Scale-Free Trees : Skeleton of Complex Networks

"Nonequilibrium Statistical Physics of Complex Systems"
Satellite meeting of STATPHYS 2004 in Seoul, Korea (July 2004)

Hawoong Jeong
(Dept. of Physics, KAIST)
Dong-Hee Kim, Seung-Woo Son

http://stat.kaist.ac.kr

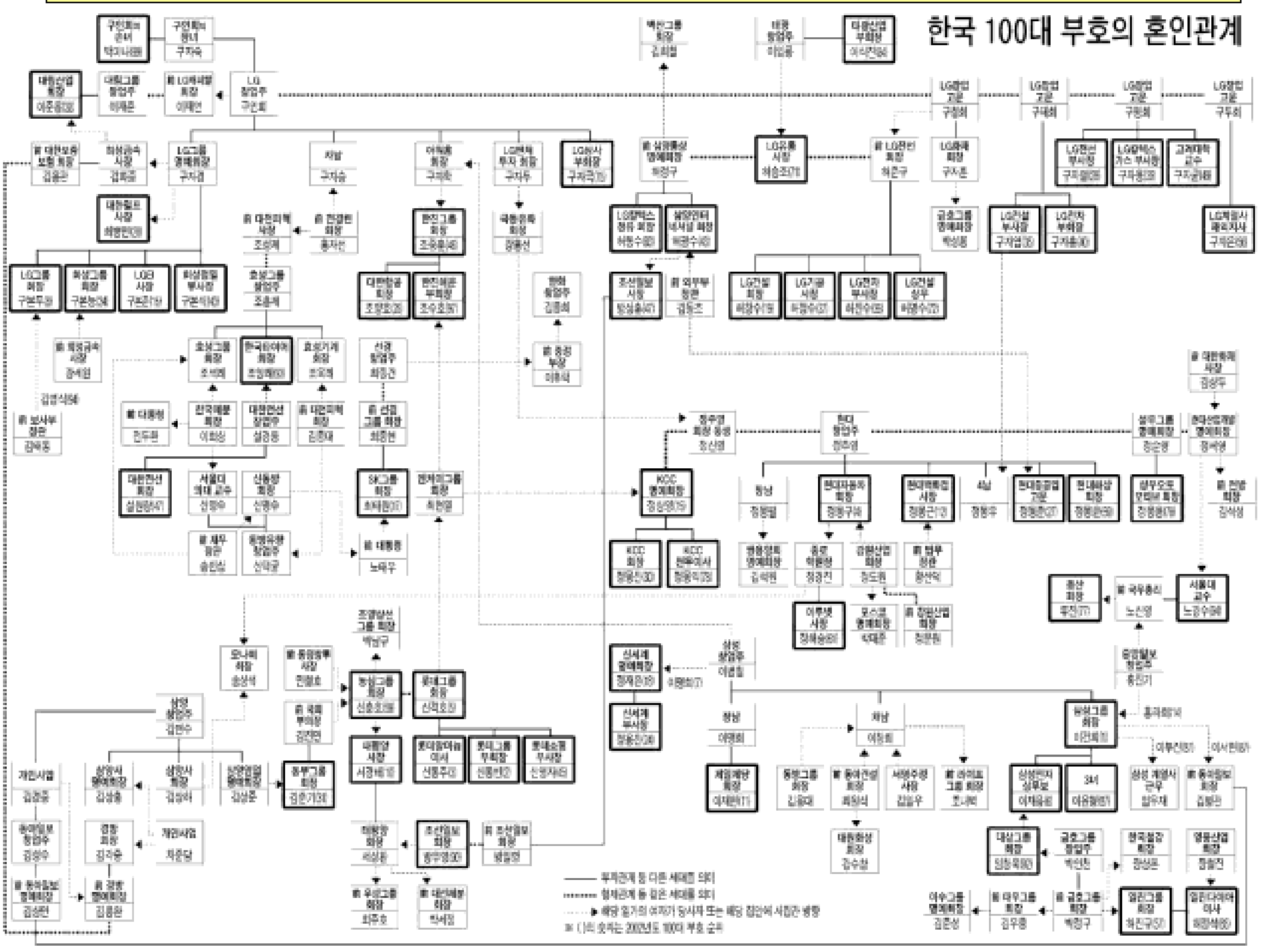
Ubiquitous networks are around us.

NETWORKS ARE EVERYWHERE!

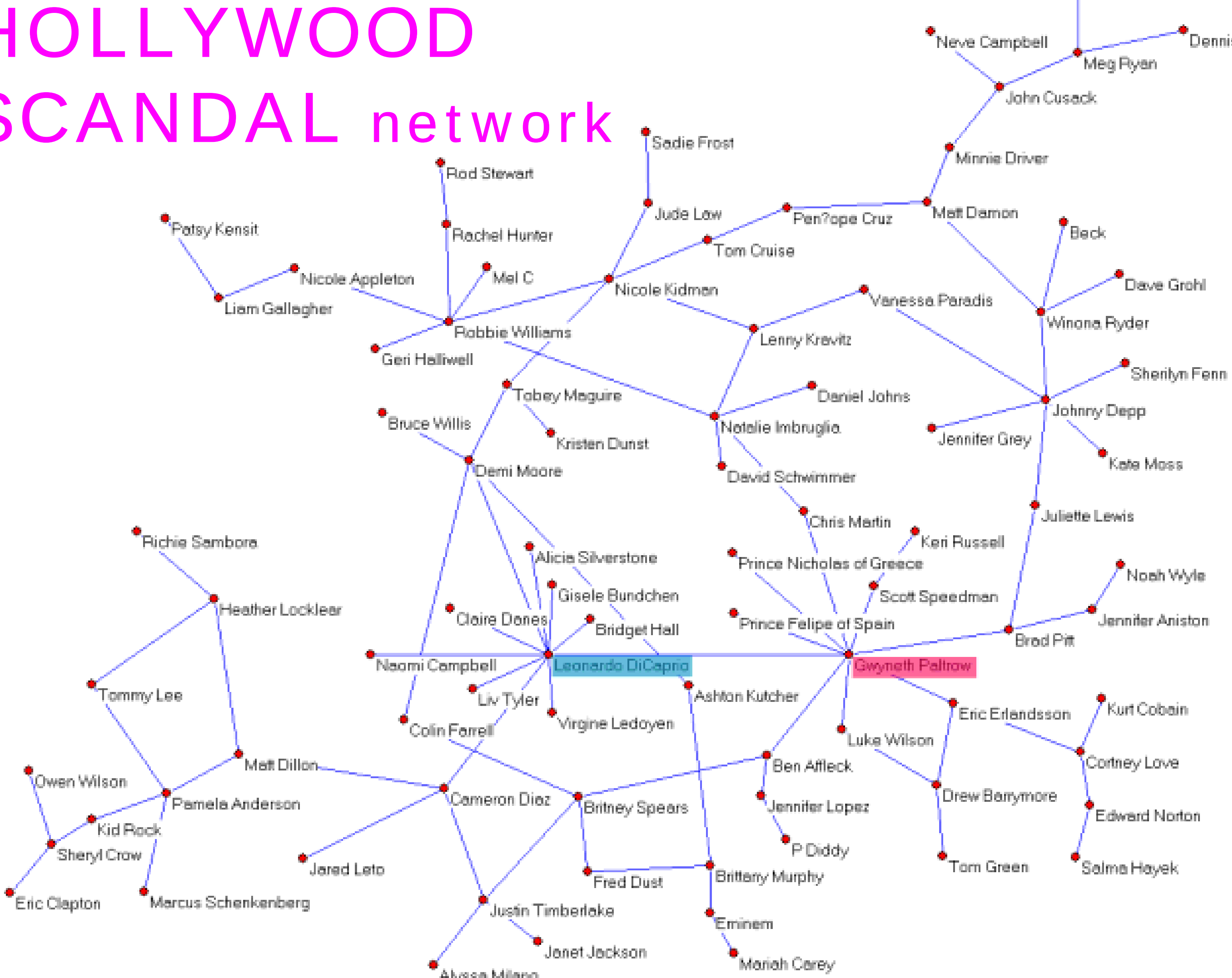
- Society = Social networks
  (Family tree, friendship network etc.)
- Money flow = Economic networks
- Highway, Subway = Transportation networks
- Metabolism, Cell cycle = Biological networks
- Food web = Ecological network
- Internet, WWW = Communication networks

We're living in a "connected world".

Marriage map between 100 richest people in Korea

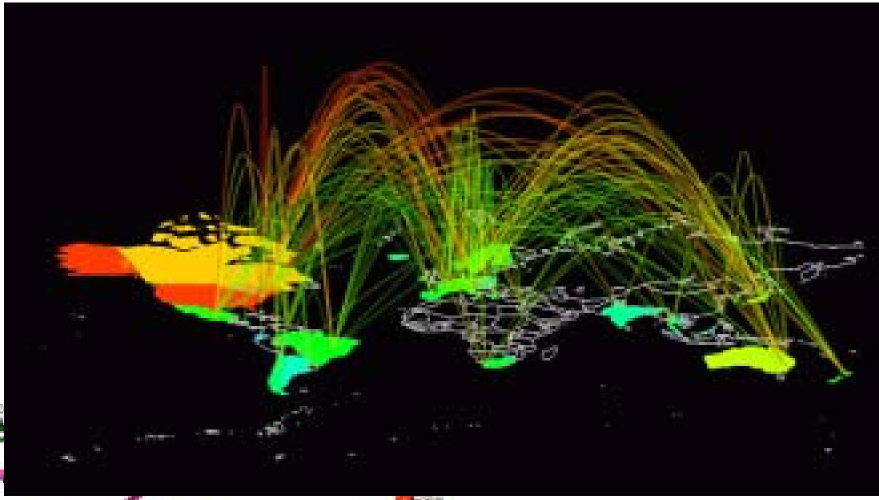


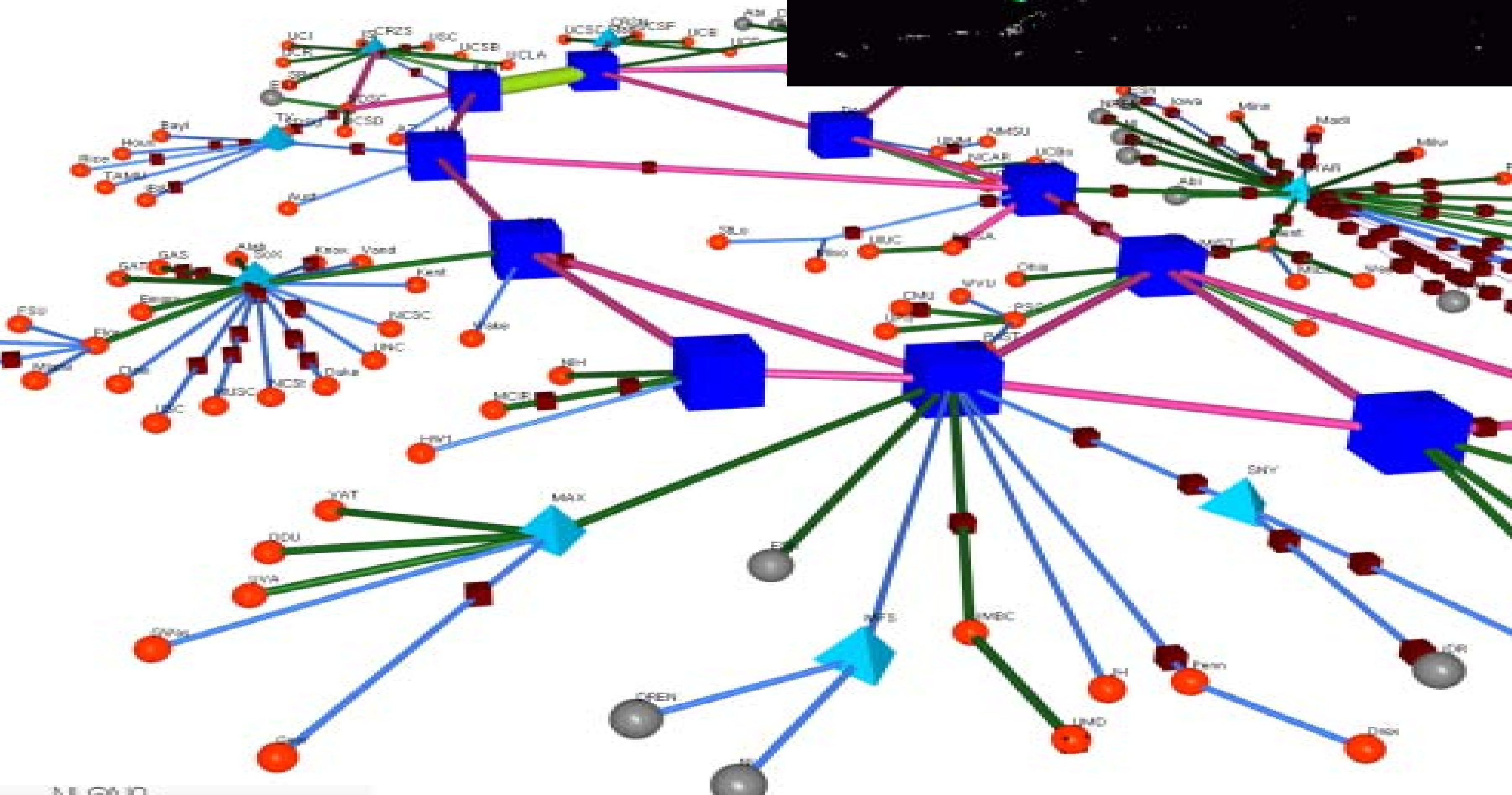
HOLLYWOOD SCANDAL network



(P. Holme, probably & hopefully not publishe

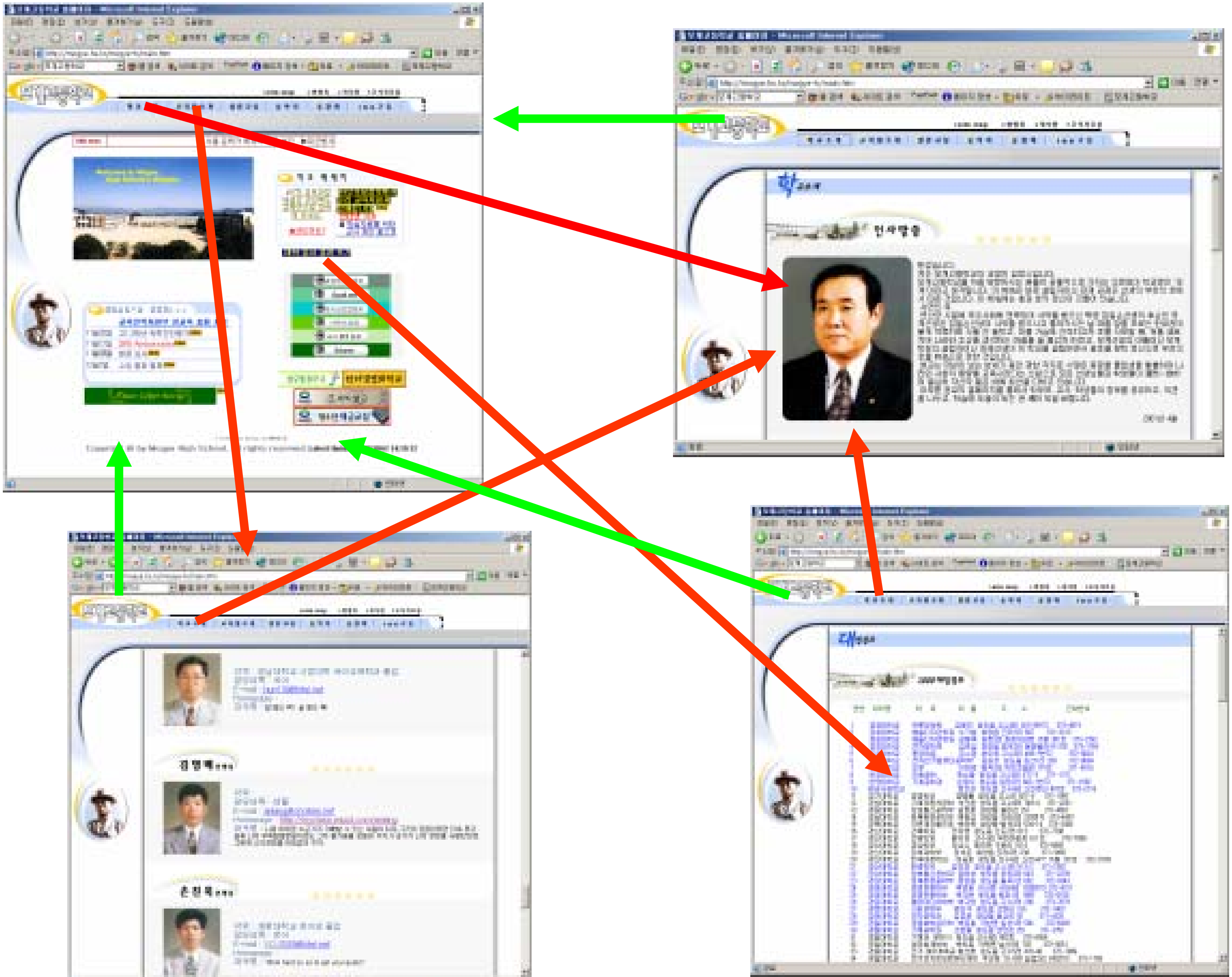
INTERNET!




UCI
USC
UCSB
UCLA
UCSC
UCSF
UCB
Iowa
NMSU
NCAR
UIUC
NIH
MAX
NCSC
UNC
Duke
SNY
UMD
Penn

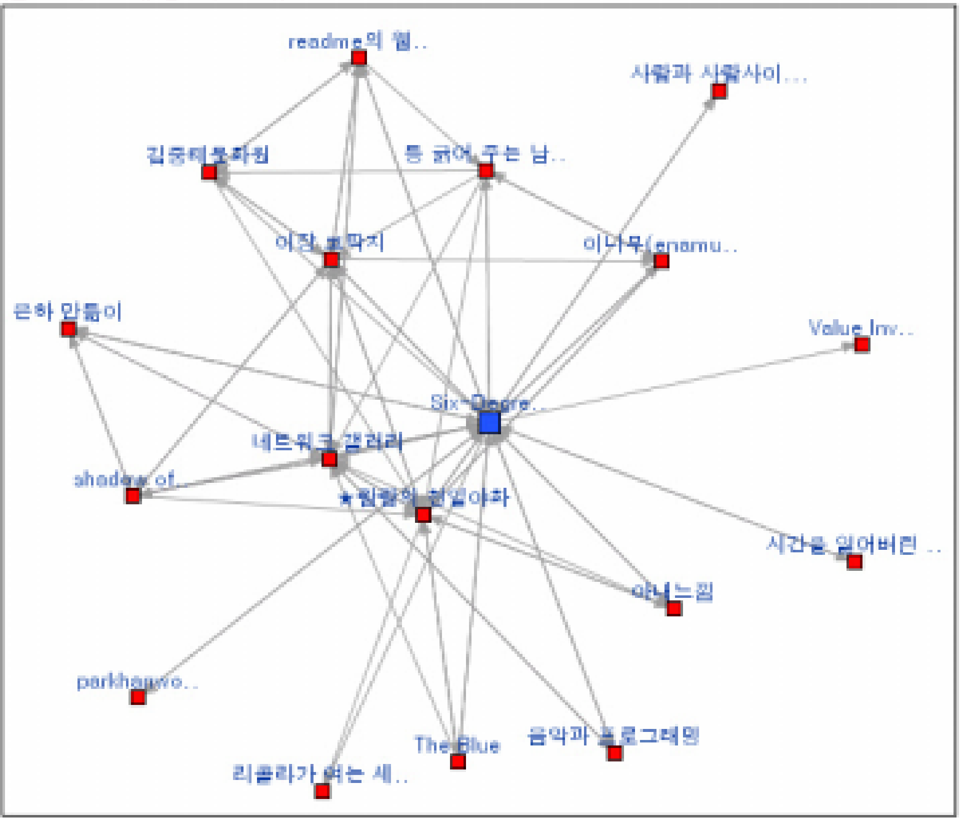
World Wide Web

node: web-page
link: hyper-link



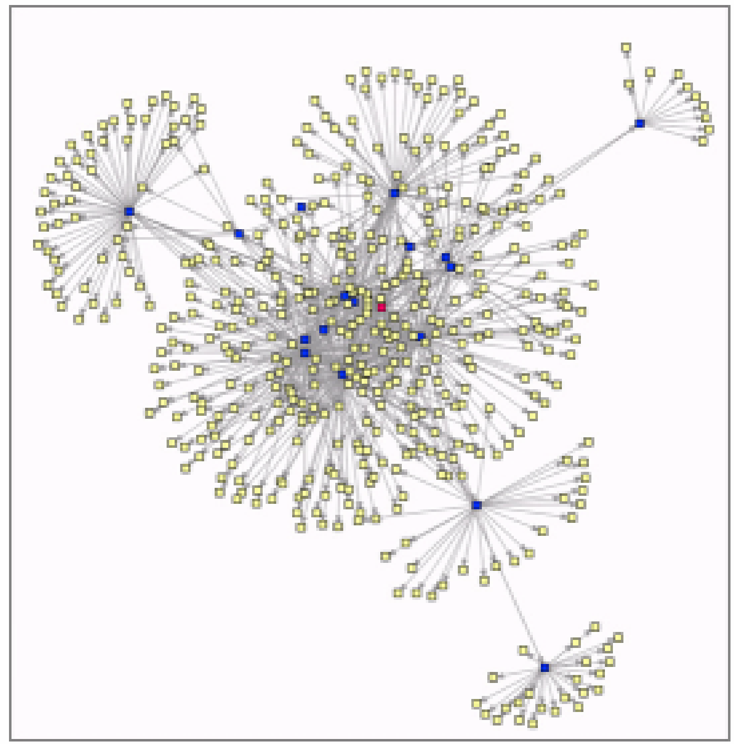
BLOG : World Wide Web? + Social Network?

