "young" nodes to get new edges.

This process is repeated m_1 times.

MLW Model
(Step (ii) continued)



Example (continued) A new node j joins the network. First, it selects the local-world C where it will locate, and then connects an existing node ($m_1=1$) in this local-world with preferential attachment probability given by (1)

MLW Model

(iii) With probability r, m_2 edges are added to a chosen local-world:

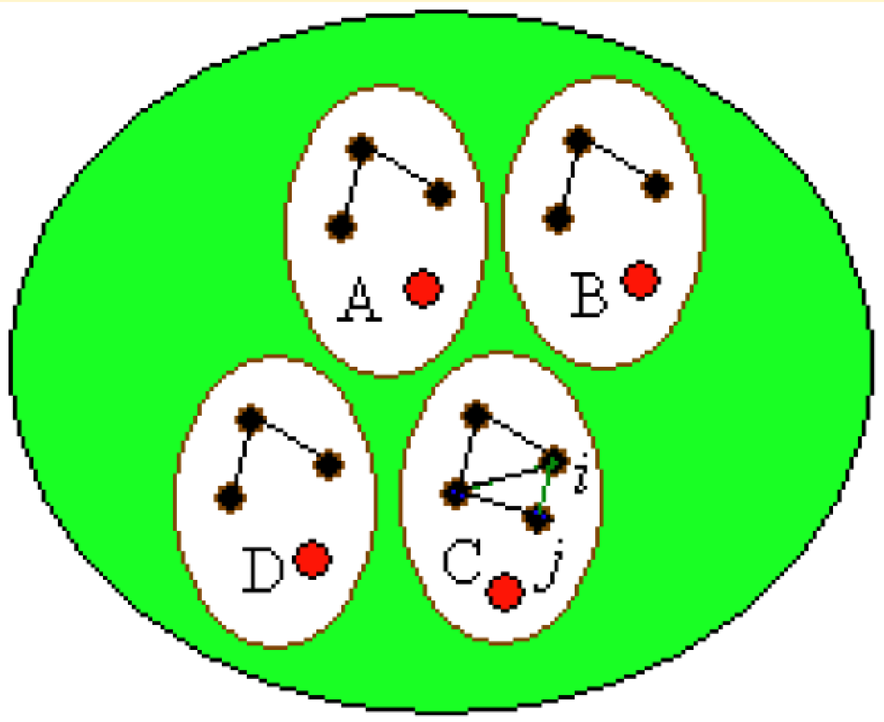
First, a local-world Ω is selected at random.

Then, one end of an edge is chosen at random, while the other end of the edge is selected by (1).

This process is repeated m_2 times.

MLW Model

(Step (iii) continued)



Example:

First, local-world C is chosen at random. Then $m_2=2$ edges are added to this local-world.

One end of an edge is selected at random, while the other end of the edge is chosen with a probability given by (1)

MLW Model

(iv) With probability s, m_3 edges are deleted within a chosen local-world:

First, a local-world Ω is selected at random.

Then, one end of an edge is chosen at random, while the other end of the edge is selected with probability

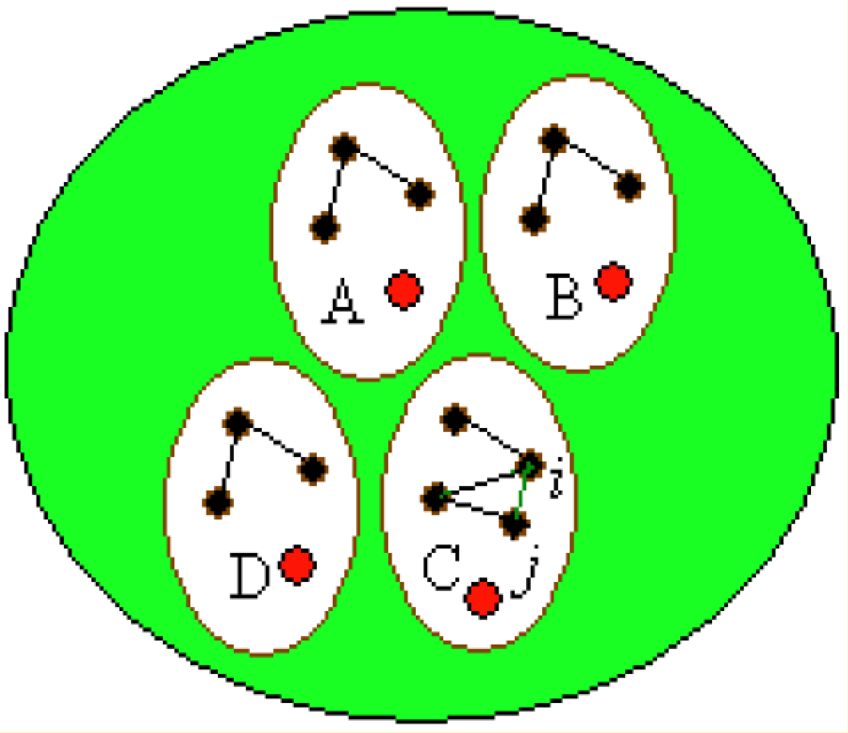
$$\Pi'(k_i) = \frac{1}{N_\Omega(t)-1}(1-\Pi(k_i)) \qquad (2)$$

where $N_\Omega(t)$ represents the number of nodes within the Ω-th local world, and $\prod(k_i)$ is determined by (1)

This process is repeated m_3 times.

MLW Model

(Step (iv) continued)



Example:

First, local-world C is chosen at random. Then $m_3=1$ edge is deleted within this chosen local-world.