(블로그수: 18개, 링크수: 74개)



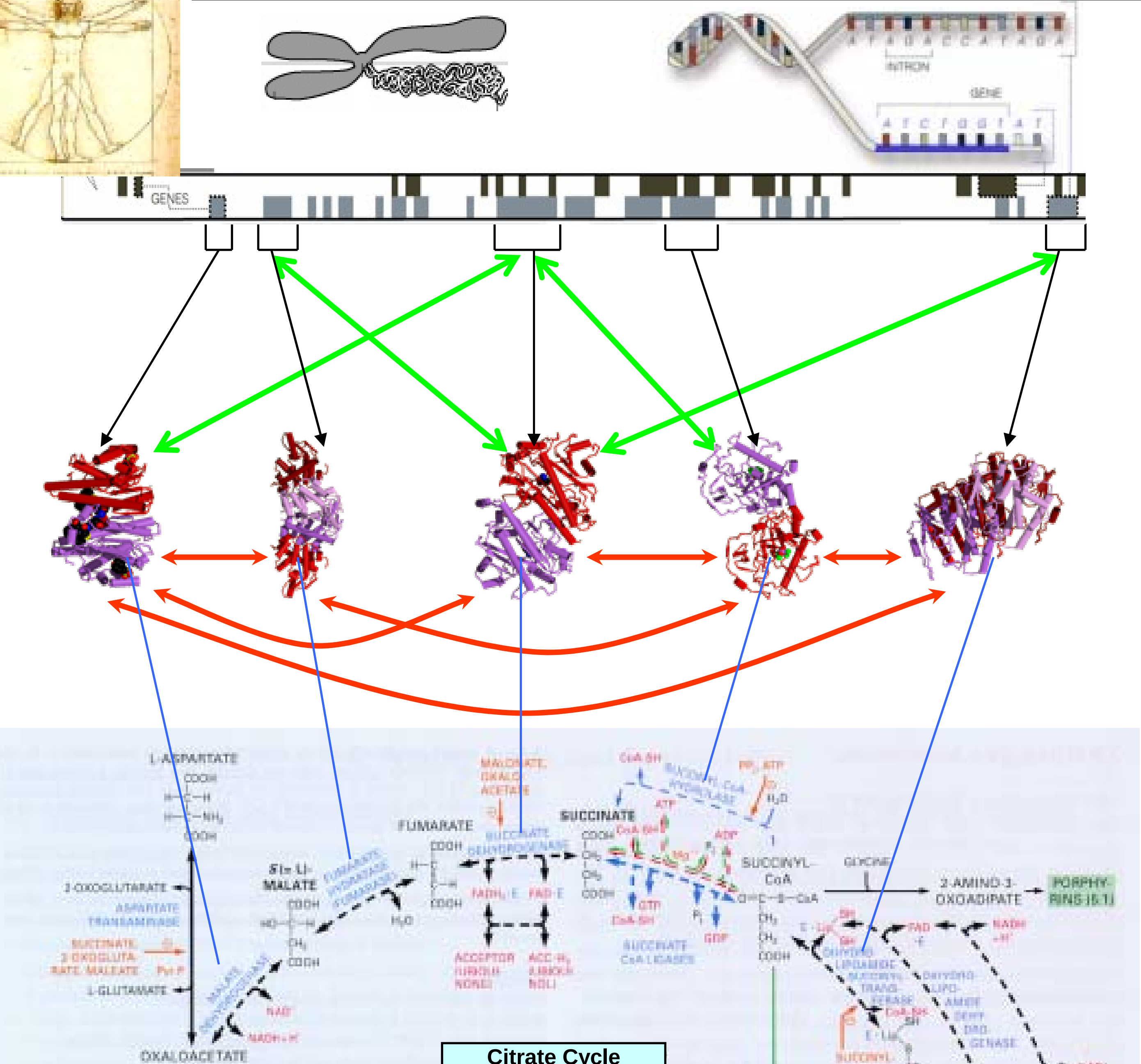
[4] 내 블로그 이웃들과 그들의 이웃들까지를 포함한 링크 차트이다. 그림에서 빨간색 노드가 Six-Degrees of Bloggers이고, 파란색이 1차 이웃, 노란색이 2차 이웃(이웃의 이웃)이다.

(블로그수: 492개, 링크수: 710개)



Web + log = b + log = blog

Biological Networks in a living organism



genes

protein- gene interaction

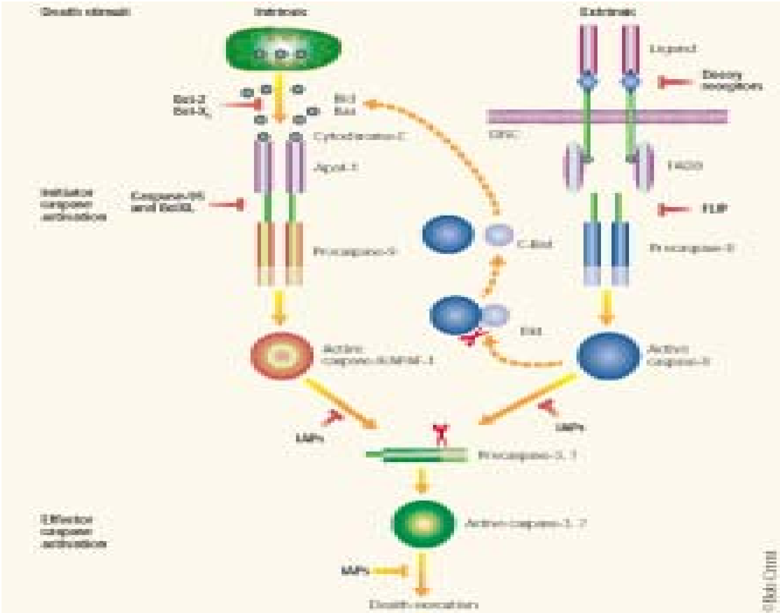
Protein

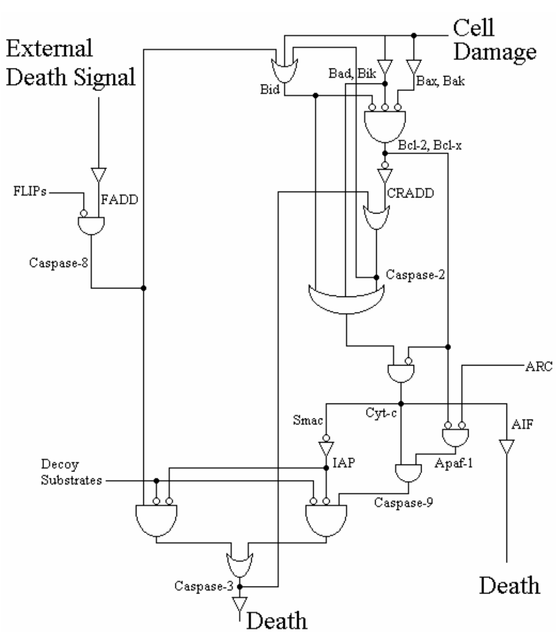
Protein-Protein Interaction

Metabolic networks

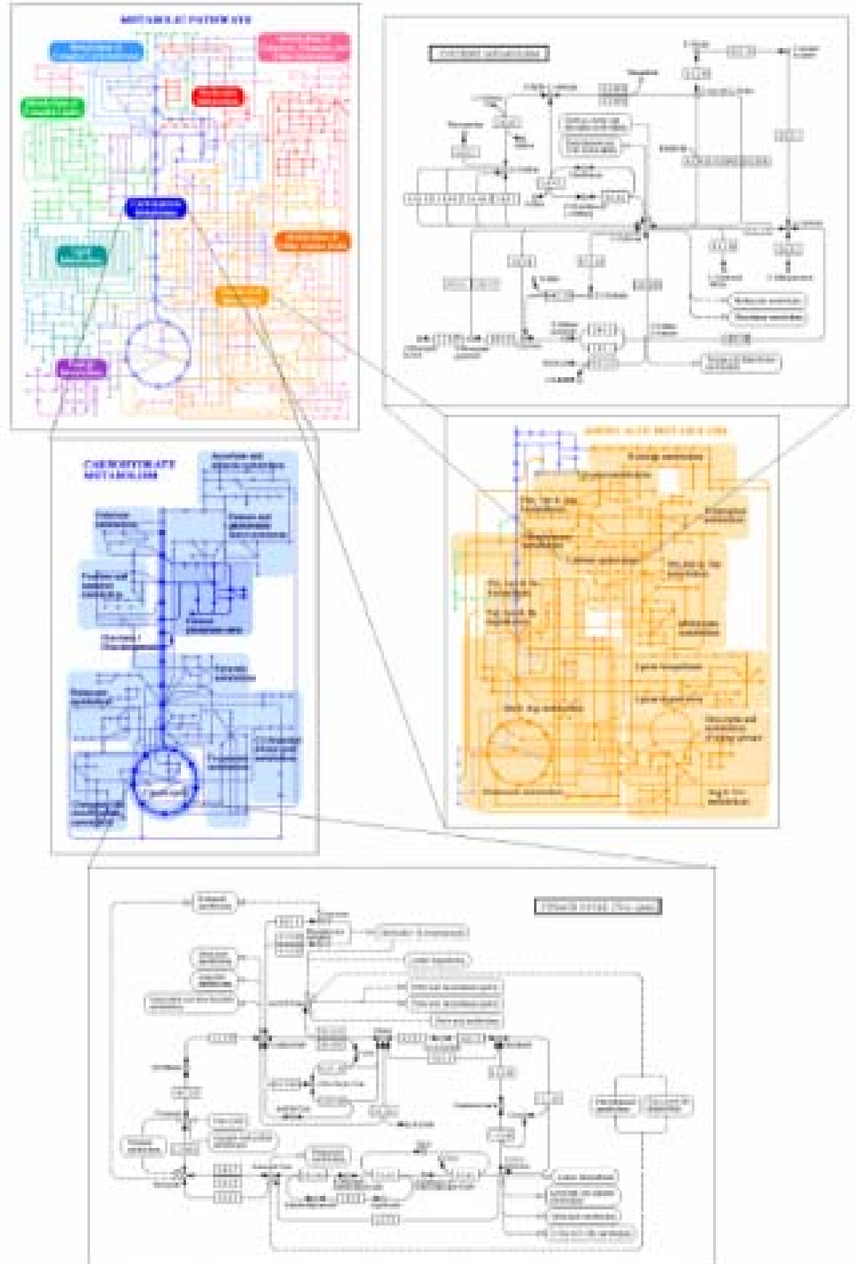
Bio-chemical interaction

Apoptosis





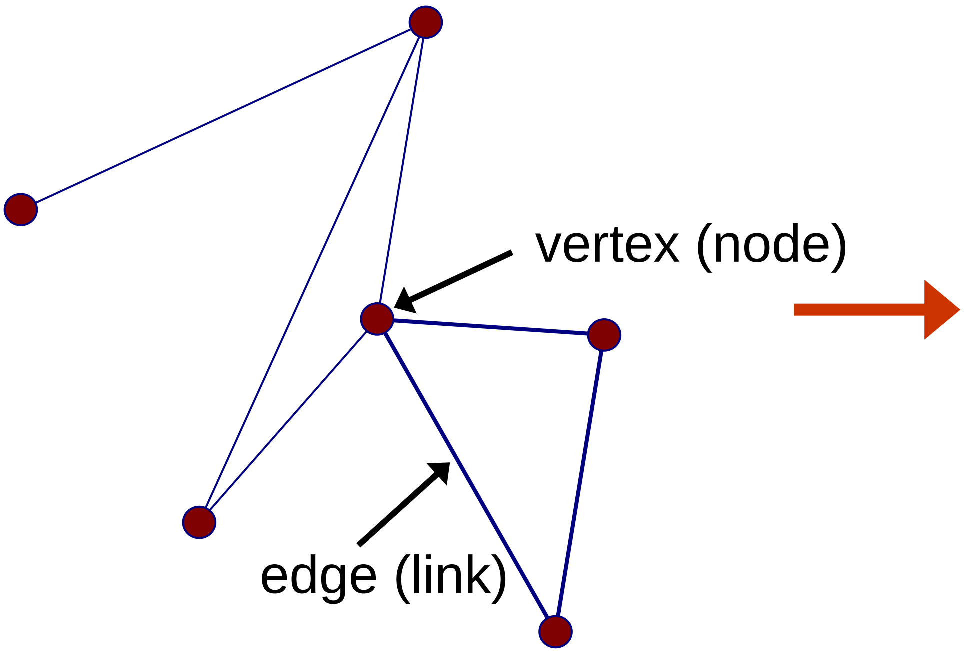
Metabolic networks



Introduction : basic concept

$G \equiv G(V, E)$: a network consists of vertices (nodes) and edges (links).

- weighted or unweighted : existence of weight on links
- directed or undirected : possible direction of information transfer

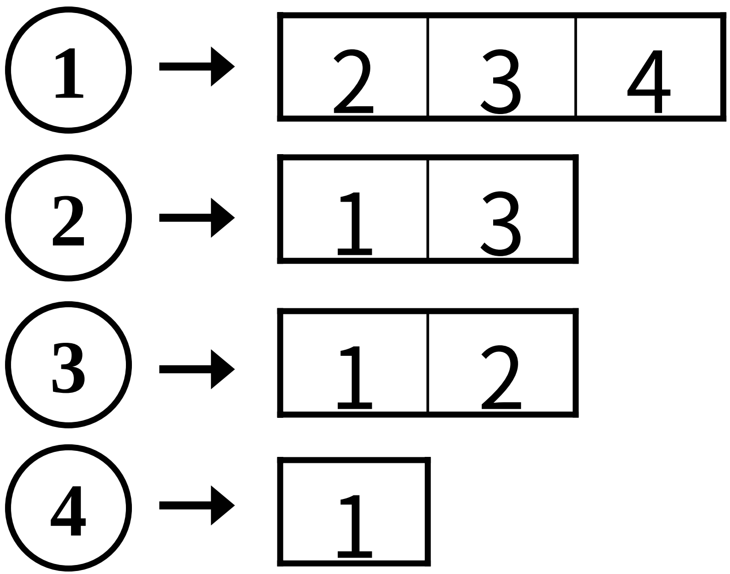


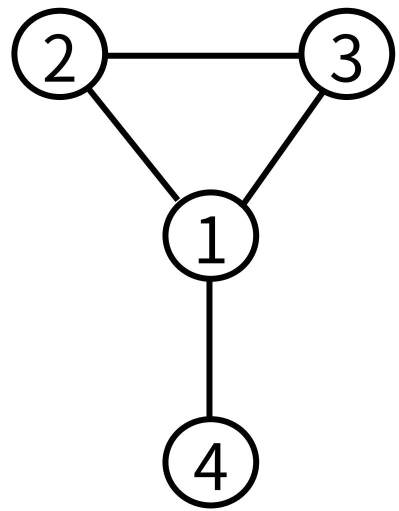
How to implement with computer language?

- **container construction**
- **algorithms for network properties**

Implementation of Network Structure

- **adjacency list**





- **adjacency matrix**

$$A_{ij} = \begin{cases} 1 & if\ \ vertex\ i\ and\ \ j\ is\ connected \\ 0 & otherwise. \end{cases}$$

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{vmatrix}$$

Introduction : quantities of interest

- Local properties of network
  - Degree
  - Clustering Coefficient
  - Average Nearest-Neighbor Degree

- Global properties of network
  - Betweenness Centrality
  - Load
  - Closeness Centrality
  - Mean Distance, etc.

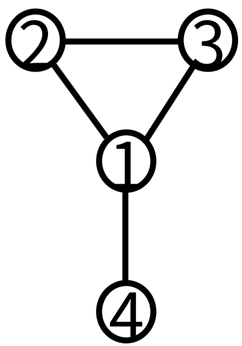
Degree & Clustering Coefficient

- **Degree (k_i) :** the number of linked neighbors of a vertex i

- **Clustering Coefficient (C)**

$$C = \frac{\text{\# of links between neighbors}}{\text{\# of possible neighbor pairs}}$$

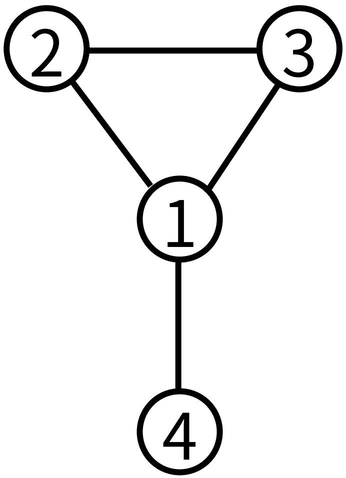
- Clustering coefficient for a network is the average of clustering coefficients of whole vertices.



$$C(1) = \frac{1}{3}$$

Average Nearest-Neighbor Degree : K_{nn}

- $k_{nn}(i) = \langle k_j \rangle$ (j : nearest-neighbor of vertex i)



$$k_{nn}(1) = \frac{k_2 + k_3 + k_4}{k_1}$$

$$= \frac{2+2+1}{3} = \frac{5}{3}$$

- **k_{nn}(k) represents the degree correlation.**
  - assortative, dissortative, neutral

Betweenness Centrality

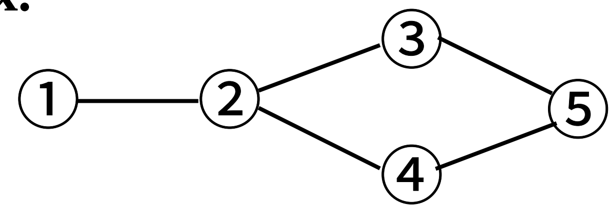
Definition

- $$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

- $$C_B(v) = \sum_{s \neq t \in V, v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

(used in Physics)

ex.



$$C_B(3) = 2 \times \frac{\sigma_{15}(3)}{\sigma_{15}} + 2 \times \frac{\sigma_{25}(3)}{\sigma_{25}} \quad (+8)$$

$$= 2 \times \frac{1}{2} + 2 \times \frac{1}{2} \quad (+8)$$

$$= 2 \quad (+8)$$

σ_{st} : the number of shortest paths from **s** to **t**.

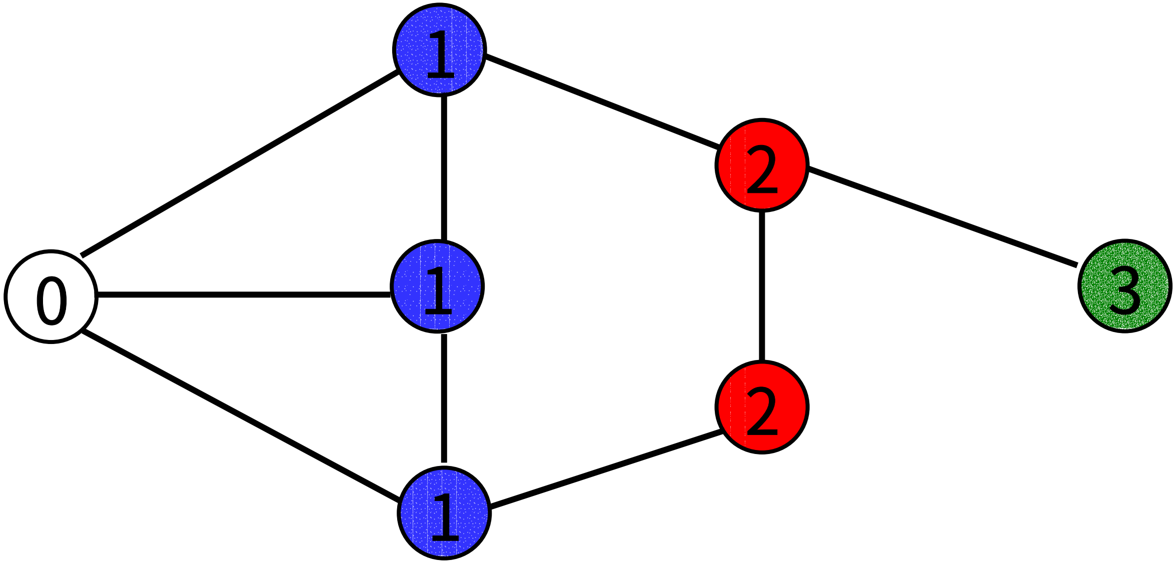
$\sigma_{st}(v)$: the number of shortest paths from **s** to **t** through **v**.

Betweenness : shortest paths

- **We have to know shortest paths to calculate the betweenness.**

- **Breadth-First searching method :**

  Start at the source vertex, visit all the neighbors, visit all the neighbors of neighbors, etc..., until you have run out of neighbors to visit.



Closeness Centrality & others

- **closeness centrality : measures closeness of a vertex from others.**

$$C_C(i) = \sum_{j \in V} \frac{1}{d_{ij}}$$

d_{ij} **: distance from i to j**

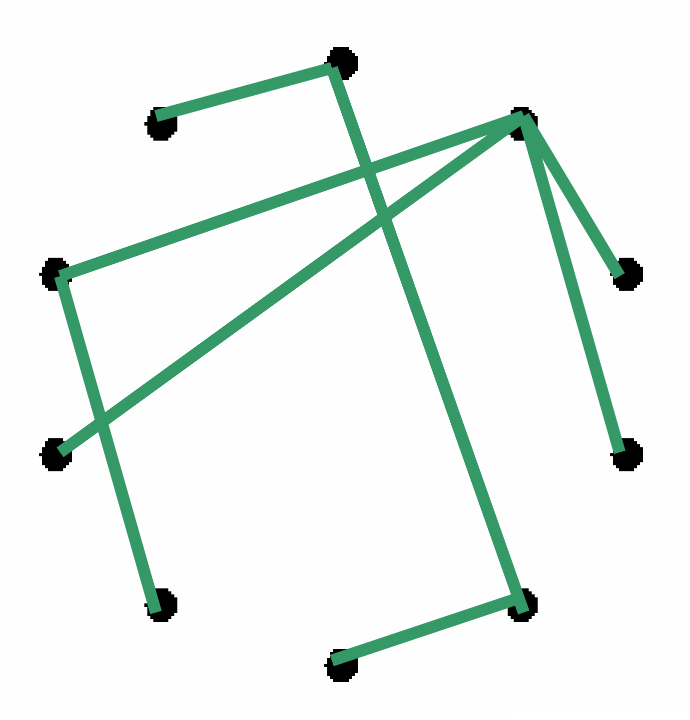
other definitions : $\frac{1}{\sum_{j \in V} d_{ij}}$, $\sum_{j \in V} d_{ij}$, etc.

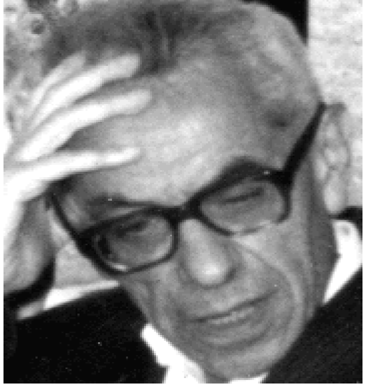
- **Mean distance and maximum distance are also related to closeness.**

Short history of
Complex Network Model

- Random Network model
- Small-World Network model
- Scale-Free Network model

Erdös-Rényi model
(1960)

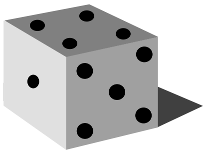




Pál Erdös
(1913-1996)

Connect with probability p

p=1/6
N=10
⟨k⟩ ~ 1.5



- Democratic

- Random

Degree distribution of a random graph

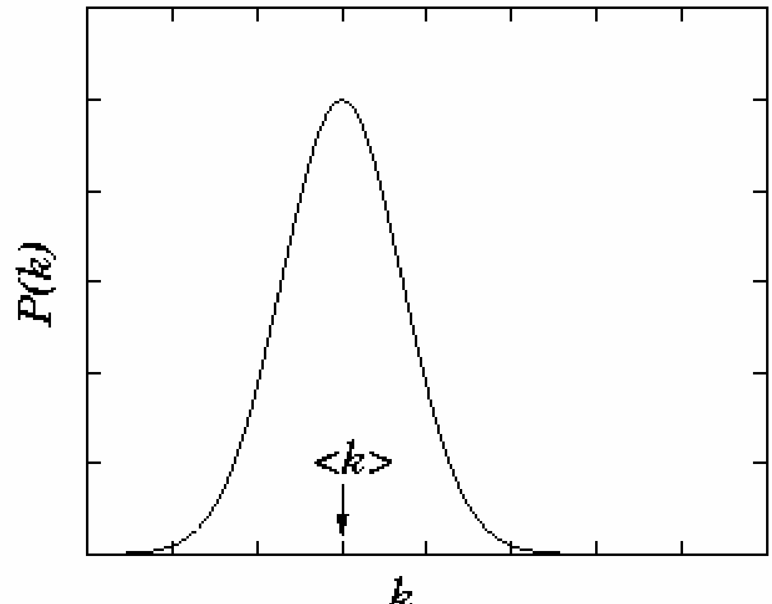
***P(k)*: the probability that a node has *k* links**

$$P(k) = C^{k}_{N-1}\, p^{k} (1-p)^{N-1-k}$$

For large N, *P(k)* can be replaced by a Poisson distribution:

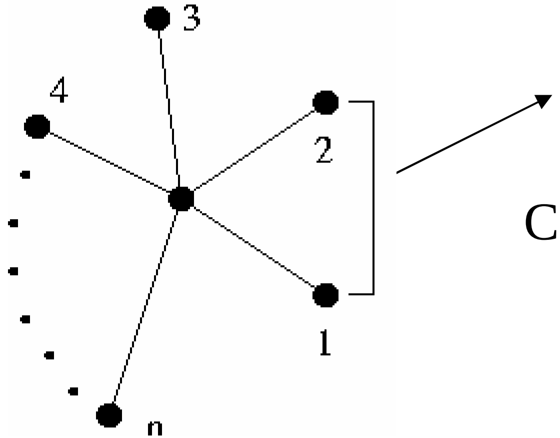
$$P(k) \sim e^{-\langle k \rangle} \langle k \rangle^{k} / k!$$

Poisson distribution



BUT THERE IS **Clustering**!!!

My friends will know each other with high probability!

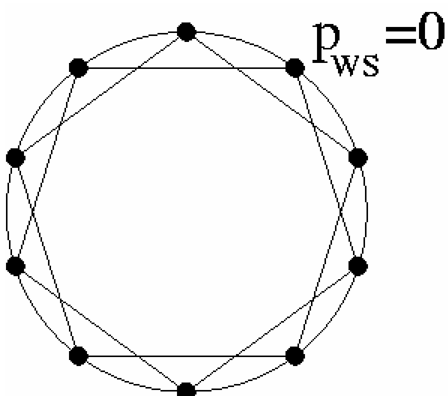


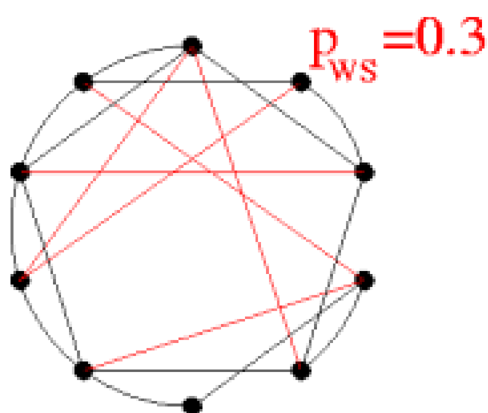
Probability to be connected $C \gg p$

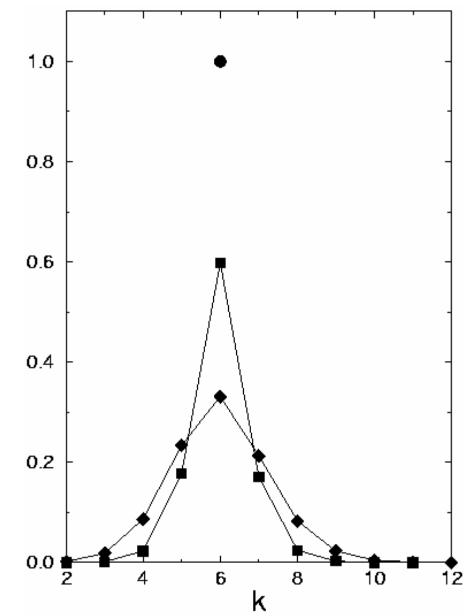
$$C = \frac{\text{\# of links between 1,2,\ldots n neighbors}}{n(n-1)/2}$$

Watts-Strogatz Small-World model

N nodes forms a regular lattice.
With probability p,
each edge is rewired randomly.



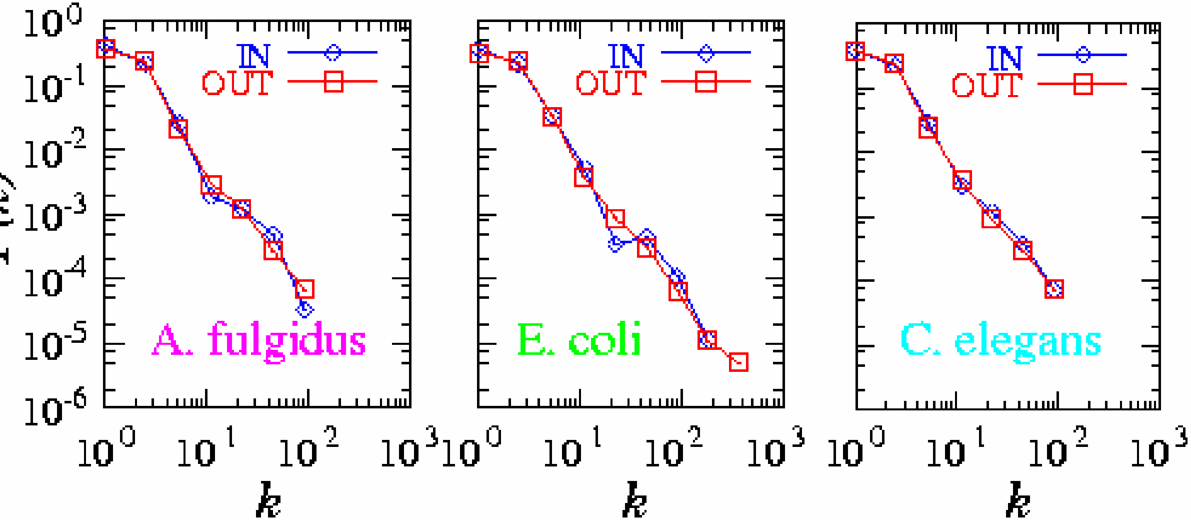




(Nature **393,** 440 (1998))

BUT THERE IS POWER-LAW!!!

Archaea Bacteria Eukaryotes



P(k)=degree (# of connection) distribution

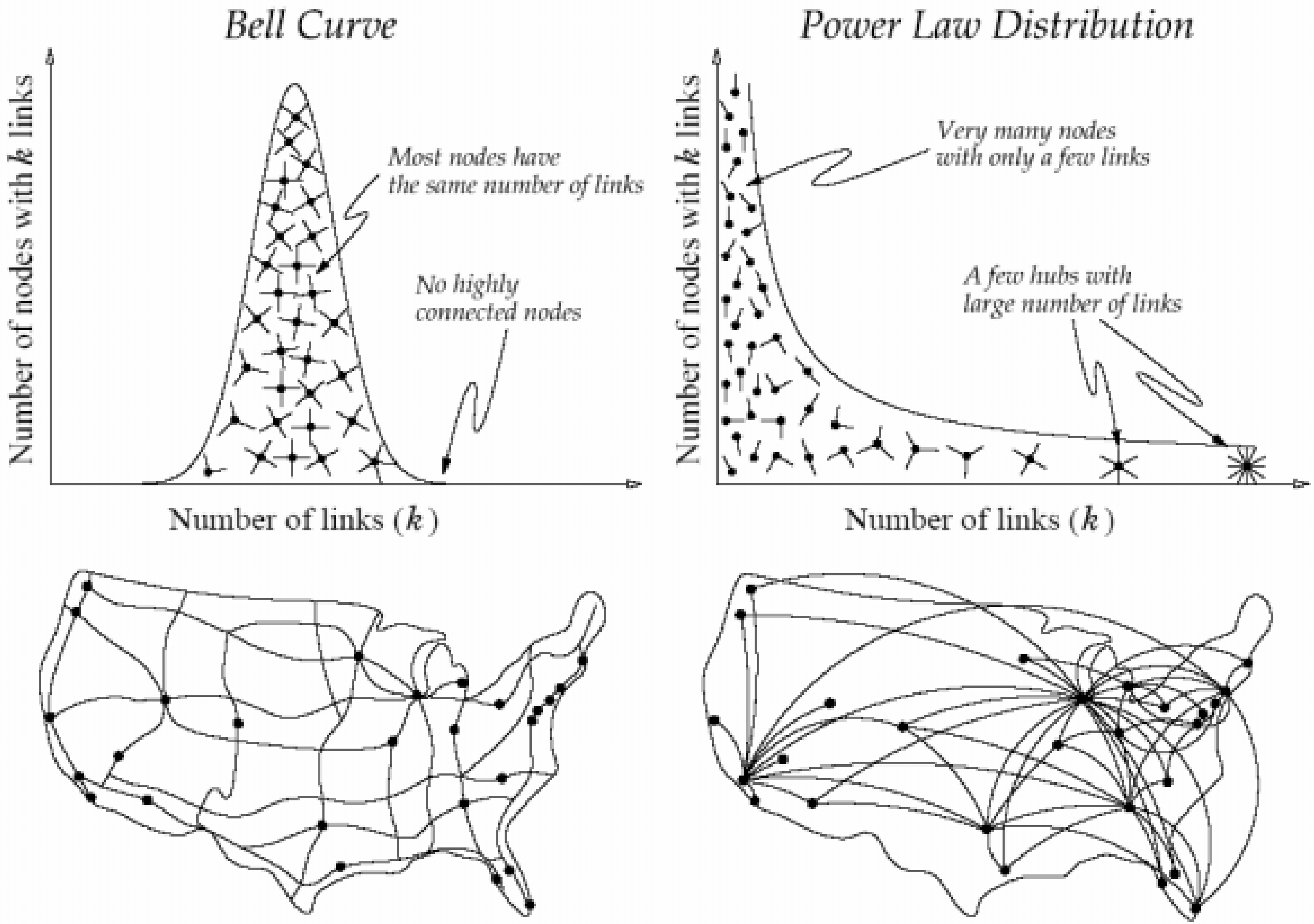
P(k) follows power-law and they are **scale-free** networks!

Scale-free network

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

Network		
WWW	2.45	2.1
WWW	2.38	2.1
WWW, site	2.72	2.1
Internet, domain*	2.2	2.2
Internet, router*	2.48	2.48
Internet, router*	2.4	2.4
Movie actors*	2.3	2.3
Coauthors SPIRES*	1.2	1.2
Coauthors NEURO*	2.1	2.1
Coauthors MATH*	2.5	2.5
Sexual contacts*	3.4	3.4
Metabolic, E. coli*	2.2	2.2
Protein, S. cerev.*	2.4	2.4
Protein, S. cerev.*	2.3	2.3
Ythan estuary*	1.05	1.05
Silwood park*	1.13	1.13
Citation		3
Phone-call	2.1	2.1
Words occurrence*	2.7	2.7
Words synonyms*	2.8	2.8

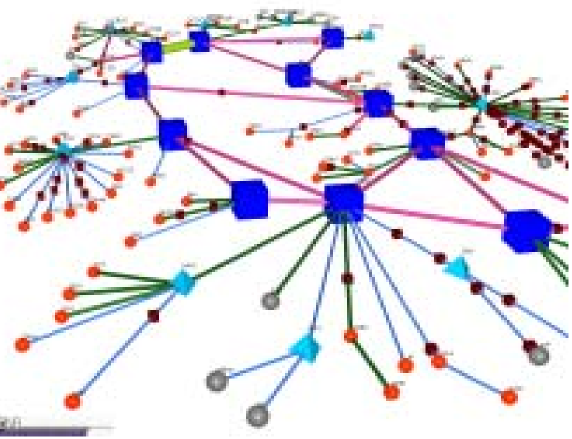
Two kinds of networks!



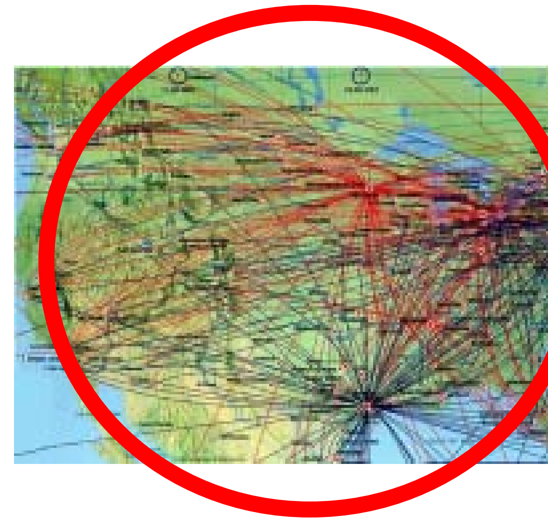
Random Network
Small-World Network

Scale-Free Network

Through a series of investigations, it is found that
Real-world network looks like inhomogeneous air-line map
They are Scale-Free Networks!



=



Examples: Too many to write down on this page... :-)

Exceptions: Power-grid, food-web, ...

Scale-free model

(1) Networks continuously expand by the addition of new nodes

WWW : addition of new documents
Citation : publication of new papers

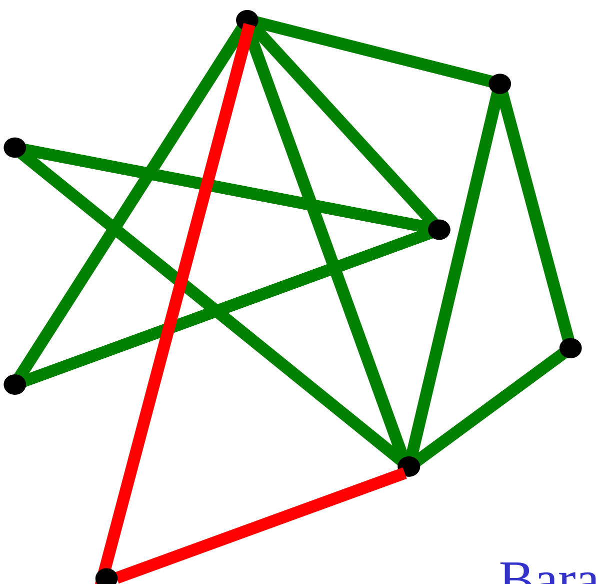
(2) New nodes prefer to link to highly connected nodes.

WWW : linking to well known sites
Citation : citing again highly cited papers

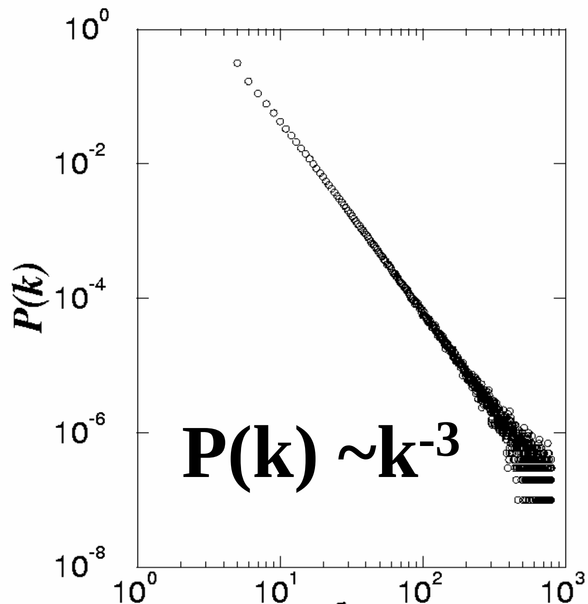
GROWTH:
add a new node with m links

PREFERENTIAL ATTACHMENT: th
probability that a node connects to
node with k links is proportional to k

$$\Pi(k_i) = \frac{k_i}{\Sigma_j k_j}$$



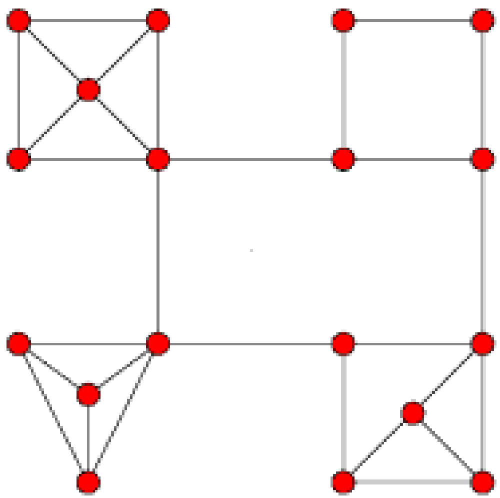
Barabási & Albert,
Science **286**, 509 (1999)

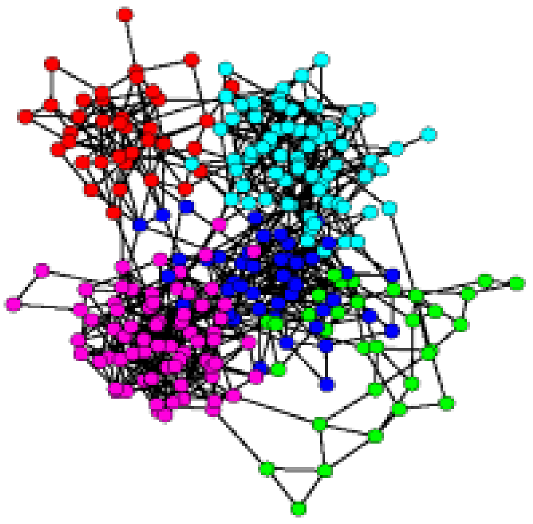


BUT THERE IS Modularity!!!

- High C → real networks are fragmented into group or modules
  - Internet: Vasquez, Pastor-Satorras, Vespignani (2001).
  - Society: Granovetter, M. S. (1973) ; Girvan, M., & Newman, M.E.J. (2001); Watts, D. J., Dodds, P. S., & Newman, M. E. J. (2002).
  - WWW: Flake, G. W., Lawrence, S., & Giles. C. L. (2000).
  - Biology: Hartwell, L.-H., Hopfield, J. J., Leibler, S., & Murray, A. W. (1999).