An end of the edge is selected at random, while the other end of the edge is chosen with probability given by (2)

MLW Model

(v) With probability u, a selected local-world has m_4 edges with the other existing local-worlds:

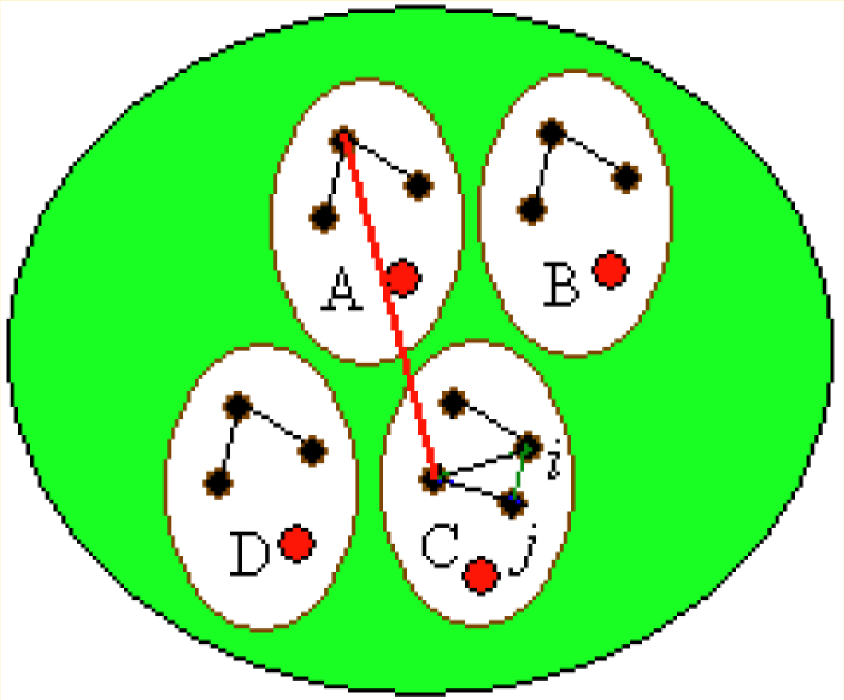
First, randomly select a local-world and a node in this local-world with probability given by (1).

Then, the selected node is acted as one end of an edge, while the other node of the edge, which is in another local-world chosen at random, is selected with probability given by (1).

This process is repeated m_4 times.

MLW Model

(Step (v) continued)

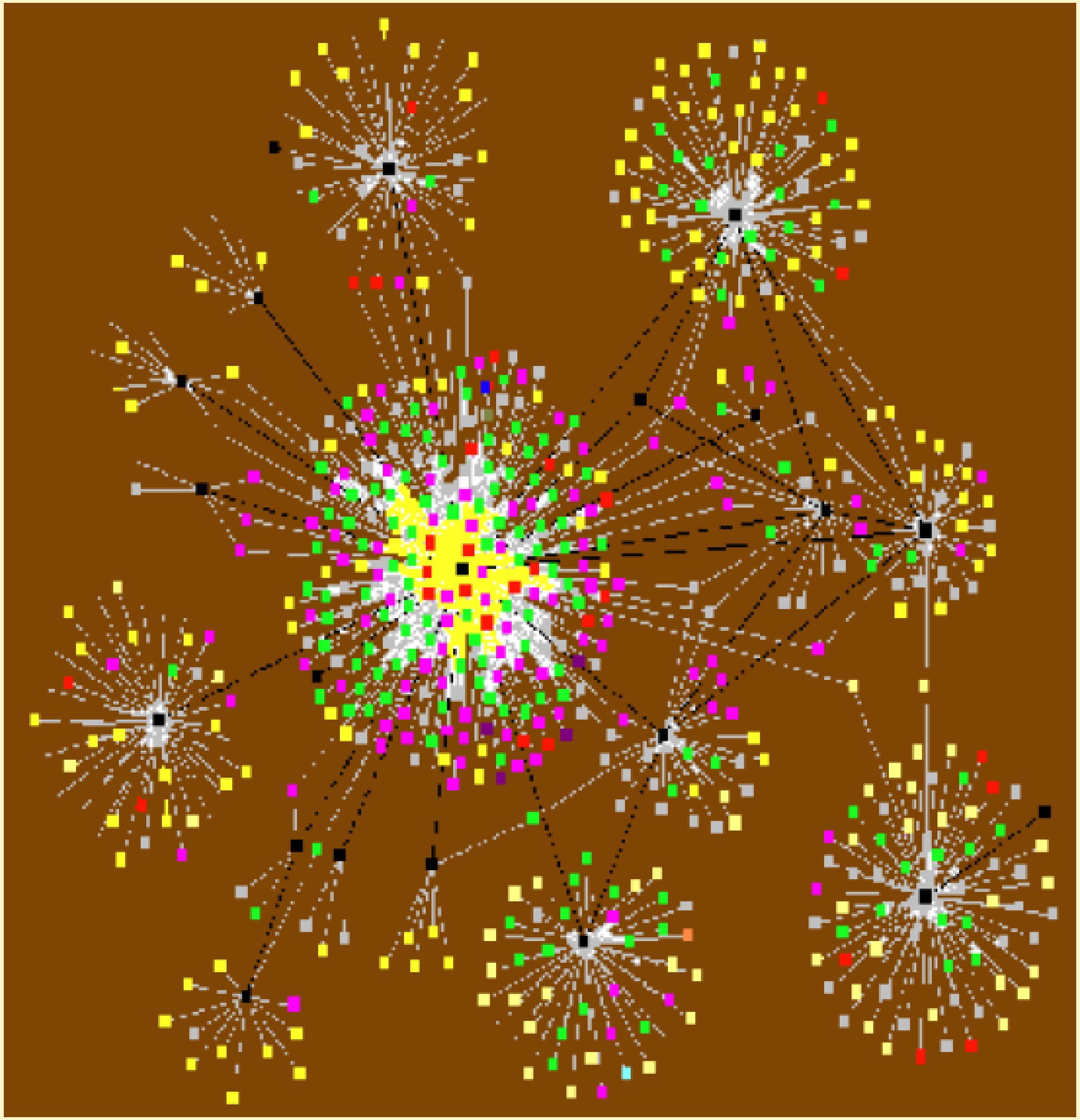


Example:

Depending on the probability *u*, $m_4=1$ link is added between two nodes in two different local-worlds.