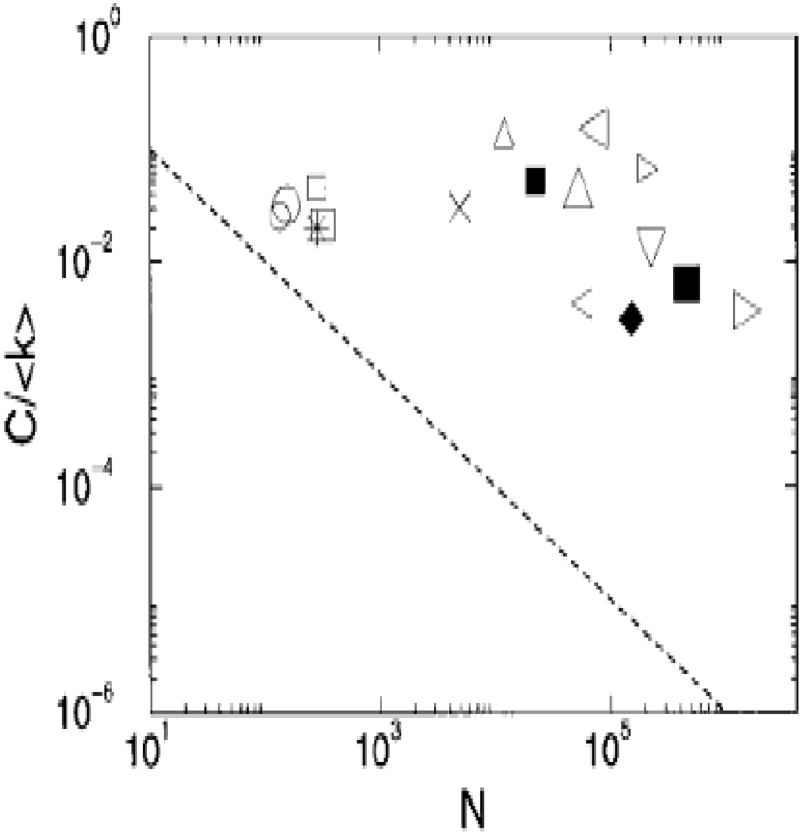
- Traditional view of modularity:



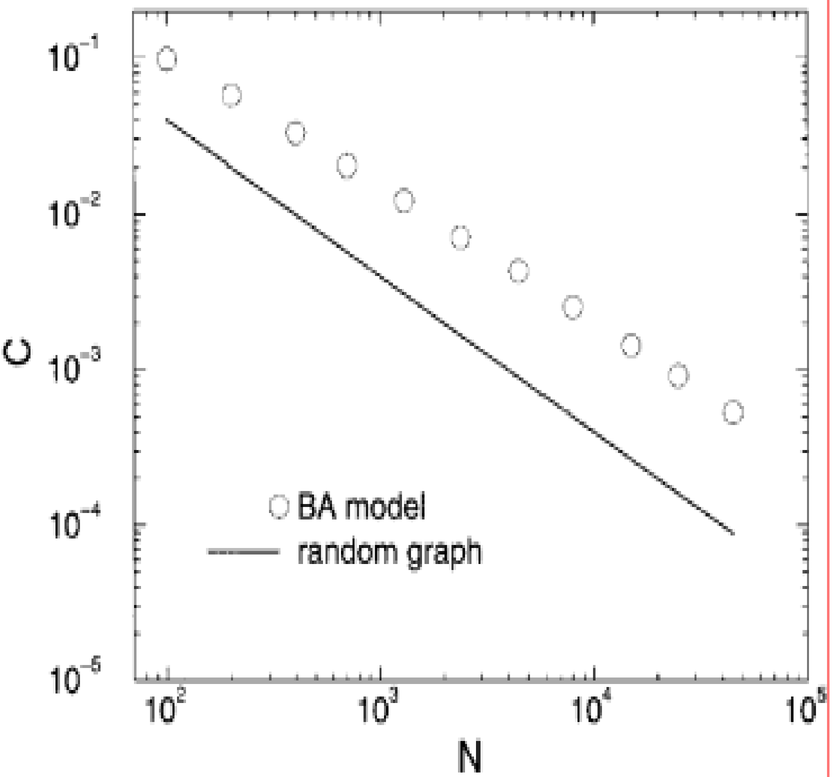


Clustering in non-biological networks

C is independent of N



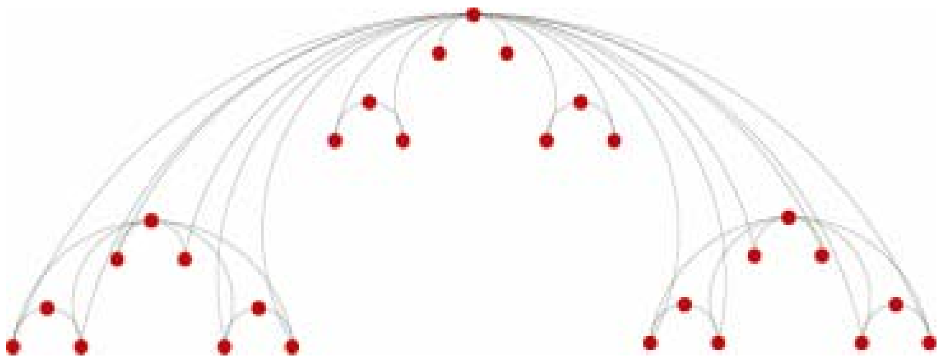
C decreases with N



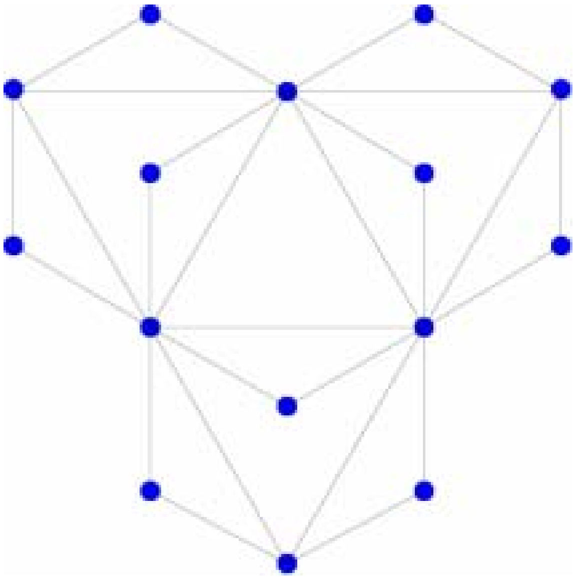
$C_i = 2n_i / k_i(k_i - 1)$

Combining Modularity and the Scale-free Property

Deterministic Scale-Free Networks



Barabási, A.-L., Ravasz, E., & Vicsek, T. (2001) *Physica A* **299**, 559.



Dorogovtsev, S. N., Goltsev, A. V., & Mendes, J. F. F. (2001) cond-mat/0112143. (DGM)

BUT THERE IS Hierarchy!!!

- **WWW:** **Moses and Eckmann, 2001.**
- **Internet:** **Vazquez, Pastor-Satorras, Vespignani, cond-mat/0112400**

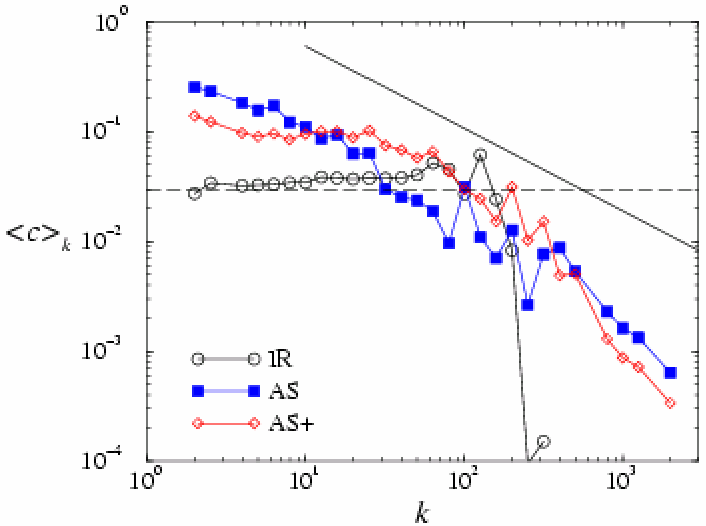


Figure 5: Average clustering coefficient as a function of the node connectivity for the AS, AS+, and IR maps. The solid line is given by the power law decay $\langle c \rangle_k \sim k^{-0.75}$. The horizontal dashed line marks the average clustering coefficient $\langle c \rangle = 0.03$ computed for the IR map.

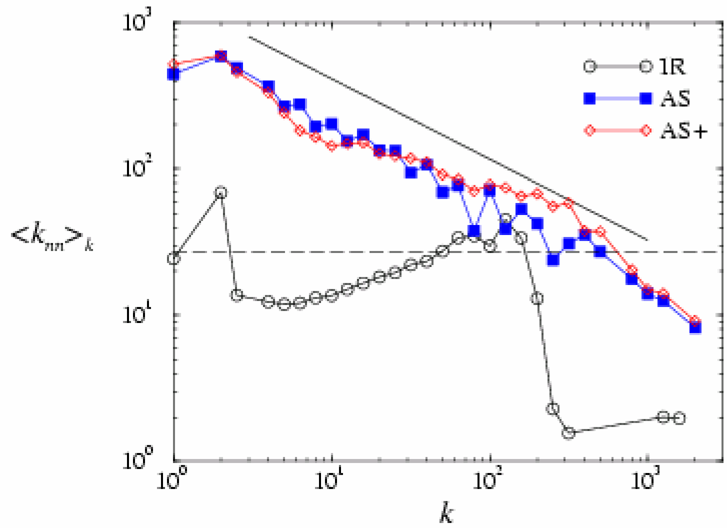
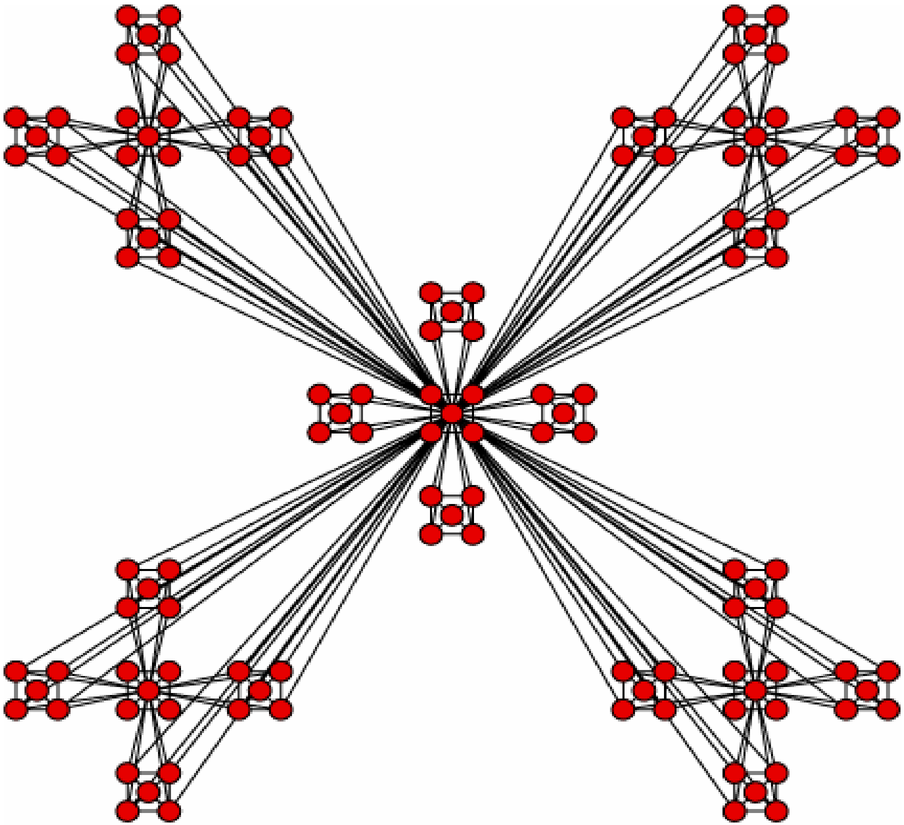


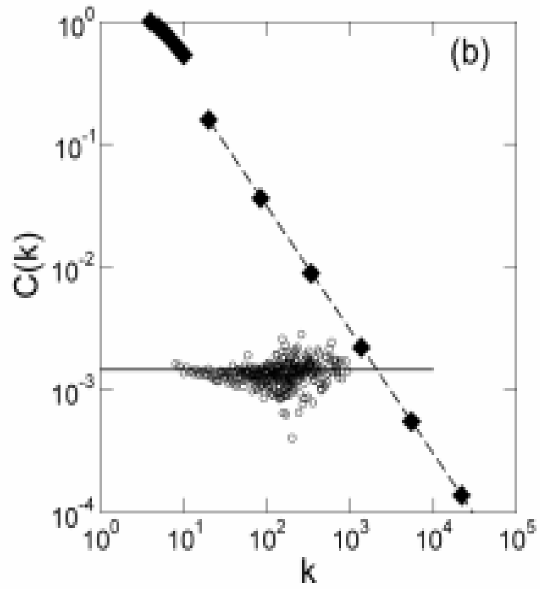
Figure 6: Nearest neighbors average connectivity for the AS, AS+, and IR maps. The solid line is given by the power law decay $\langle k_{nn} \rangle_k \sim k^{-0.55}$. The horizontal dashed line marks the value in the absence of correlations, $\langle k_{nn} \rangle_k^0 = \langle k^2 \rangle / \langle k \rangle = 26.9$, computed for the IR map.

Hierarchical Networks



1. Clustering coefficient depends on *k*

$$C(k) \sim k^{-1}$$

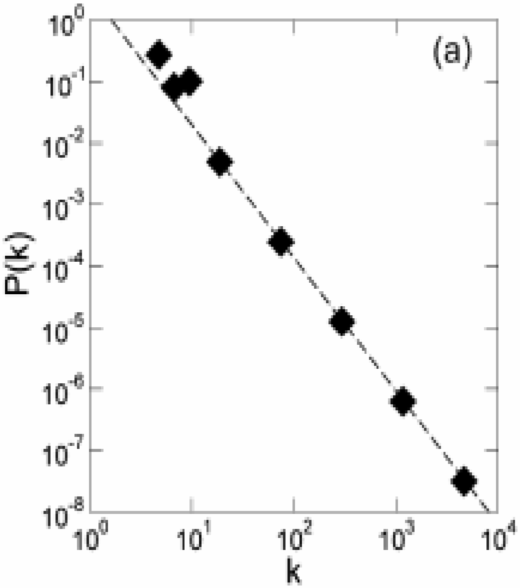


$$C(k)= \frac{\#\ \text{links between k neighbor}}{k(k-1)/2}$$

Properties of hierarchical networks

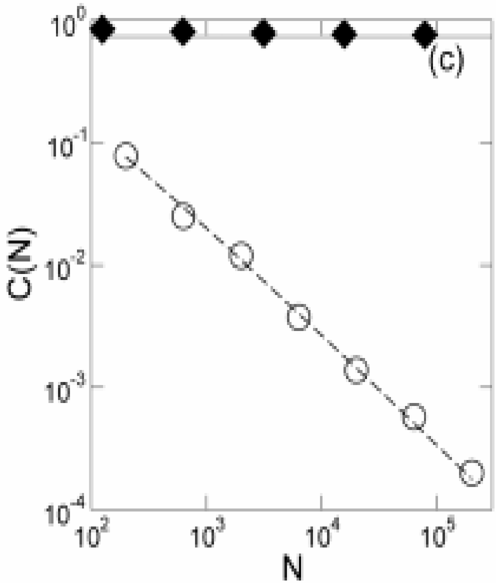
2. Scale-free

$$\gamma = 1 + \frac{\ln 5}{\ln 4}$$
$$= 2.161$$

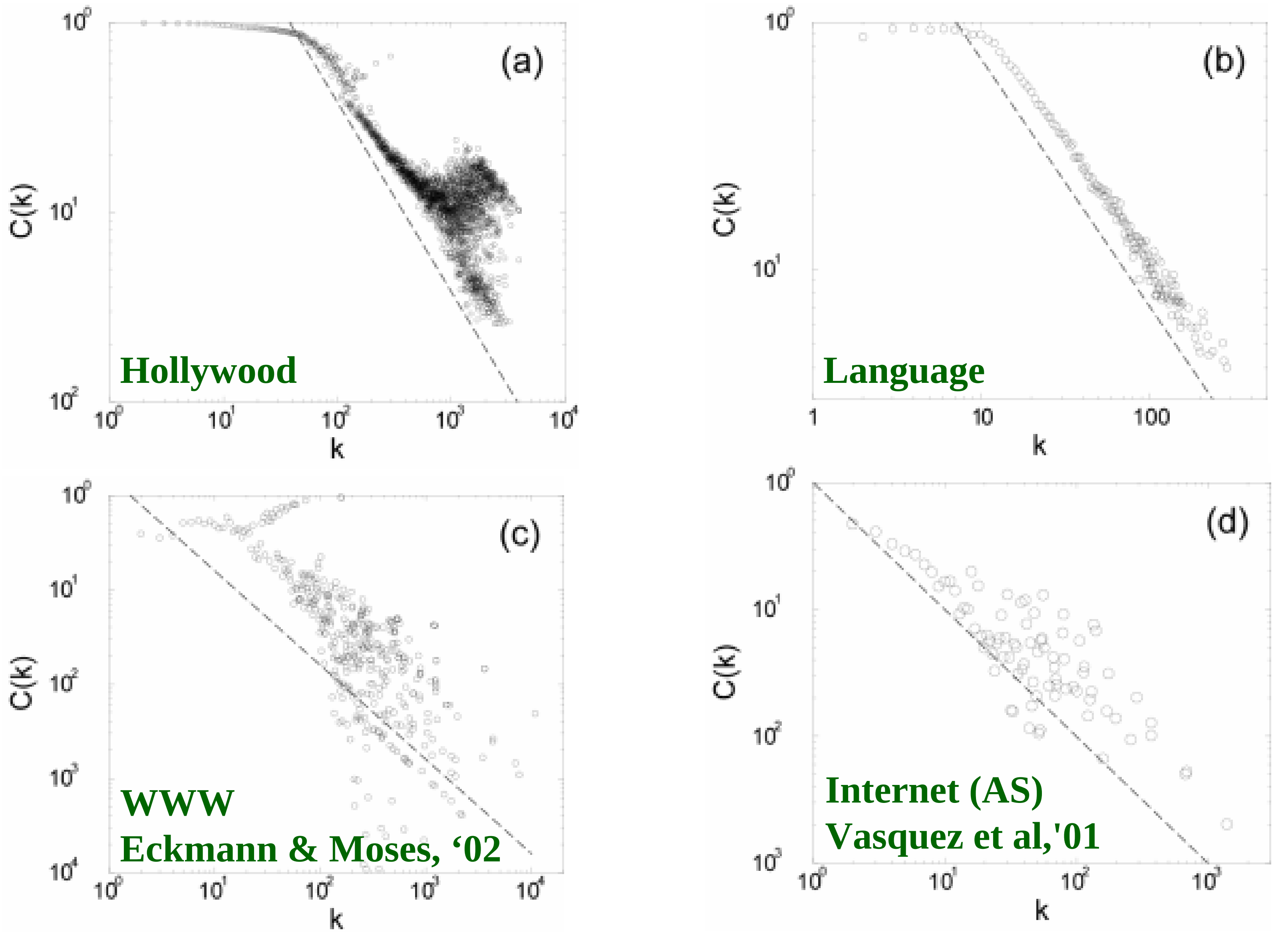


3. Clustering coefficient independent of N

$$C(N) = \text{const.}$$

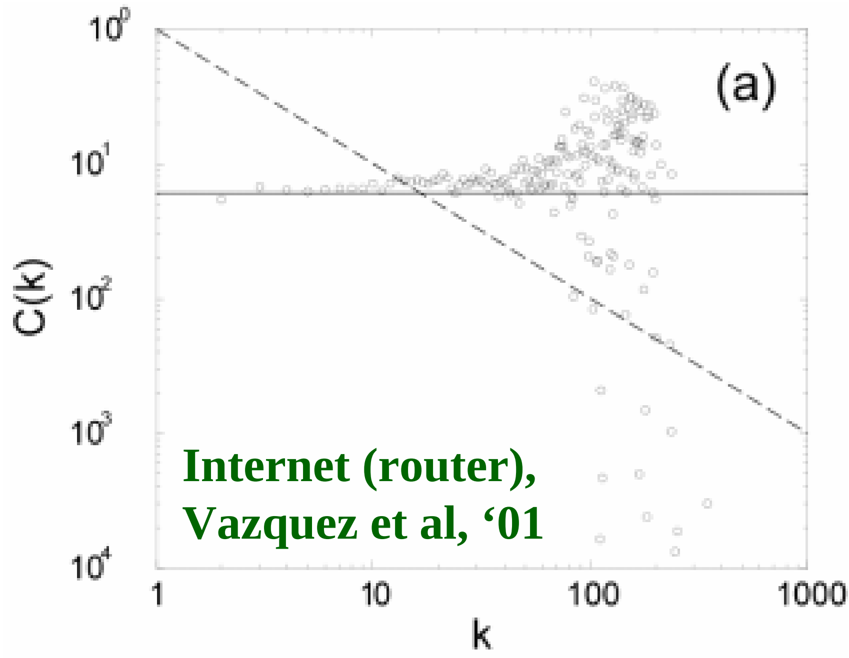


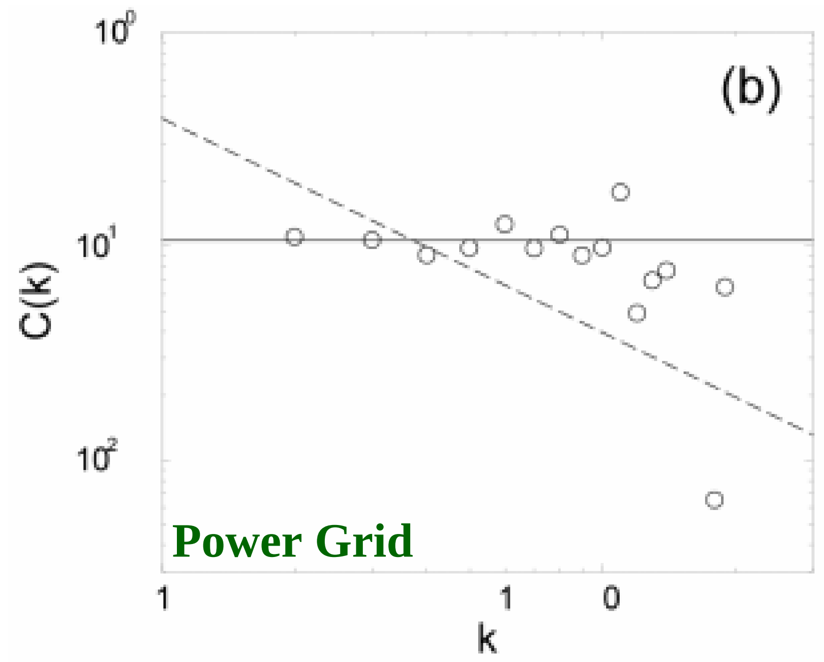
Real Networks



E. Ravasz and A.-L. Barabási, cond-mat/0206130

Exceptions:
Geographically Organized Networks:





Common feature:

economic pressures towards shorter links

Is hierarchy present in network models?