Both ends of the link are chosen with preferential attachment according to a probability given by (1)

Illustration of a Resultant Network



MLW Model

Degree Distribution has a power-law form:

$$P(k) = \frac{t}{a(3m + t(1+2p))}\left(m_1 + b/a\right)^{1/a}\left(k + b/a\right)^{-\gamma}$$

Here:

$$0 \le p, r, s, u \le 1, \quad 0 < q \le 1, \quad p + q + r + s + u = 1$$

$$\gamma = 1 + 1/a$$

$$a = \frac{qm_1}{c} + \frac{rm_2(q + m_0 p - p)}{(q + m_0 p)c} + \frac{sm_3 p}{(q + m_0 p)c} + \frac{2um_4}{c}$$

$$b = \frac{q\alpha m_1}{c} + \frac{rm_2}{(q + m_0 p)} + \frac{rm_2(q + m_0 p - p)\alpha}{(q + m_0 p)c} + \frac{sm_3 p\alpha}{(q + m_0 p)c} - \frac{2sm_3}{(q + m_0 p)} + \frac{2um_4\alpha}{c}$$

Evaluating the Internet models

- Internet Models at the AS-level:

  - Waxman model
  - Transit-stub model

  Poisson distribution

  - Fluctuation-driven model
  - BA model
  - Generalization BA (GBA) model
  - Fitness model
  - HOT model

  Power-law distribution

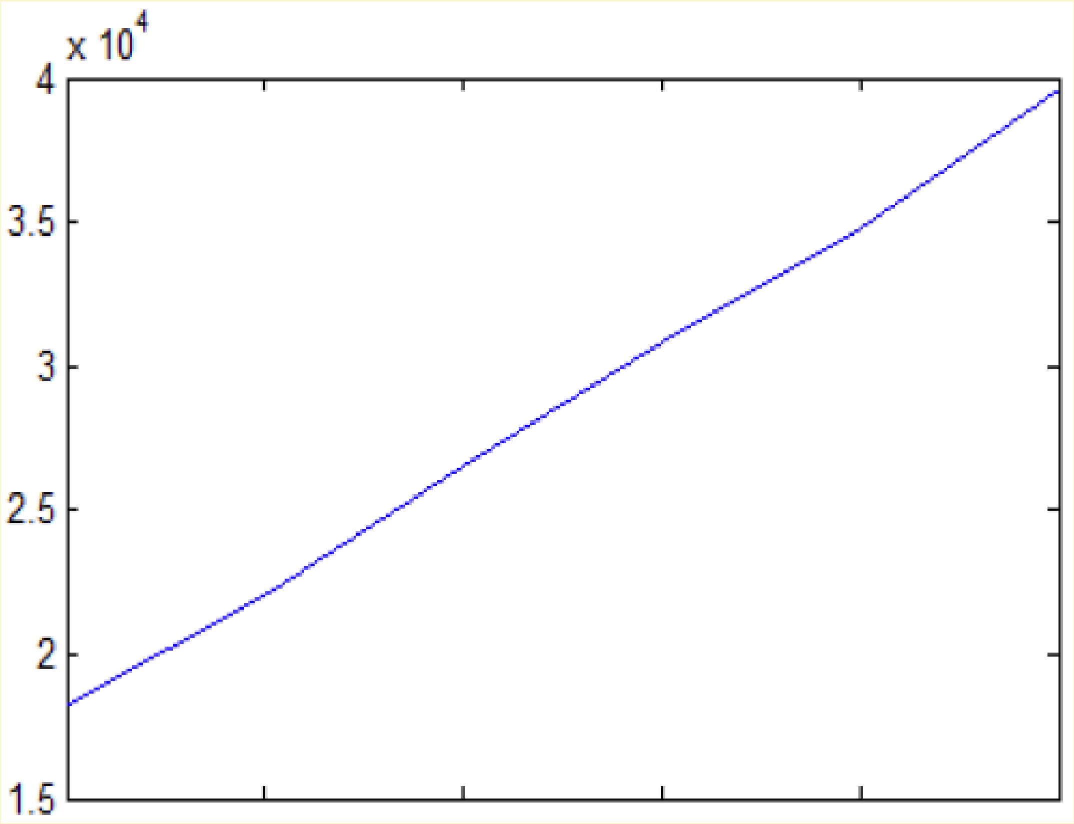
Evaluating the Internet models (cont.)

- Fluctuation-driven model - Exponentially growing network

- BA model
- Generalization BA (GBA) model
- Fitness model
- HOT model

Linearly growing network

Evaluating the Internet models (cont.)