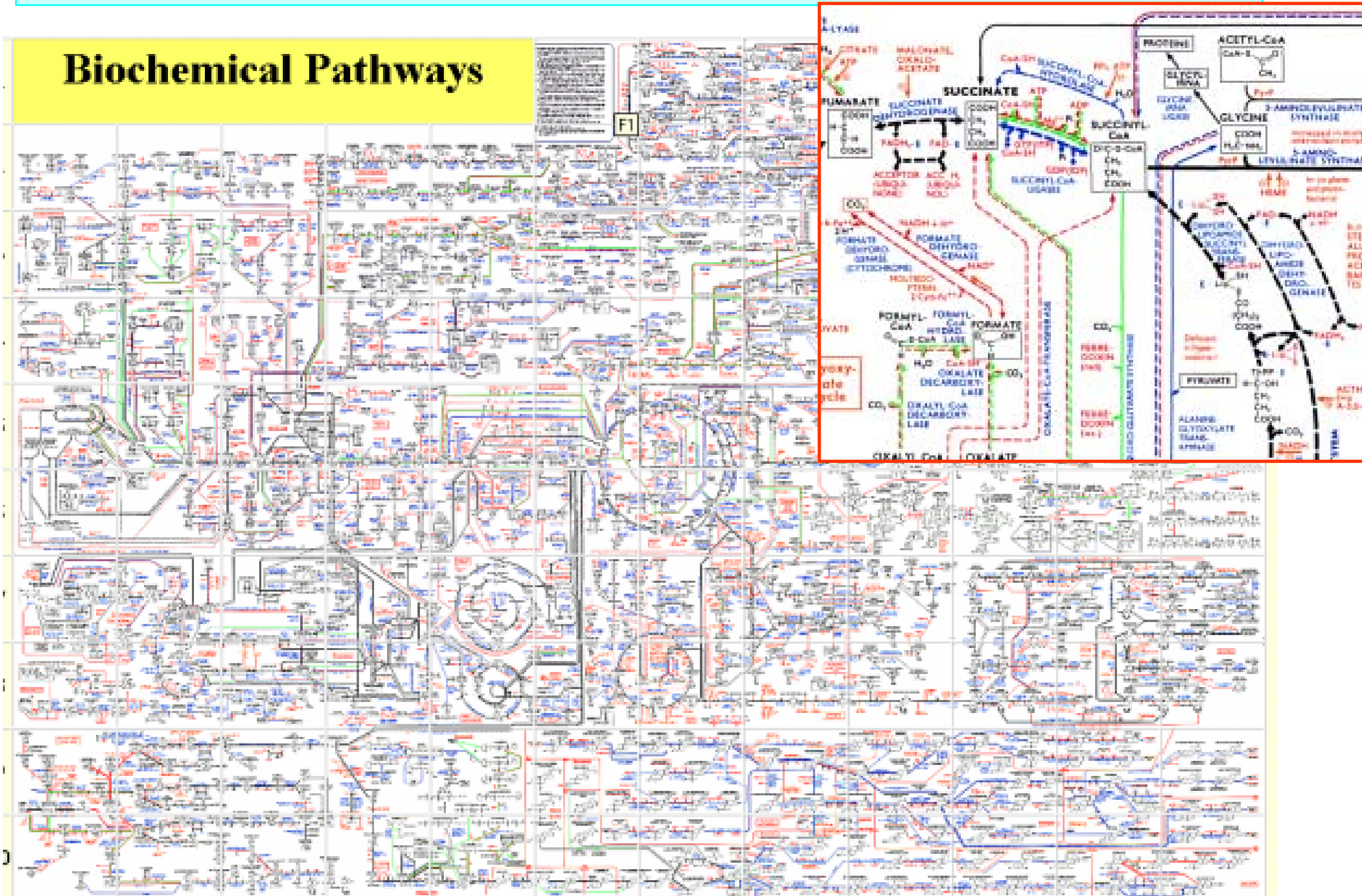
NO:

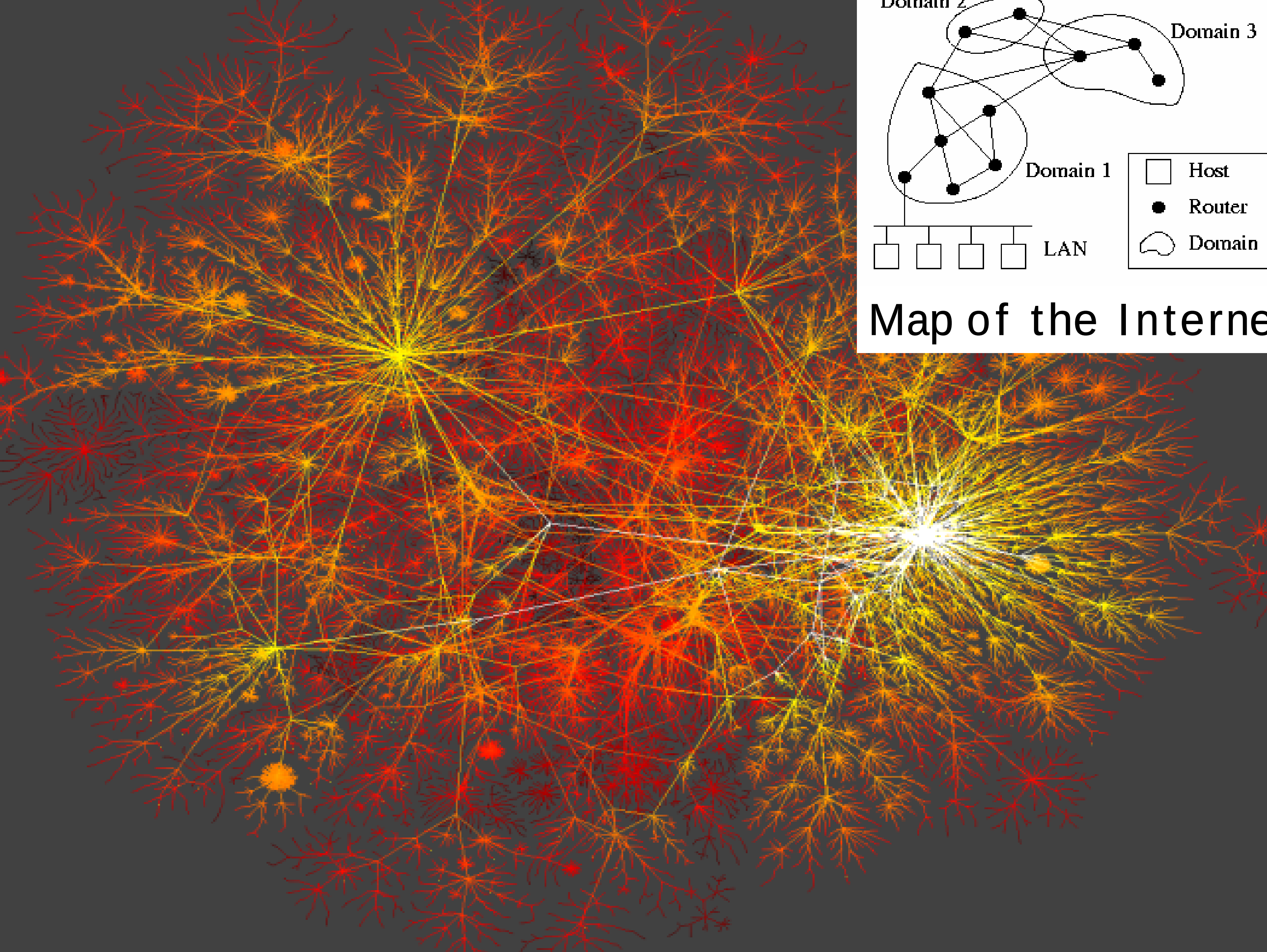
-Scale-free model (Barabasi & Albert,1999)
-Erdos-Renyi model (1959)
-Watts-Strogatz (1998)

YES:

-Dorogovtsev, Goltsev, Mendes, 2001 (determ.)
-Klemm and Eguiluz, 2002
-Vazquez, Pastor-Satorras,Vespignani (2001)
$\Rightarrow$ Bianconi & Barabasi (fitnesss model) (2001)
-Szabo, Alava & Kertesz, cond-mat/0208551:
$\Rightarrow$Holme-Kim model (2002).

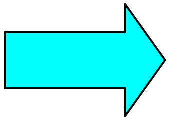
Isn't the real network still too complex??





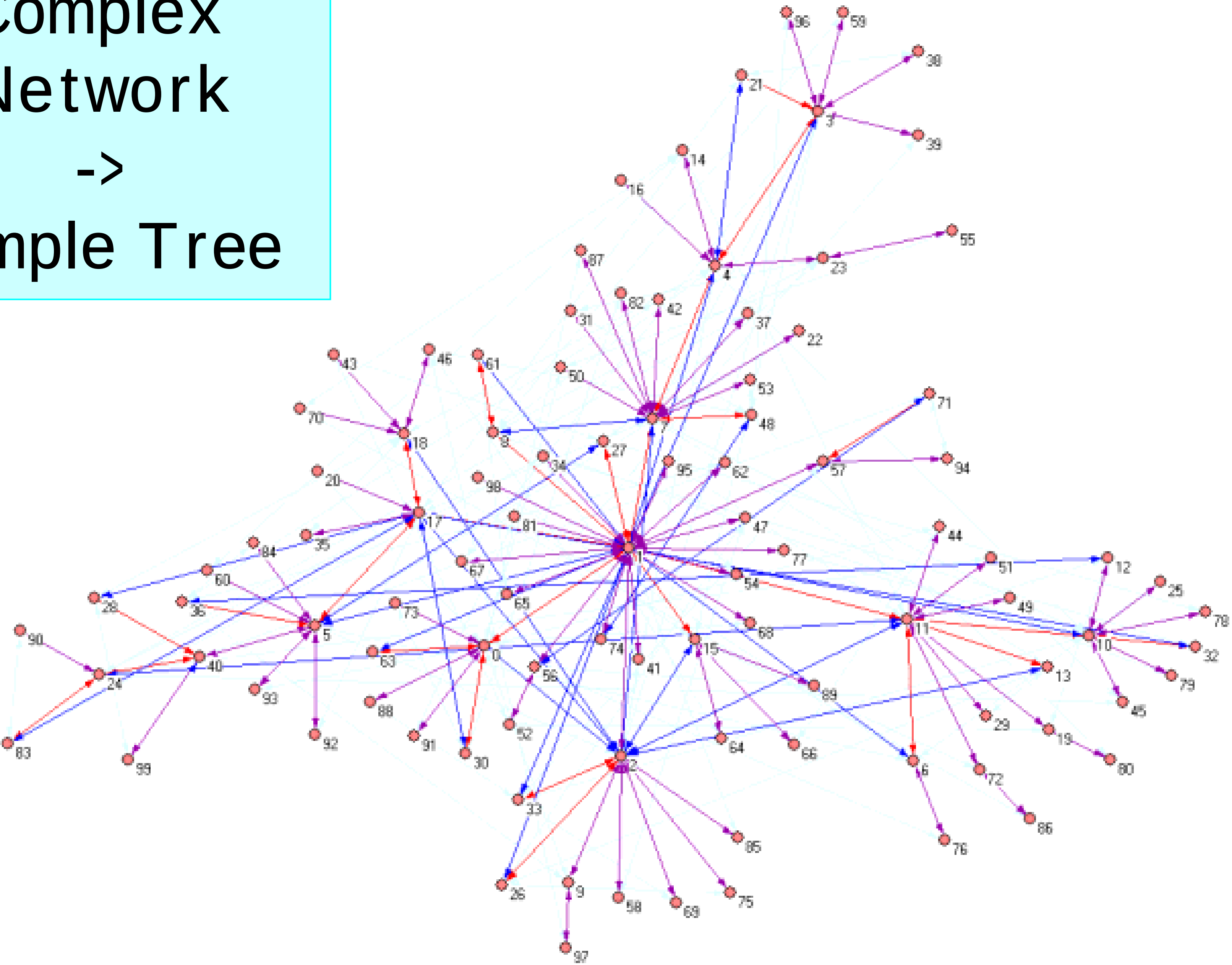
Map of the Interne

Let's make it simple!



Find the skeleton of complex network!

Complex Network -> Simple Tree



Network vs. Tree

- **A tree is the simplest network.**
  - A tree is a connected acyclic network; there is no cycle (loop) in tree.

- **A spanning tree is a subgraph connecting all the vertices with no loop.**
  - Optimization: cost minimization while connecting the whole system

    Minimum spanning trees of weighted networks

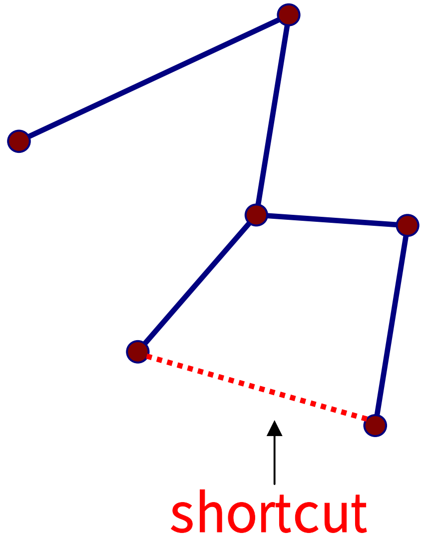
  - Percolation approach

    A spanning tree has the minimum number of links to connect all vertices.

SF Network ≟ (SF) Tree + Shortcuts

- **Question: does (SF) network decompose into (SF) tree and shortcuts ?**

  - method of decomposition
  - properties of the extracted trees
  - meaning of the extracted trees in entire networks



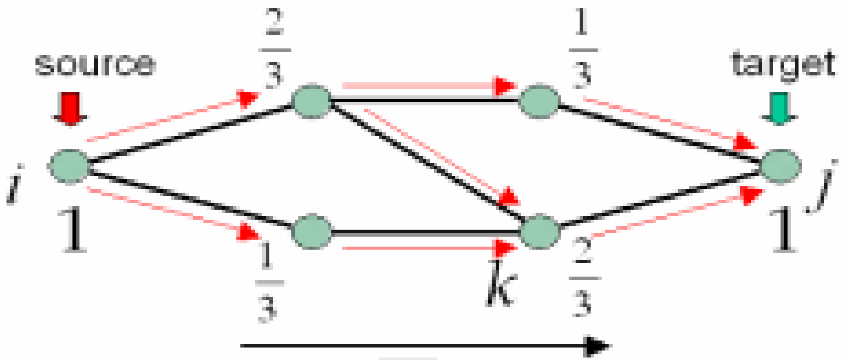
- **Strategy to extract tree structures from networks:**

1. Determine the importance of each links.
   - The edge-betweenness is a good candidate.

2. Assign the weight (the importance) to each links.

3. Extract tree structures with maximum total weight.
   - Minimum spanning tree technique is used.

Edge-Betweenness Centrality



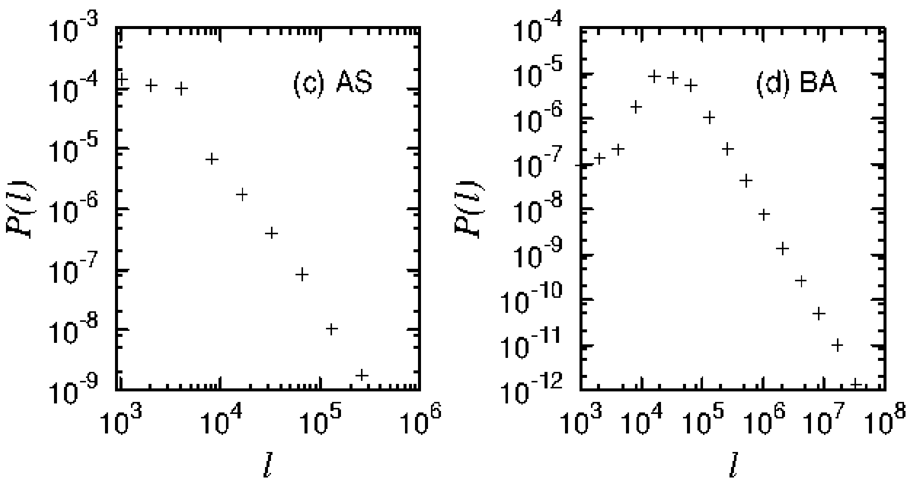
$$b(k)=\sum_{i,j} b_{i\to j}(k)$$

Definition:

$$l_e=\sum_{i\neq j} l_e(i,j)=\sum_{i\neq j}\frac{c_e(i,j)}{c(i,j)}$$

the effective number of paths through an edge ***e***

(**<u>average traffic through an edge</u> *e***)



- **The distribution of edge-betweenness shows scale-free or similar behavior in asymptotic limit.**
  - The selection of links with large edge-betweenness is likely to construct structures following scale-free behavior.

Minimum Spanning Tree : communication kernel

- **We extract embedded trees using the minimum spanning tree technique.**

  Kruskal Algorithm:

  1. Select a link with largest edge-betweenness.
  2. If the endpoints of the links are disconnected in tree, add the link to tree. Otherwise, forget it. Then, return to the step 1 until the tree contains all connected nodes in networks.

- **Extracted trees are the kernels of SF networks in communication.**

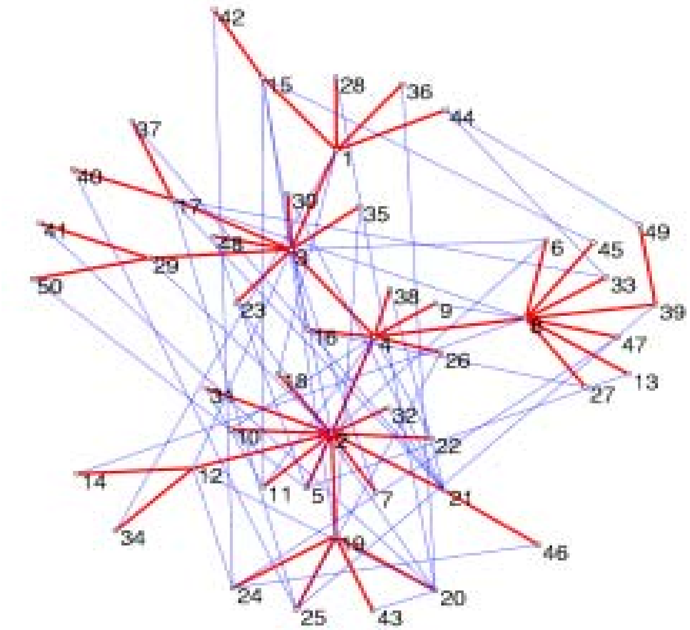
network	$<k>$	f_l	f_b
coauthorship	11.58	0.16	0.46
internet AS	3.88	0.50	0.68
PIN	6.55	0.30	0.54
PIN model	$O(1)$	0.50	0.74
BA model (m=2)	4	0.50	0.71
fitness model	4	0.50	0.61
static model	4	0.50	0.66
adaptation model	$O(1)$	0.11	0.57

$f_l < f_b$

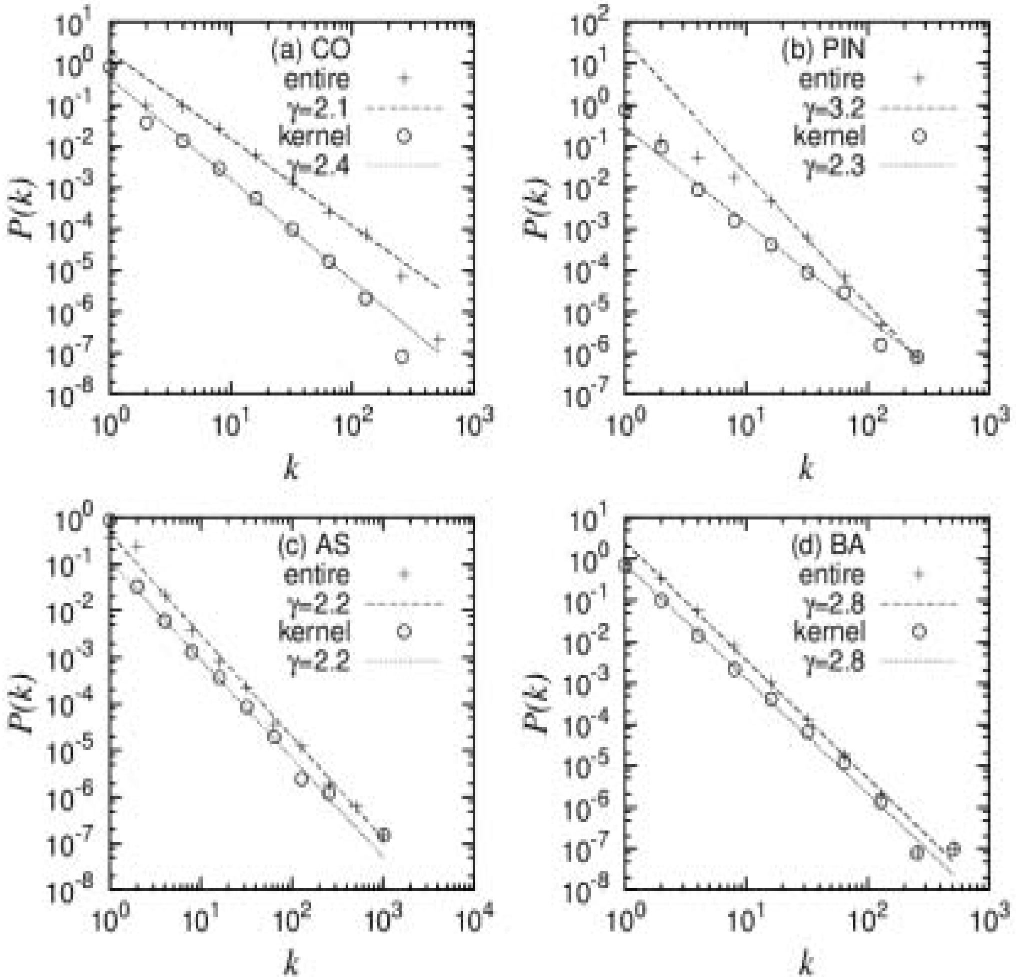
n: # of nodes
m: # of links

$$f_l = (n-1)/m$$

f_b : ratio of edge-betweenness summation of selected links to entire links.



Results : degree distribution



The embedded trees show the scale-free behavior.

SF network

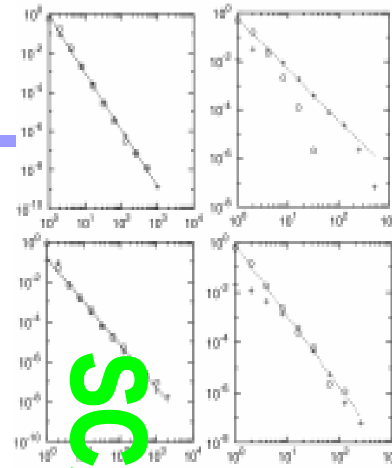
=

SF tree

+

Shortcuts

Other methods : making trees



1. PERC : random removal of edges with retaining connection between vertices.

Staring from an undamaged network,

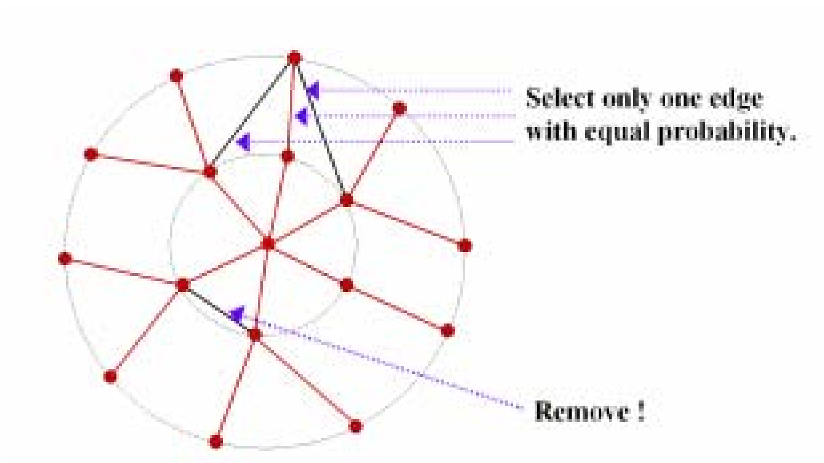
At each step one of remaining edges is randomly chosen for removal.

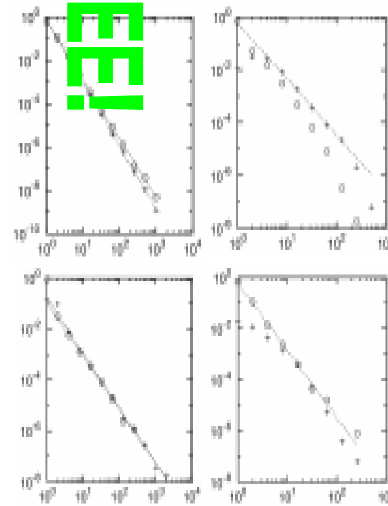
If the network becomes disjoint after the removal, the edge is restored.

(A. S. Ioselevich and D. S. Lyubshin, cond-mat/0306177)

Again SCALE-FREE!

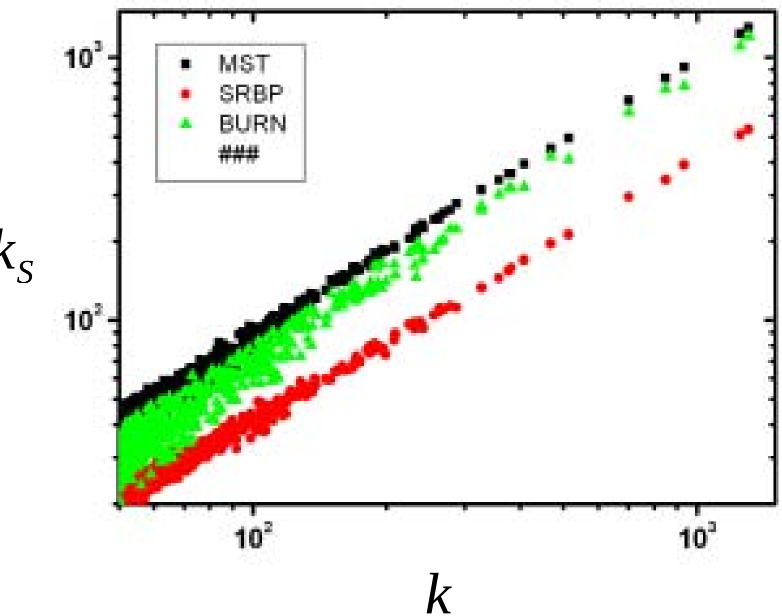
2. BURN: Burning from a randomly selected vertex



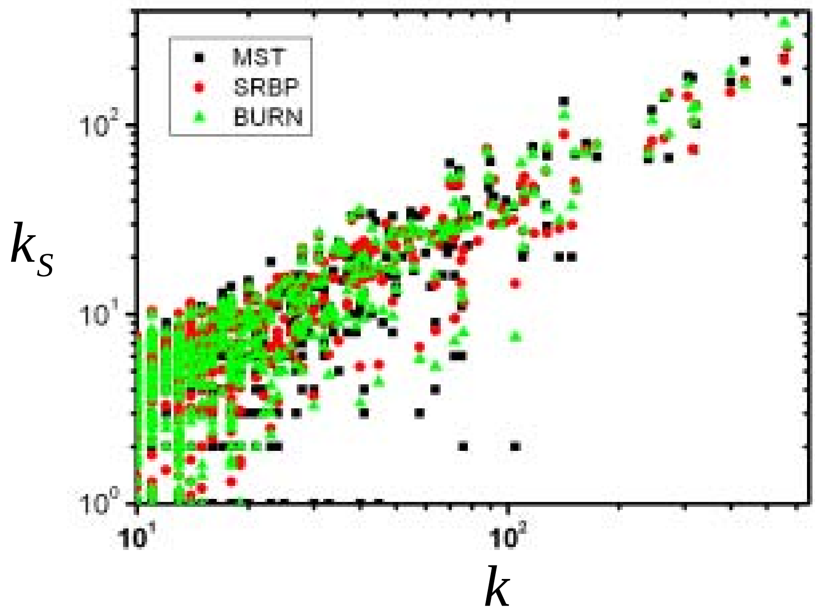


Discussion : strong correlation

BA model



Internet AS



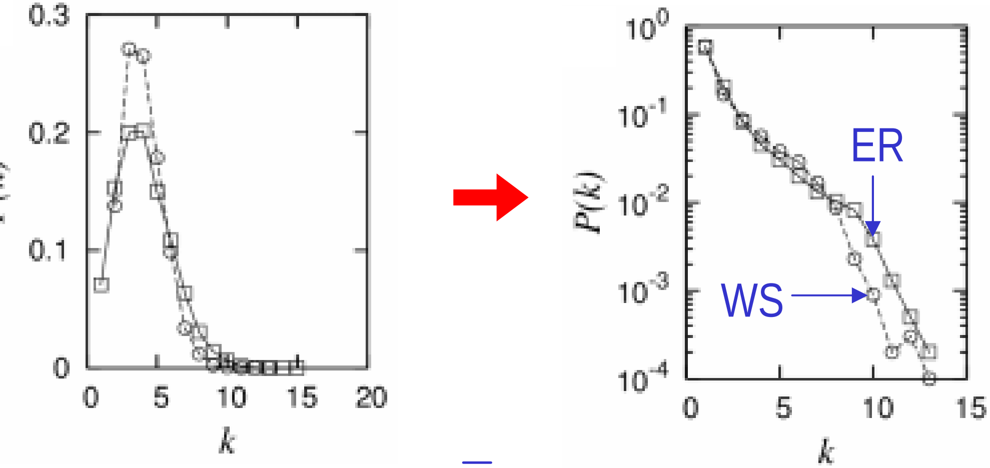
$$k_S \propto k^{\alpha}$$

$$\alpha \approx 1.0(1)$$

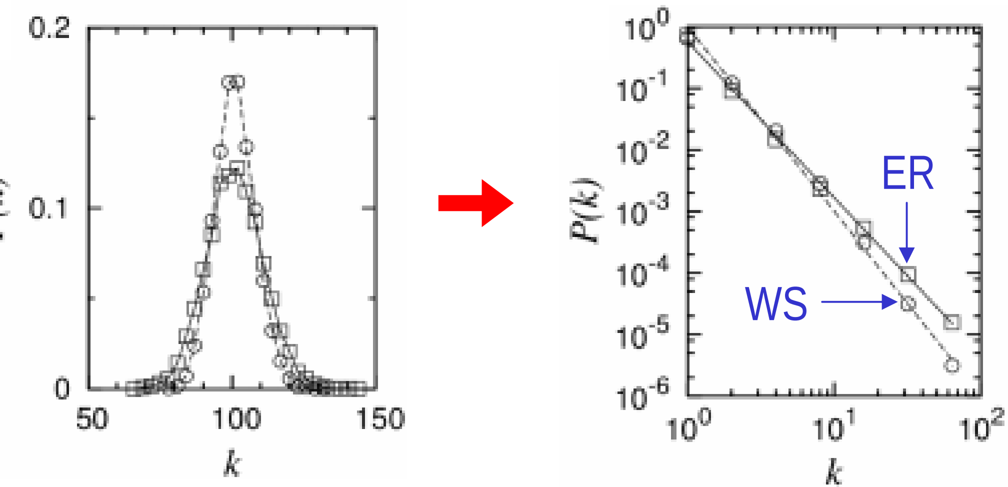
There is strong degree correlation between the spanning trees and their original networks.

The Spanning Trees of Homogeneous Networks (I)

[sparse version] $\bar{k} = 4,\ N = 10000$



[dense version] $\bar{k} = 100,\ N = 10000$



We examine two popular homogeneous networks, the Erdos-Renyi (ER) network, and Watts-Strogatz (WS) network (p=1).

In sparse version, P(k) of the spanning tree is not clear but similar to the exponential rather than scale-free.

The degree of the spanning tree is smaller than that of the original network.

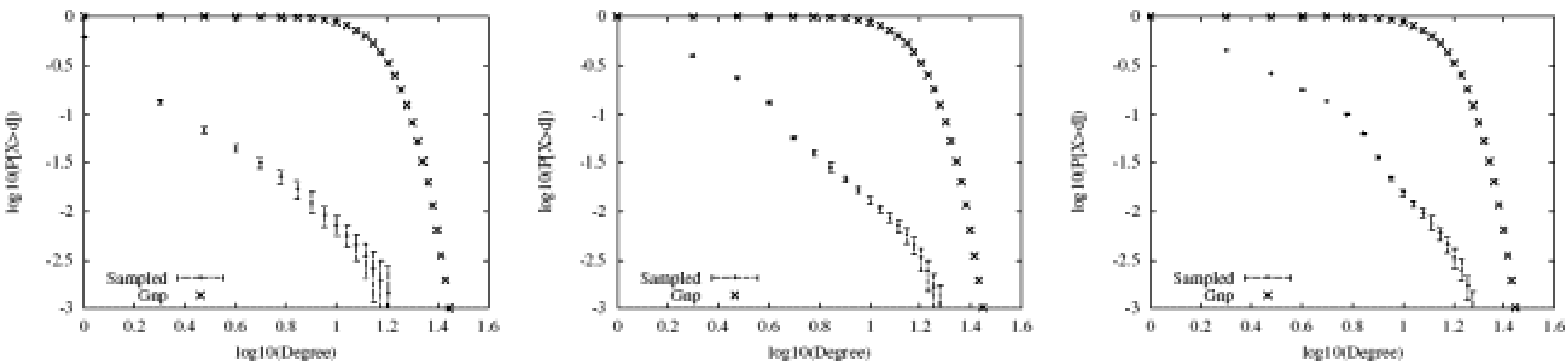
However, as it gets *denser*, the spanning tree becomes the *scale-free* network.

The Spanning Trees of Homogeneous Networks (II)

Lakhina, et al. presented that the emergence of the scale-free degree distribution of the internet topology is caused by the sampling bias of the real topology.

(A. Lakhina, J. Byers, M. Crovella, and P. Xie, "Sampling Biases in IP Topology Measurements," in *Proceedings of IEEE Infocom 2003*,San Francisco, California, April 2003)

- In the simulation, they obtained scale-free topology from sampling the shortest paths from few sources and many targets on the ER model.



(a) 1 source, 1000 destinations (b) 5 sources, 1000 destinations (c) 10 sources, 1000 destinations

Fig. 2. Degree Distribution of subgraph sampled from Erdős-Rényi random graph ($N = 100,000$, $p = 0.00015$)

The Spanning Trees of Homogeneous Networks (III)

But, though it is scale-free, the spanning tree of a homogeneous network is different from the scale-free spanning tree of a scale-free one.

- In case of the dense version, selecting high edge-BC links is not much different from the random selection.
  - It is not the communication kernel any more.

- Degree correlation between the spanning tree and the original network is much weaker than that for the scale-free networks.

$$r_p = \frac{\langle k_s k \rangle - \langle k_s \rangle \langle k \rangle}{\sqrt{\langle k_s^2 \rangle - \langle k_s \rangle^2} \sqrt{\langle k^2 \rangle - \langle k \rangle^2}}$$

0.97 for BA model

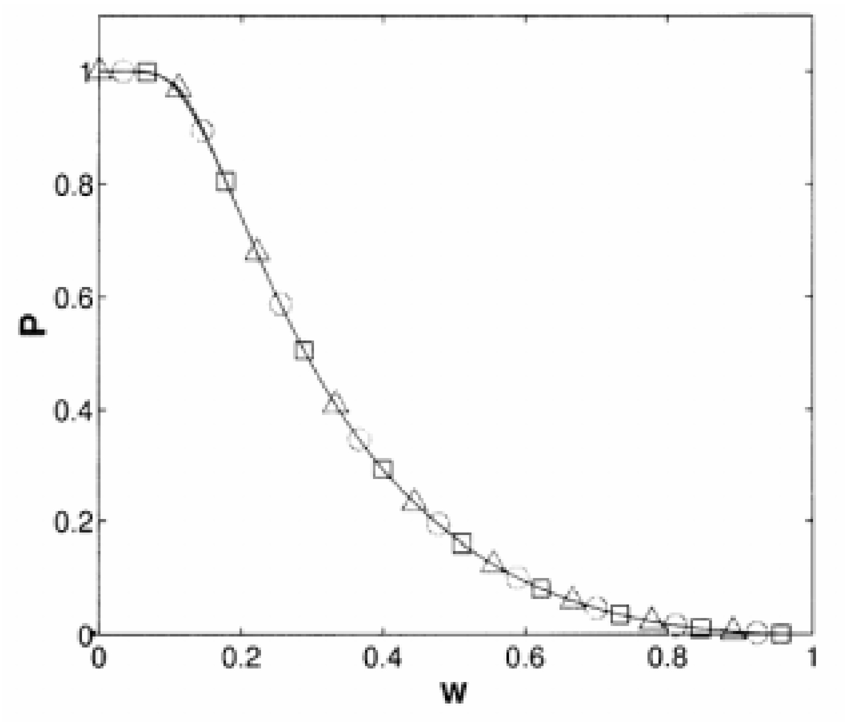
0.45 for ER model

- The BC exponent is not 2.0, one for the scale-free tree :

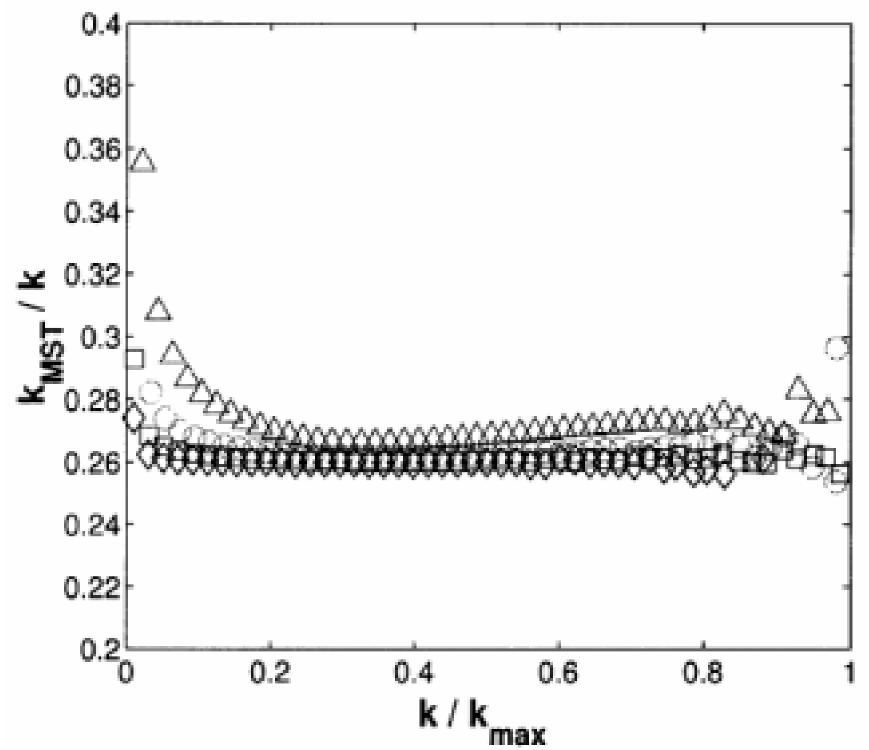
~1.7 for the ER and WS models.

Geometry of minimum spanning trees
on scale-free networks
Gabor J. Szaboa, Mikko Alavab, Janos Kertesz
Physica A (2003)

Assign random weight to the link & select minimum weight



Weight distribution on MST



Degree correlation

Isn’t the tree too simple??

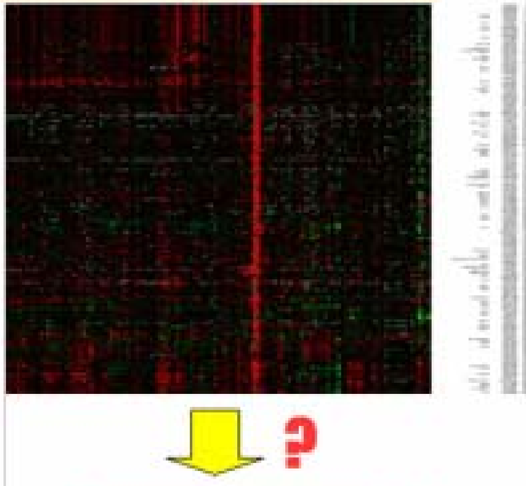
Tree can be a good starting point for unknown biological network.