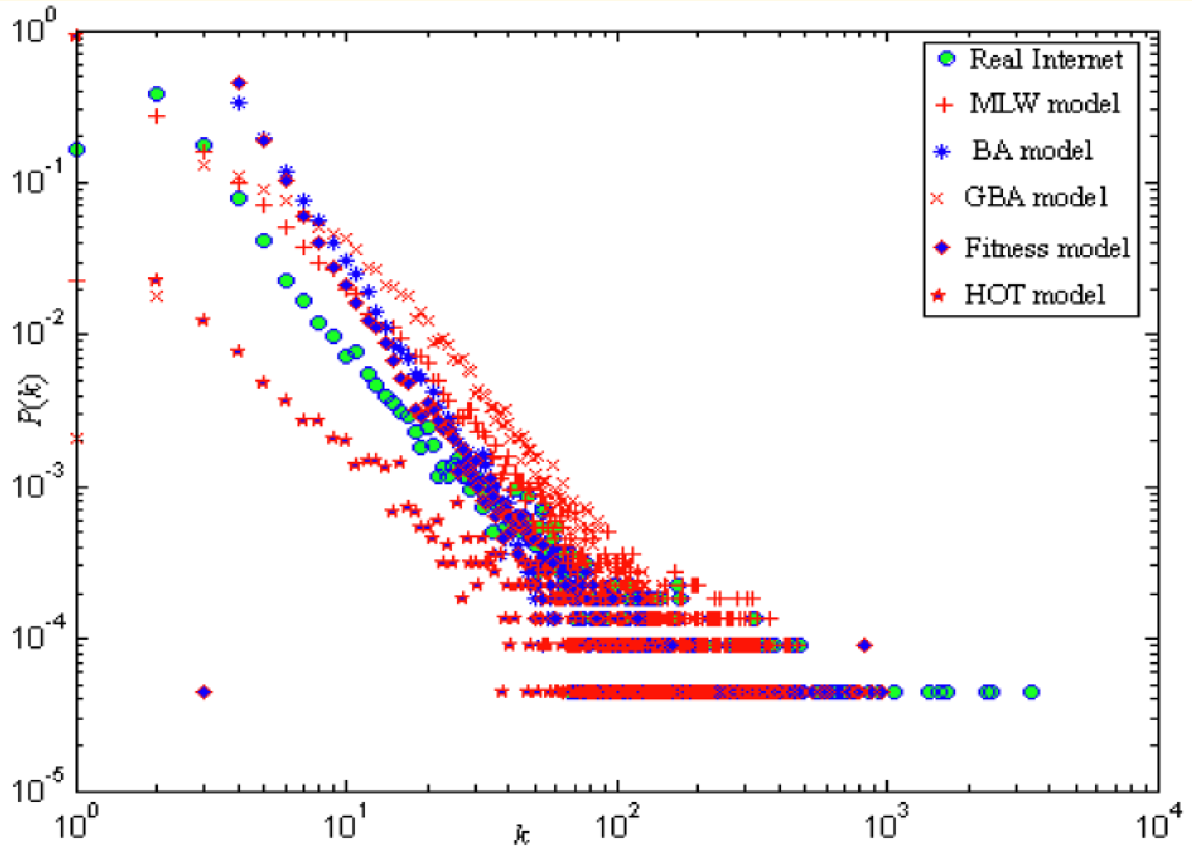
Number of Nodes

Linearly growing network

Fluctuation-driven model is NOT suitable for the AS-level Internet

Data (from 2004 to 2010)

Evaluating the Internet models (cont.)



Internet snapshot on May 15, 2005

Power Exponent:

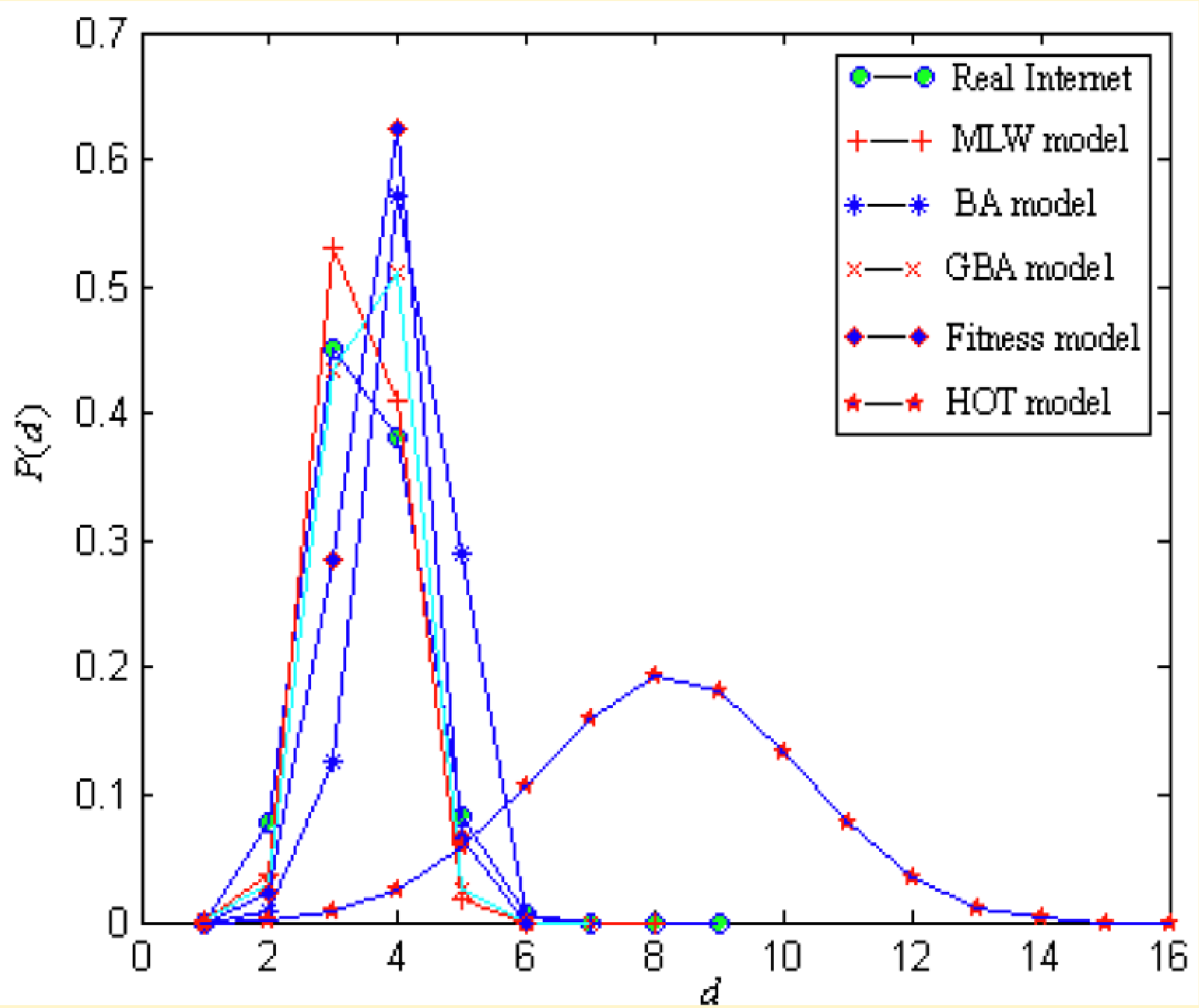
r=2.2 for real Internet

r=3.0 for BA model

r=1.5 for HOT model

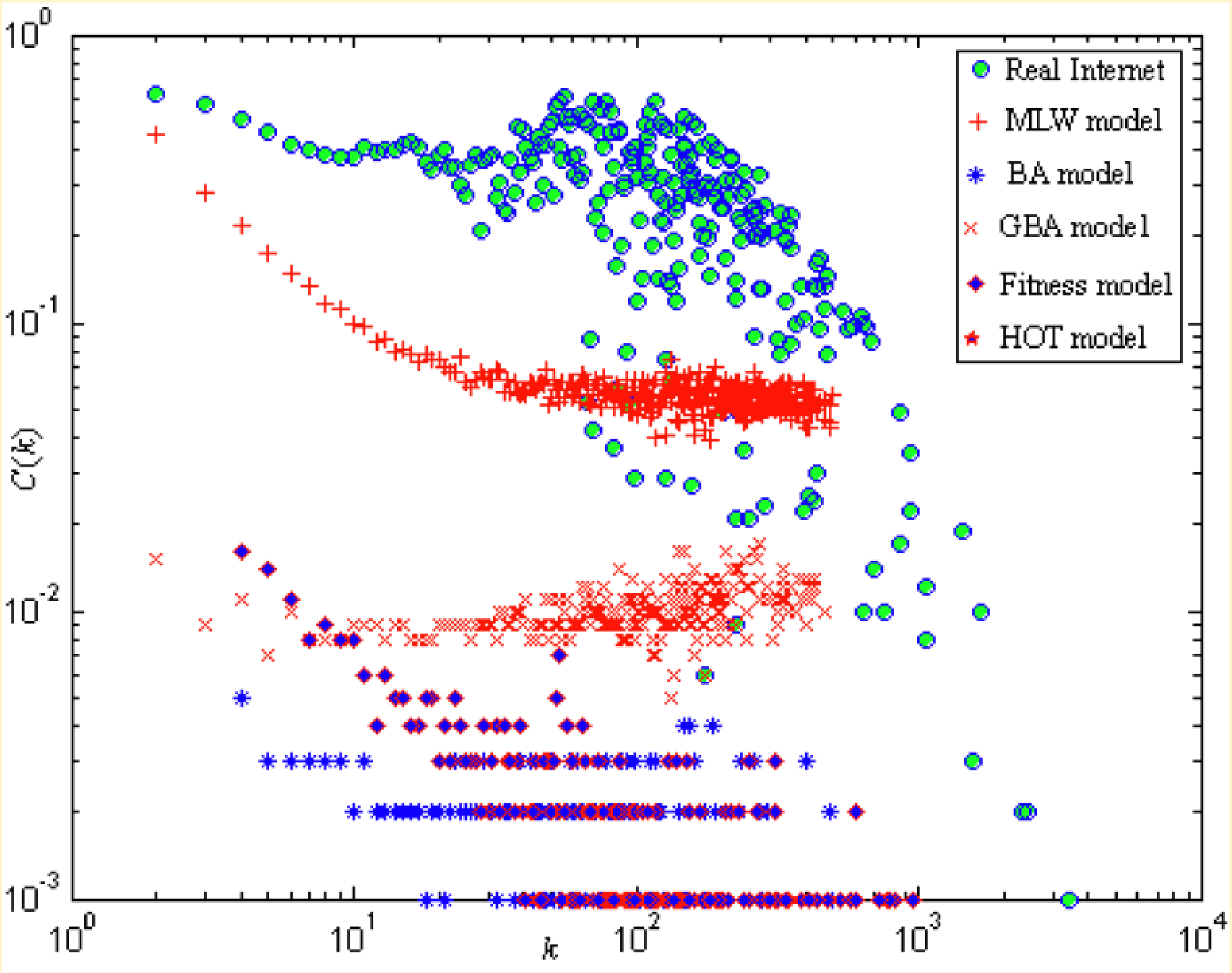
BA and HOT models can NOT capture the scale-free feature of the Internet

Evaluating the Internet models (cont.)



Distance distribution of the Internet and of different scale-free models

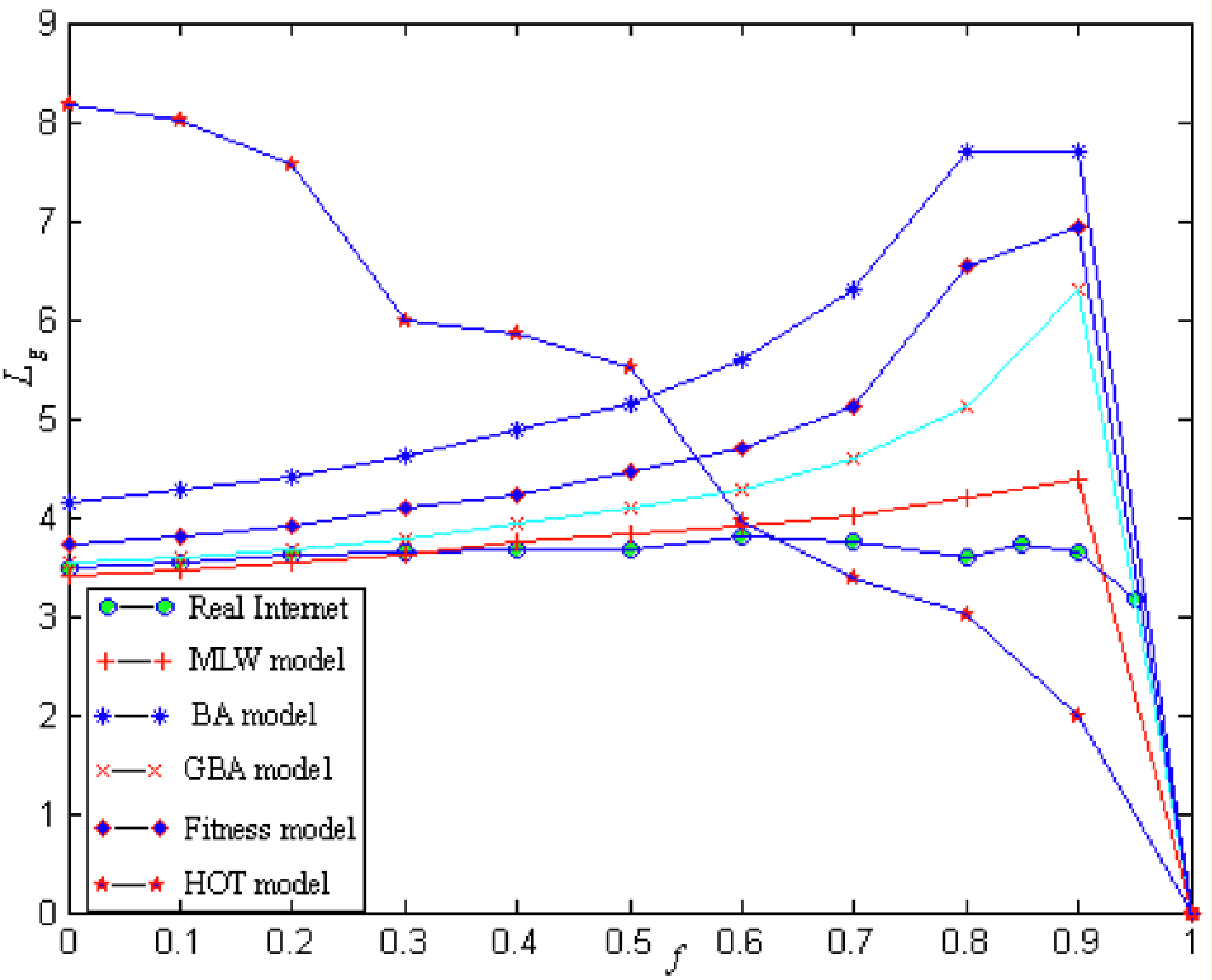
Evaluating the Internet models (cont.)



BA, GBA, Fitness, and HOT models can NOT capture the small-world feature of the Internet

Clustering coefficients as functions of the degree for the real Internet and the BA, GBA, Fitness, HOT, and MLW models.

Evaluating the Internet models (cont.)



Comparison of the **average shortest path-lengths** in the giant component for the real Internet and the five models studied.

Evaluating the Internet models (cont.)

Comparison: Four Models vs Real

	BA	GBA	Fitness	MLW	**Real Internet** (May 15,2005)
N	21999	21999	21999	21999	21999
$\overline{C}$	0.003	0.01	0.01	**0.24**	**0.46**
$\overline{d}$	4.14	**3.49**	3.71	3.45	**3.49**
γ	3	2.69	2.45	**2.36**	**2.18**
$\lambda_{\max}$	27.82	62.83	39.16	**111.87**	**141.12**

Comparison of MLW Model with Other Models

	Structural Features of the Internet	
	Scale-free feature	Small-world feature
BA model	Yes	No
EBA model	Yes	No
Fitness model	Yes	No
HOT model	Yes	No
MLW model	YES	YES

MLW model is better than the BA, GBA, and Fitness models in capturing the scale-free and small-world features of the Internet

Further Evaluating the Internet Models

Summary -

MLW model is better than BA, GBA, Fitness, and HOT models in capturing the scale-free and small- world features of the Internet