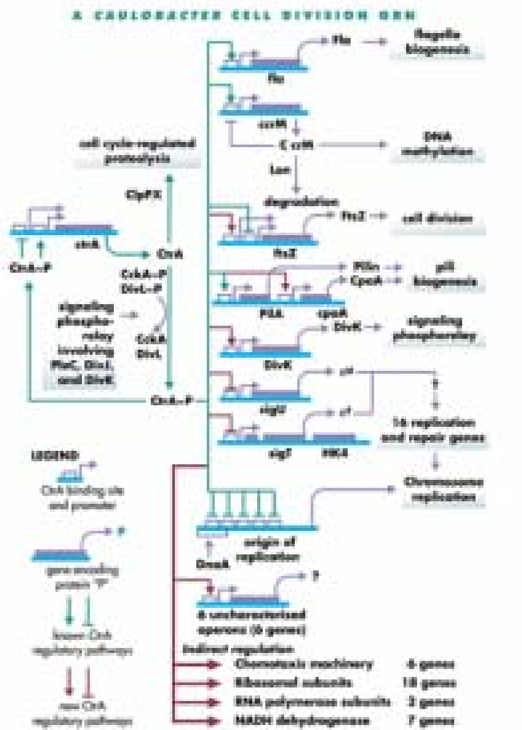
Q:What if we don’t know about the network at all? (in many biological network cases)

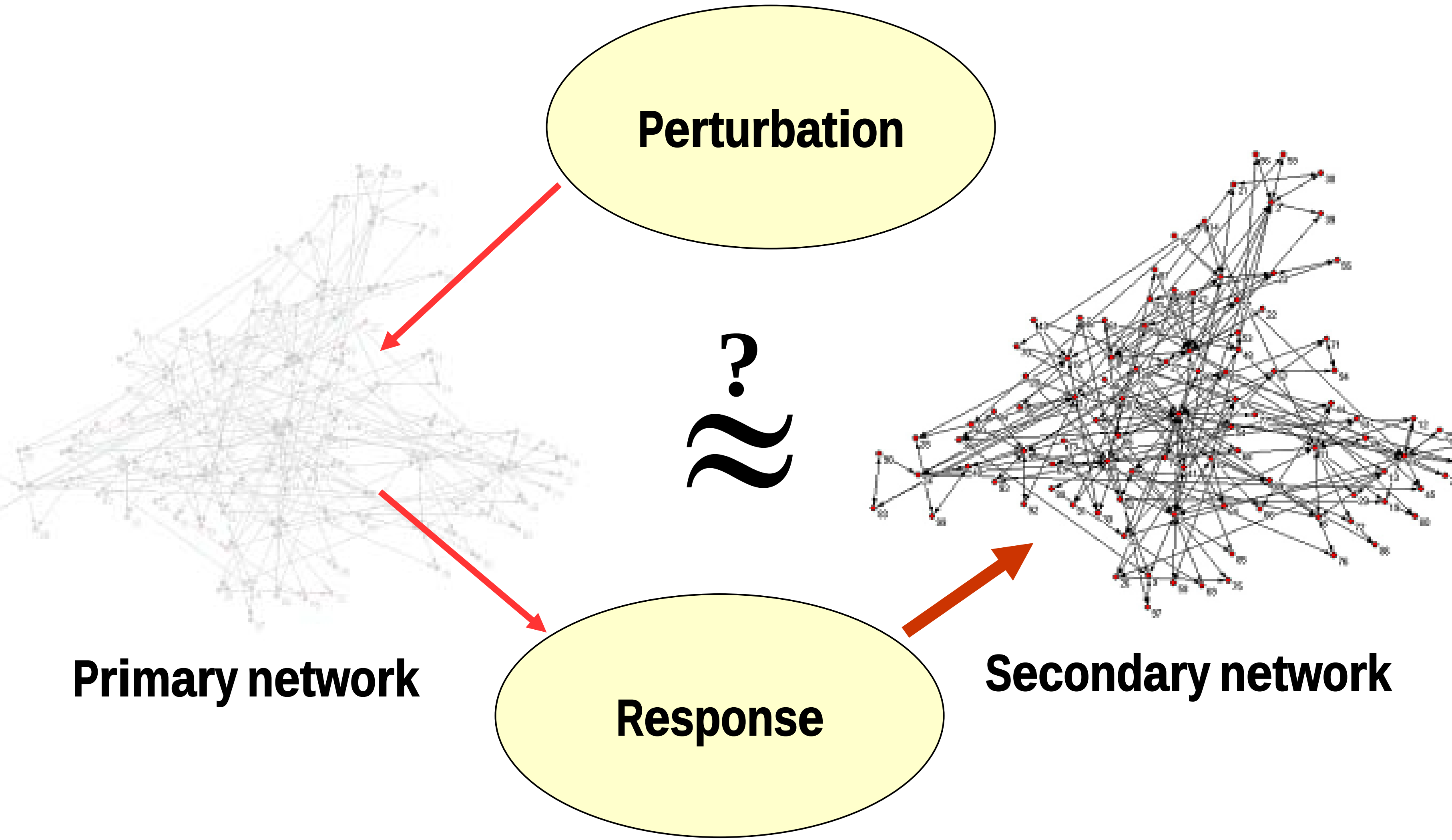
Motivation : Microarray Data

- Microarray data show the response of each gene to an experiment, which is a kind of perturbation to the genetic network.
  ex) gene deletion, temperature change etc.

- From microarray data, we try to understand the underlying genetic networks.






Perturbation
?
≈
Primary network
Response
Secondary network

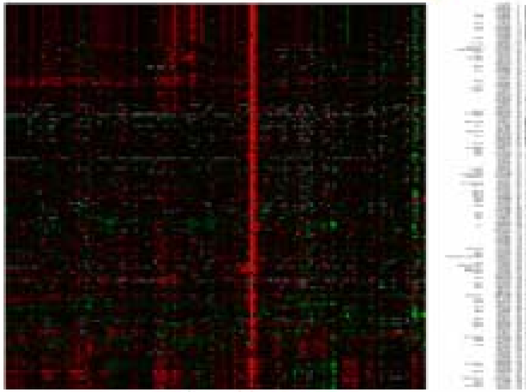
Motivation : Microarray Data

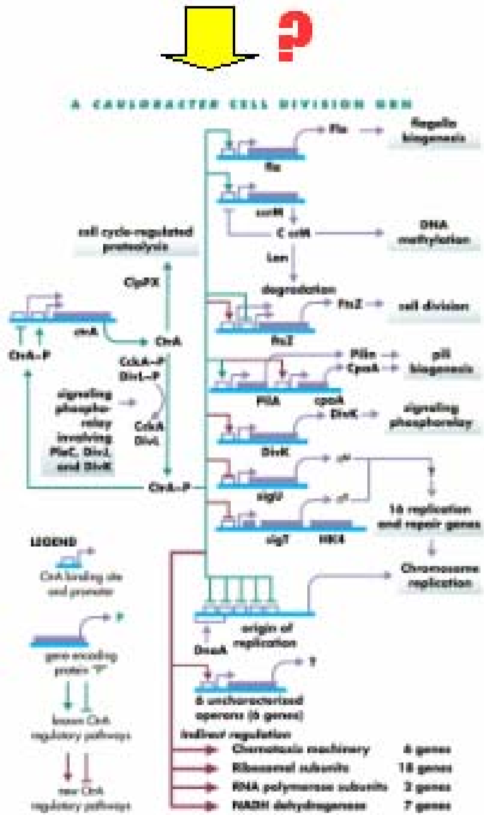
- Microarray data show the response of each gene to an experiment, which is a kind of perturbation to the genetic network.

  ex) gene deletion, temperature change etc.

- Like building the genetic network from microarray data, the secondary network can be constructed from the response of primary network under perturbation.

  ex) node removal (?)



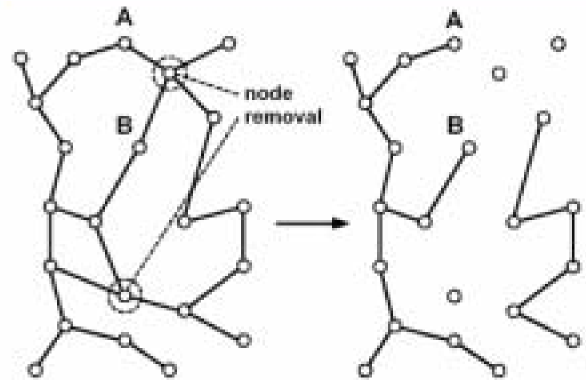


“ How can we construct secondary network from simple perturbations and responses ?”

“ Can the secondary network represents the primary network correctly ? ”

“ Ultimately, can we find out primary network from the secondary network ? ”

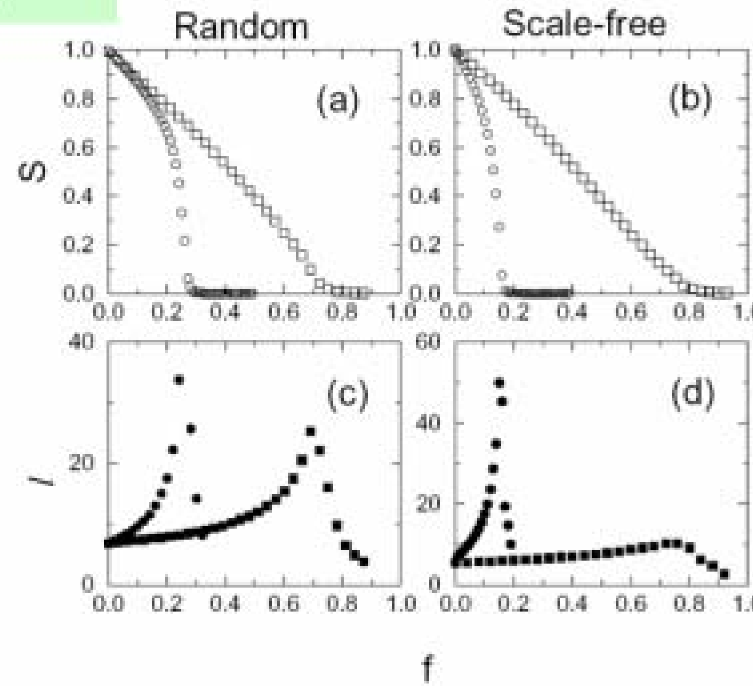
Introduction : Node Removal Perturbations



- When a node is removed, network structure changes. The network can break into several isolated clusters.
- Giant cluster size decreases gradually and the average path length increases.

R. Albert and A.-L.Barabási, *Reviews of Modern Physics*, **74**, 47 (2002)

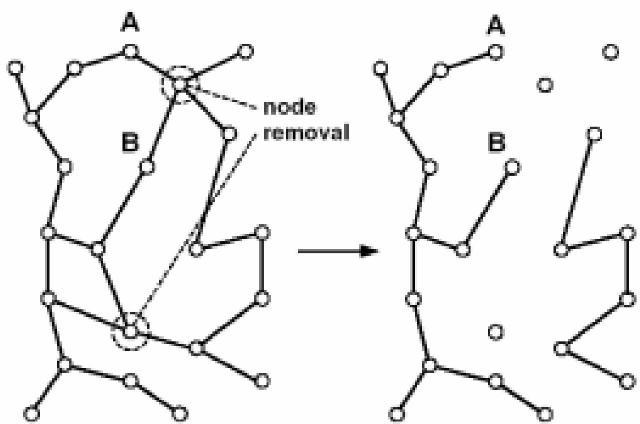
- SF network is more tolerant against random removal better than random network.
- But SF network is more vulnerable against designed attack than random network..



Node Removal & Betweenness Centrality Changes

Perturbation?

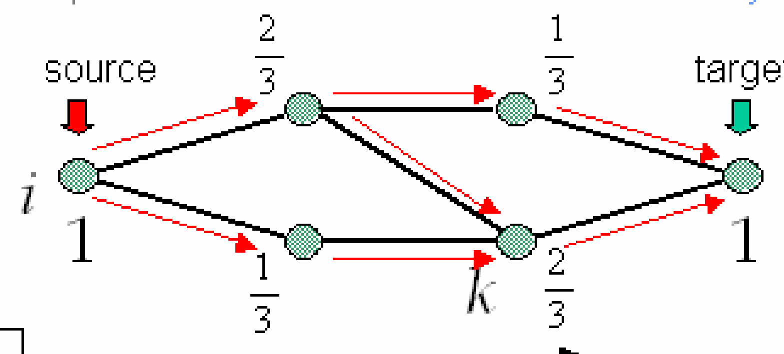
Node removal (e.g. gene knock-out in microarray data)



What will be the change of traffic through node j when node i is removed?

Response?

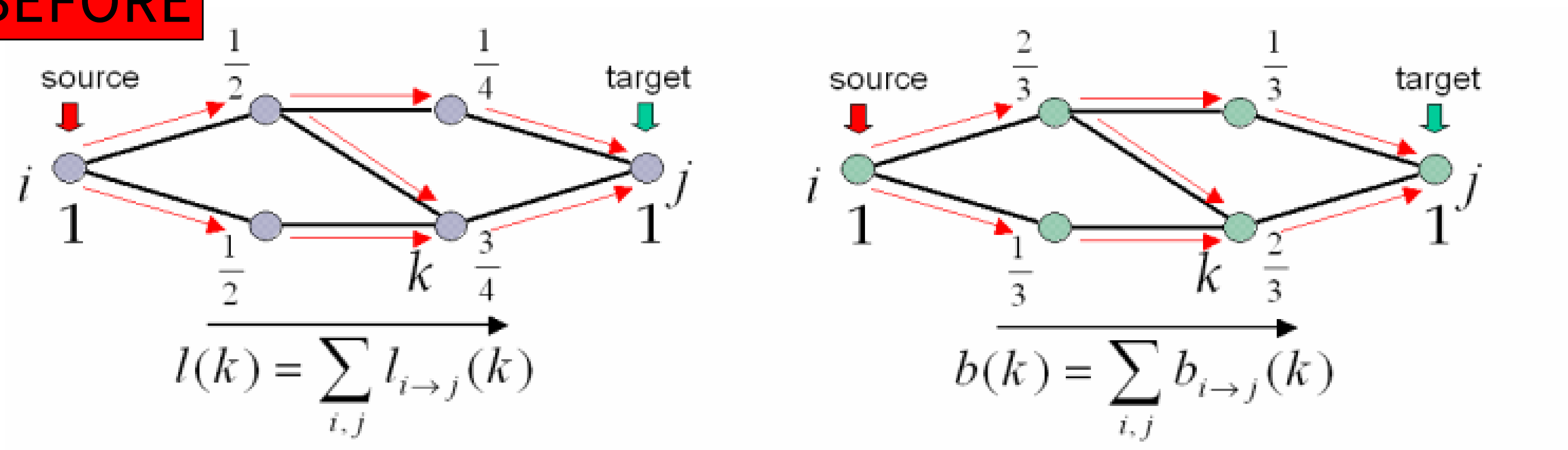
- **Betweenness Centrality BC** [~average traffic over a node](Freeman, 1977
  - if $g_{i\to j}$ is the number of geodesic paths from i to j and $g^k_{i\to j}$ is the number of paths from i to j that pass through k, then $g^k_{i\to j}/g_{i\to j}$ is the proportion of geodesic path from i to j that pass through k. The sum $b_{i\to j}(k)\equiv g^k_{i\to j}/g_{i\to j}$ for all i, j pairs is betweenness centrality.



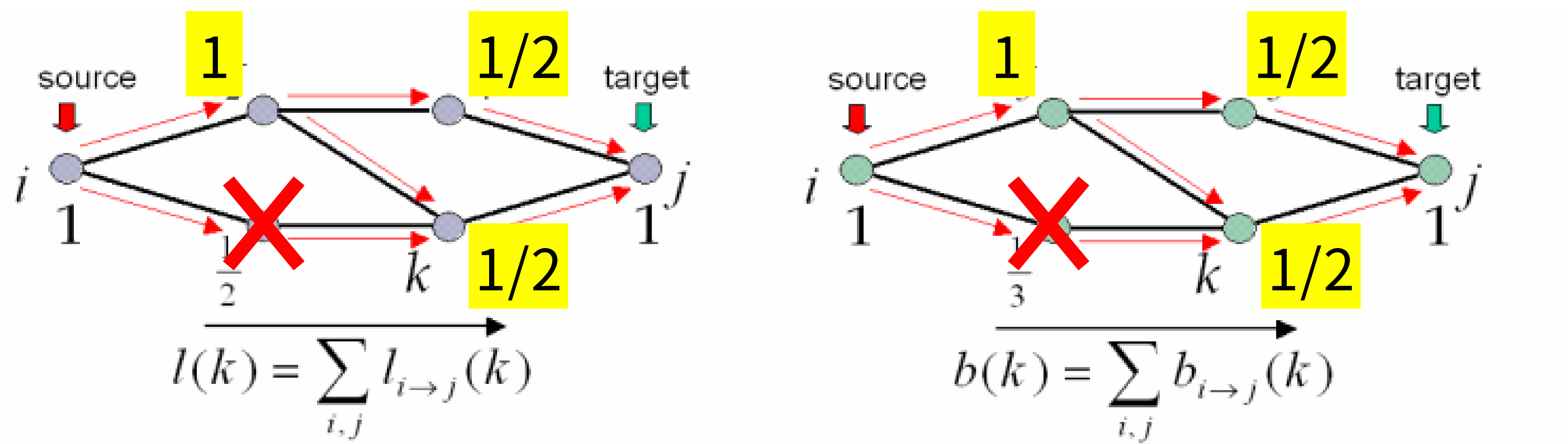
$$b(k) = \sum_{i,j} b_{i\to j}(k)$$

Effect of node removal

BEFORE



AFTER



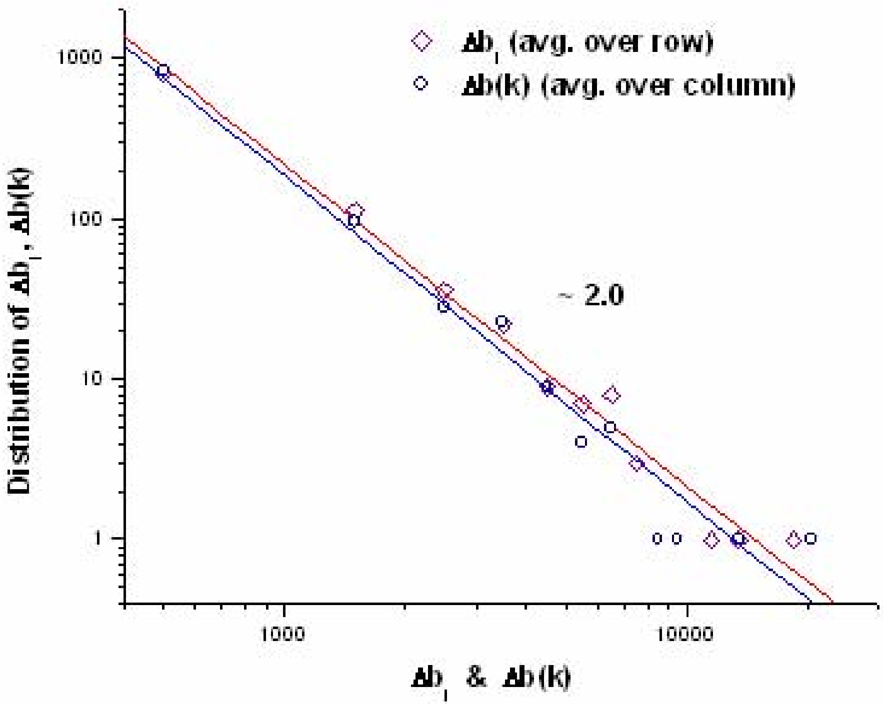
BC Changes - BA model

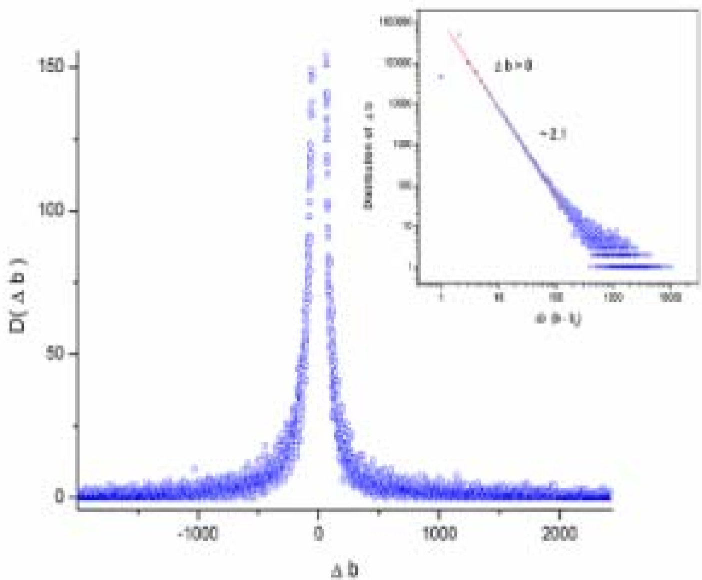
$$\Delta b_i(k) = b_i(k) - b_o(k)$$

$b_i(k) = \text{BC of k-th node after i-th node removal}$

$b_o = \text{BC of k-th node after the whole network}$

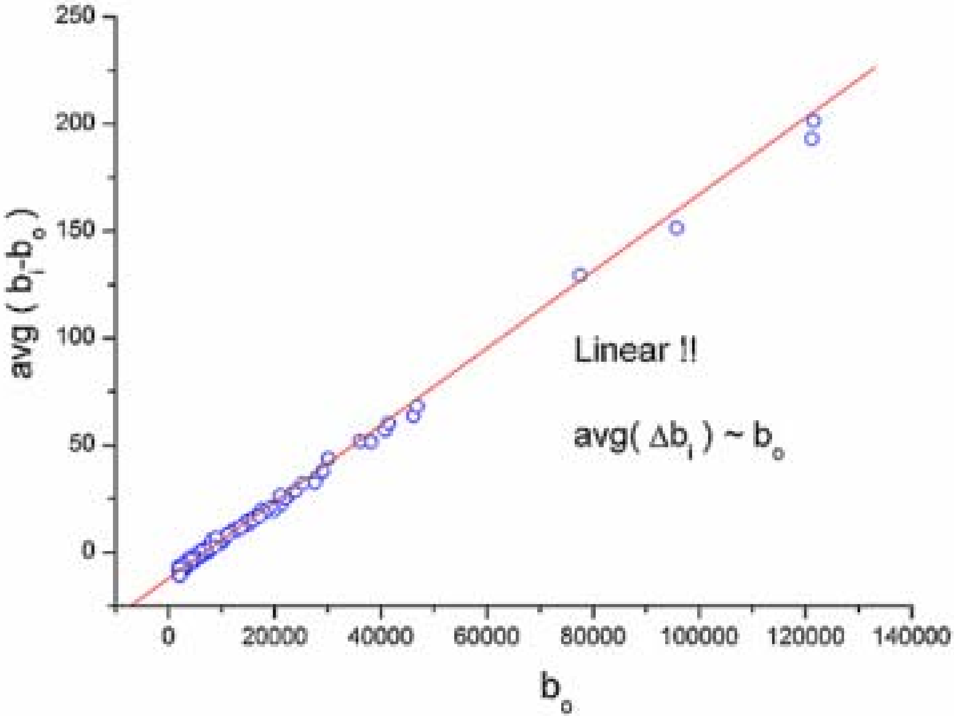
$$\Delta b = \begin{bmatrix} \Delta b_1(1) & \cdots & \Delta b_1(j) & \cdots \\ \vdots & \ddots & \vdots & \\ \Delta b_i(1) & \cdots & \Delta b_i(j) & \\ \vdots & & & \ddots \end{bmatrix}$$



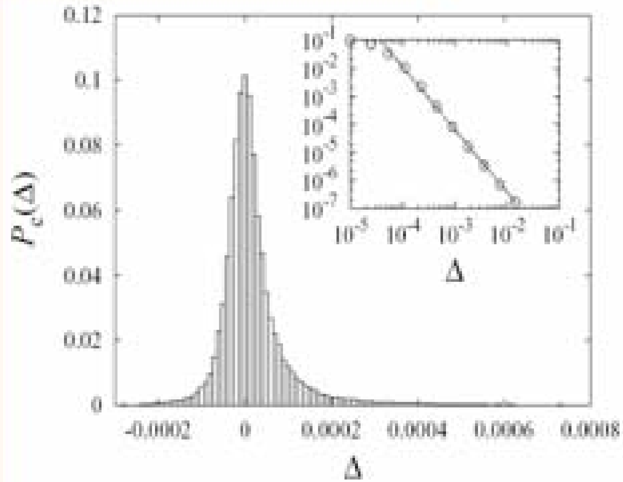


- $\Delta b_i(k)$ distribution is power law distribution with exponent 2.1

Distribution of $\Delta b_i(k)$ - BA model



cf) diameter changes



J.-H. Kim, *et al.*, *Physical Review Letters*, **91**, 5 (2003)

$$\Delta b_i = \sum_k \Delta b_i(k) = \sum_k b_i(k) - \sum_k b_o(k)$$

$$= b_i - b_o + b_o(i)$$

$$\because \quad \Delta_i \cong \frac{\Delta b_i - b(i)}{b_o}$$

- $\Delta b_i \propto b_o(i)$

The average of BC changes after i-th node removal is linearly proportional to BC of i-th node in BA model.