Topological Statistics -

degree distribution, power-law exponent,

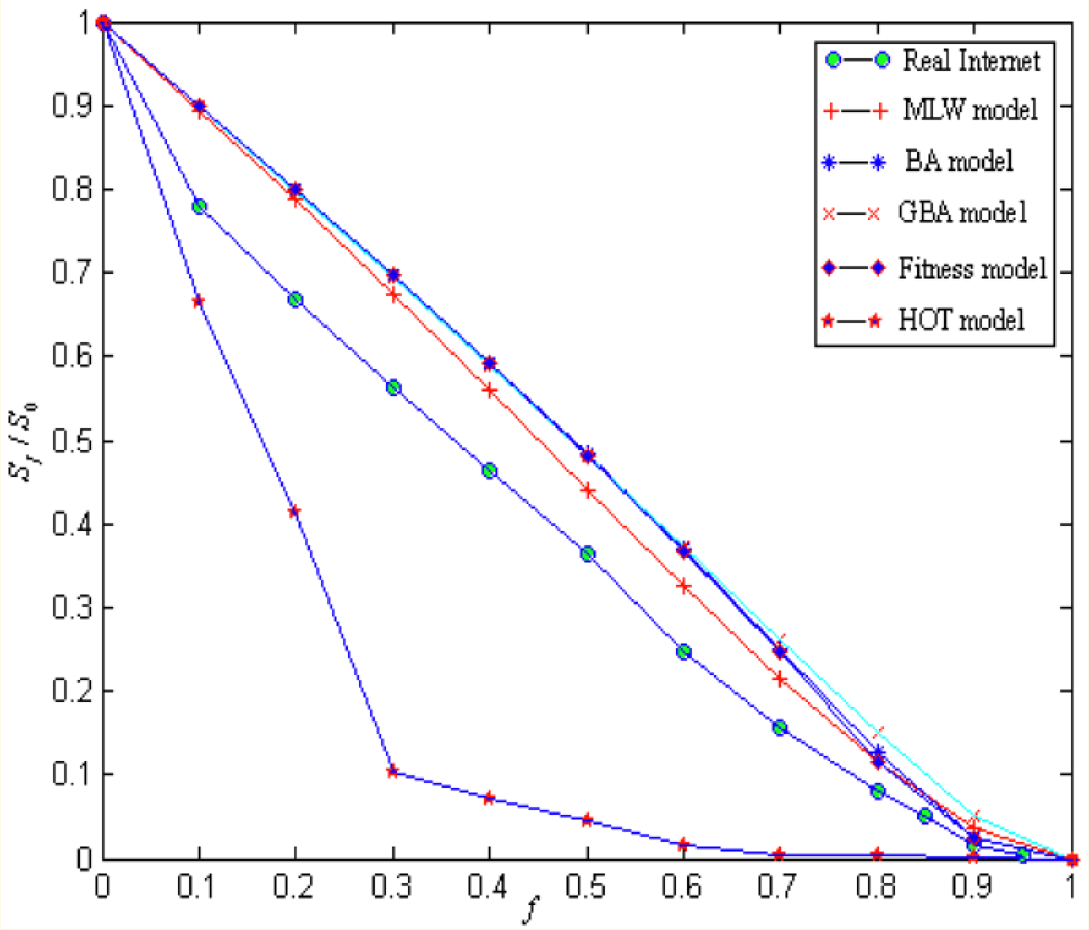
distance distribution, clustering coefficient,

average shortest path-length, hierarchical clustering

But what about performances?

Comparison: What about performances?

Robustness against random attacks



S_f: the size of the largest component after a fraction of nodes, f, in the network are randomly removed.

S_f / S_0 measures the capability of the network in which nodes still can communicate each other after the f portion of nodes has been randomly removed.

Comparison: What about performances?

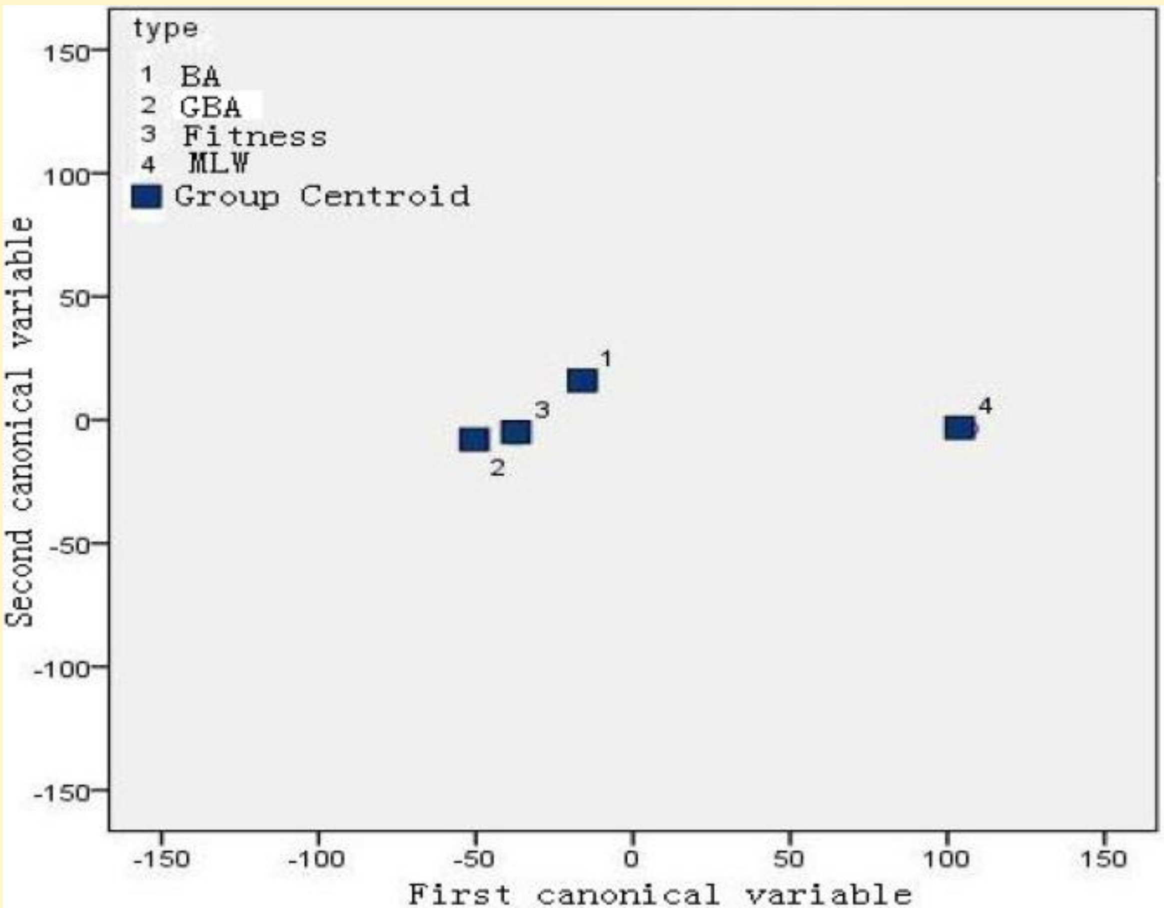
- Canonical variable analysis
- Bayesian decision theory

Topological measurements are projected into a reduced dimensional feature space by using canonical analysis, so that the Bayesian decision method can be applied onto a more representative feature space in a lower dimension.

L. F. Costa, F. A. Rodrigues, G. Travieso, P. R. V. Boas, Advances in Physics 56(2007): 167

Comparison: Bayesian Test

Consider: average clustering coefficient, average distance, and largest nonzero eigenvalue of adjacency matrix



Bayesian decision method

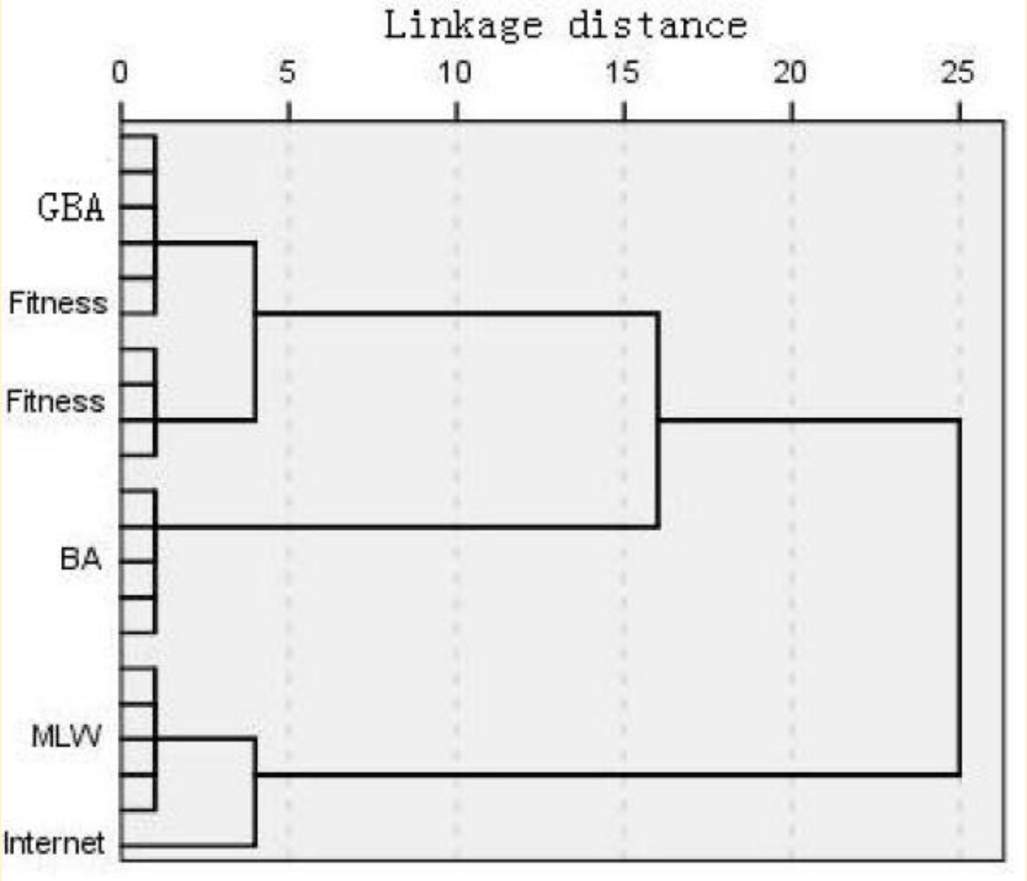
→

MLW model is most compatible with the Internet

Comparison: Hierarchical Clustering Algorithm

Applying the hierarchical clustering algorithm to evaluate different Internet models

Principle:
The sooner two networks are merged, the more similar they are



L. F. Costa, F. N. Silva, J. Stat. Phys. 125(2006): 841

MLW model is closest to the Internet

Conclusions

- MLW is the best model for the AS-level Internet as compared to Fluctuation-Driven, BA, EBA, and Fitness models
- These comparisons were performed only based on part of the Internet features:
- degree distribution
- distance distribution
- average path-length
- clustering coefficient
- robustness against random attack
- The MLW model is rather complicated
- More comparisons are needed
- Internet is too complex to comprehend. As of today, there is no commonly-agreed model of the Internet.

→ Good models are badly needed ….