MST & Percolation Network

- How to build the secondary network ?

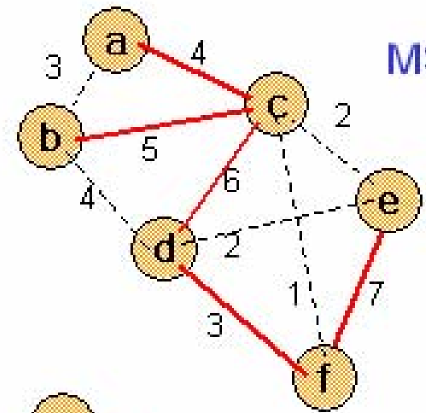
: Based on $\{\Delta b_i(j)\}$ = "correlation" between node i and j

$$\Delta b = \begin{bmatrix} \cdot & \cdots & \Delta b_1(j) & \cdots \\ \vdots & \ddots & \vdots & \\ \Delta b_i(1) & \cdots & \Delta b_i(j) & \cdots \\ \vdots & & \vdots & \ddots \end{bmatrix}$$

  - **MST** (minimum spanning tree)

    A graph G = (V,E) with weighted edges. The subset of E of G of minimum weight which forms a tree on V ≡ MST .

    A node is linked to the most influential one with constraint such that N vertices must be connected only with (N-1) edges.

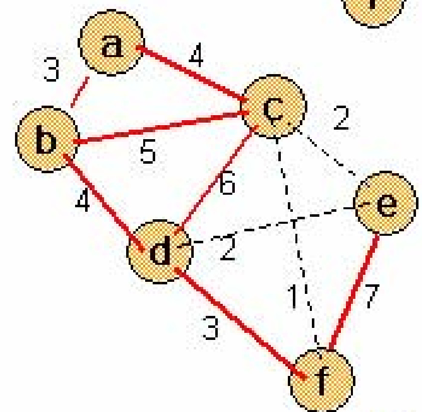


MST

  - **Percolation**

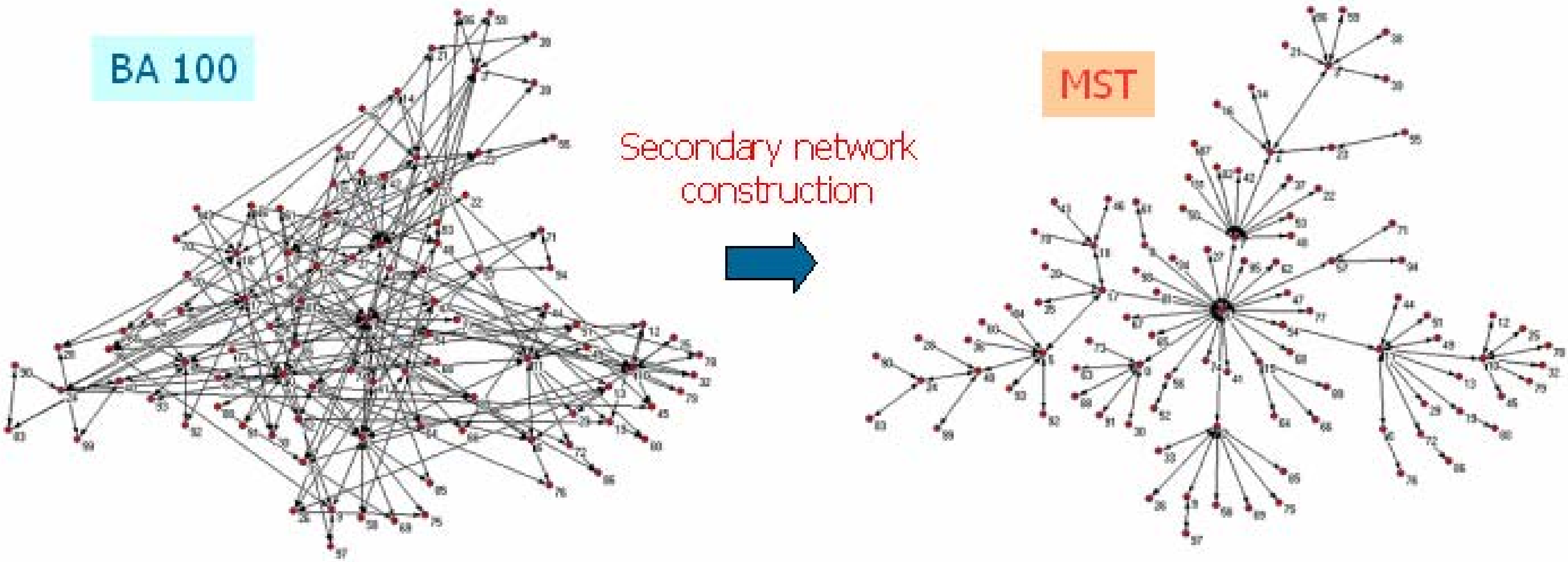
    After sorting $\Delta b_i(j)$ in descending order, add a link between i and j following that order. When all nodes make a giant cluster, stop the attachment.

    It means the links with values $\Delta b_i(j) > b^*$ (percolation threshold) are valid and connected.



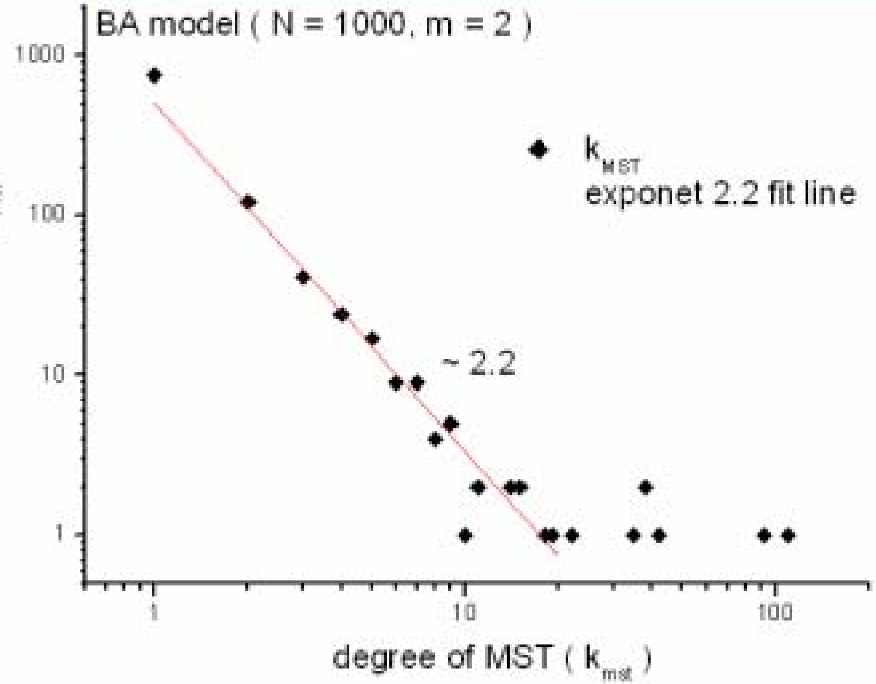
Percolatio

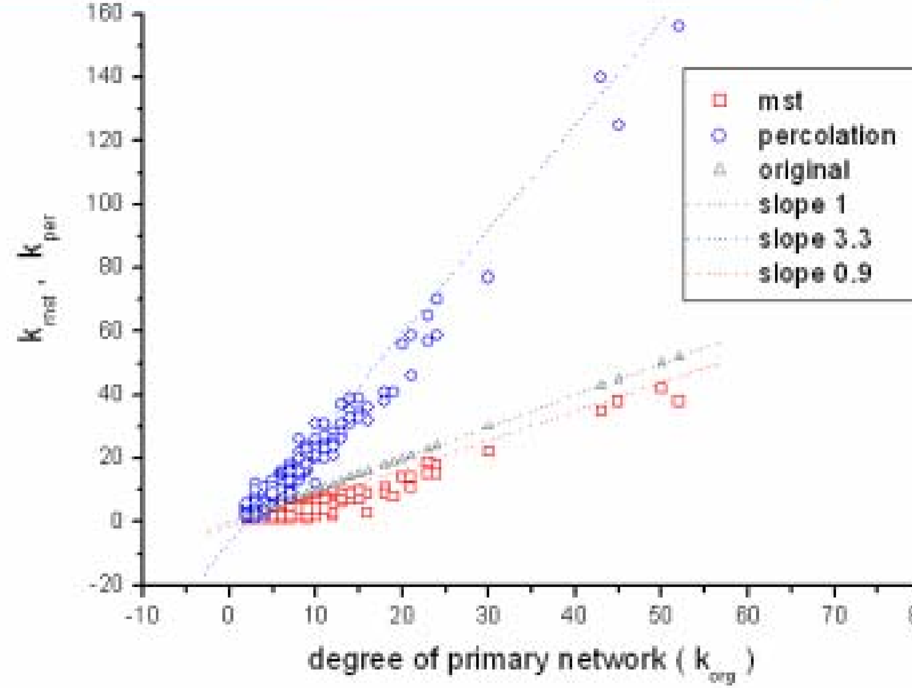
Result : Secondary Networks



- The degree k of secondary networks contain the global information of primary network, because it is constructed from BC that is calculated from the information of whole network.
- More sparse or dense networks which contain the information of original network can be constructed.
- Secondary networks represent the primary network well with significant link matches.

Result : Minimal Spanning Tree





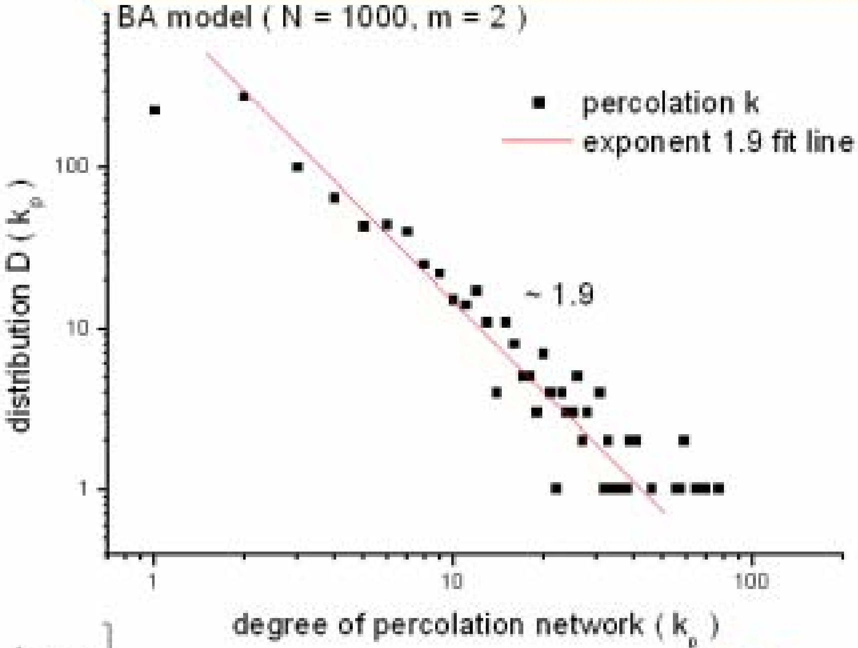
- In MST network, the degree distribution shows the power-law with exponent 2.2 not 3.0 (Scale-free)

$$D(k_{mst}) \sim k_{mst}^{-2.2}$$

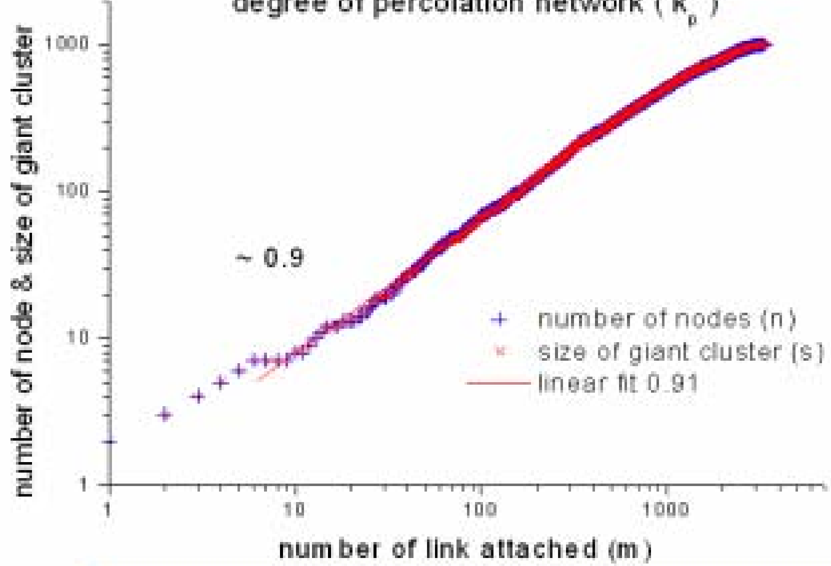
- The degree of each node in secondary network is linearly correlated to that of primary network.

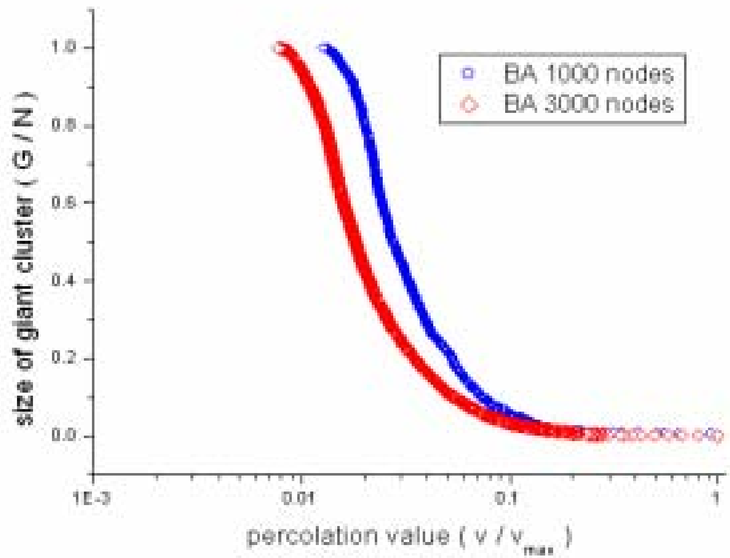
$$k_{mst}, k_{per} \propto k_{org}$$

Result : Percolation Network

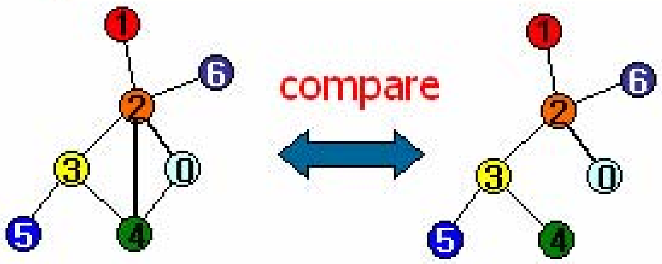


- The degree distribution of percolation network shows power-law. (exponent -1.9)
- Percolation features appear during giant cluster formation.





Similarity Measurement between Two Networks



$$v_i = (w_{i1}, w_{i2}, \cdots, w_{ij}, \cdots, w_{iN}) \quad x_i = v_i \cdot v_i^{'}$$

$$w_{ij} = \begin{cases} 1 & (\text{if i, j are linked}) \\ 0 & (\text{otherwise}) \end{cases}$$

$$\text{similarity measure } X = \frac{1}{M} \sum_{i=1}^{N} x_i$$

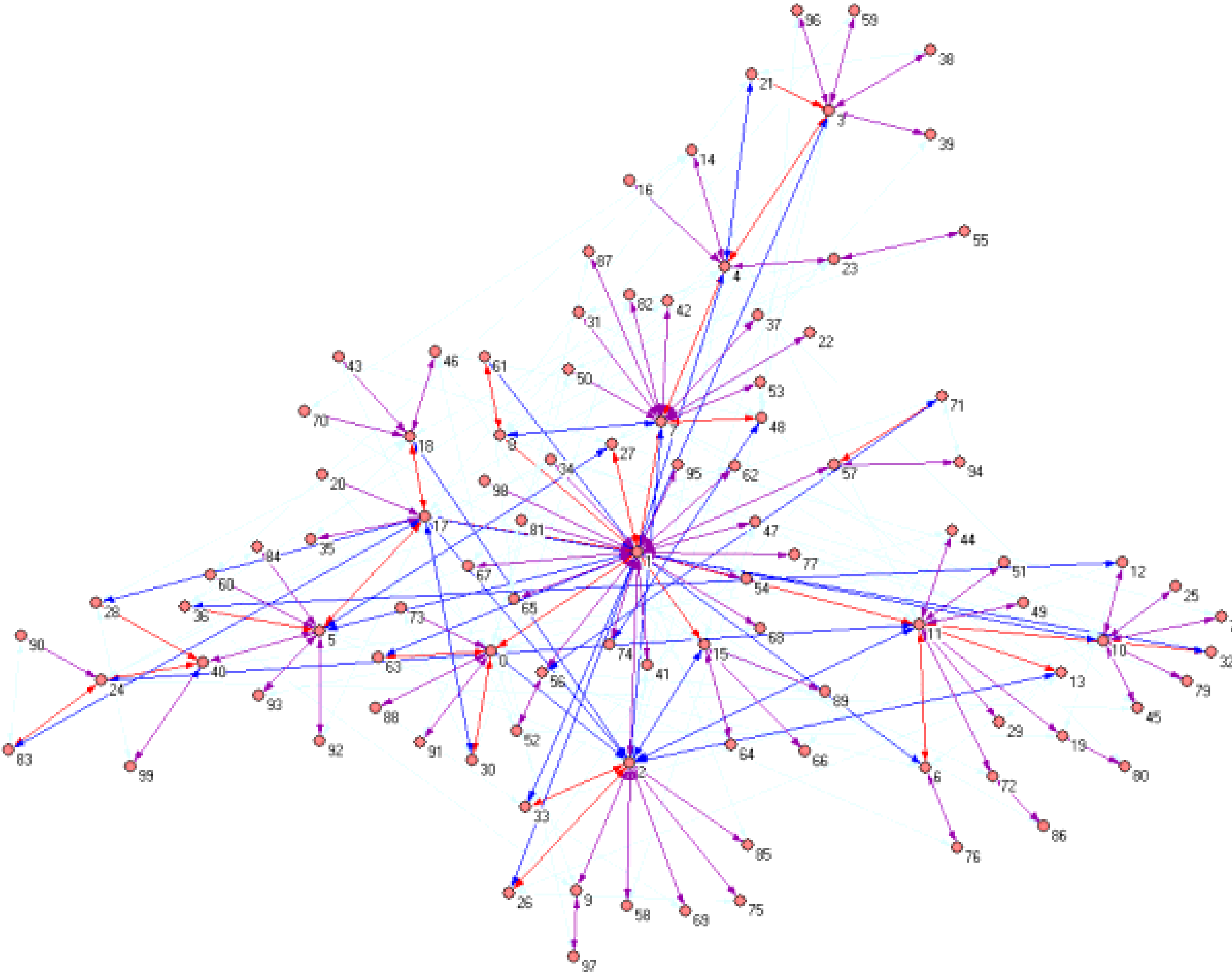
- The links of each node are regarded as vector in N dimensional vector space.
  - Vector inner product shows the similarity between two networks.
- Binary undirected network case : It means how many links are overlapping each other.

- The network similarity measure between secondary and primary networks are significantly higher than other network.

(MST : 90.8 % , percolation : 76.6 %)

The secondary networks well represent the primary network.

BA model (N = 1000 , M = 1996)			
	links	X	matches
MST network	999	0.908	907
Percolation net.	3377	0.766	1529
Other BA net.	1996	0.019	39
RG network	1996	0.012	23
Random net.	2041	0.003	5
	957	0.001	1


96
59
38
21
3
39
14
16
55
87
23
4
82
42
31
37
22
46
61
43
50
53
71
70
7
48
18
8
27
34
95
62
57
94
20
98
17
81
47
44
35
84
1
77
67
51
12
60
54
25
28
36
65
49
73
78
11
5
90
68
74
15
10
32
63
0
41
40
24
56
13
93
89
79
88
45
29
52
19
91
92
83
99
30
64
66
2
6
72
80
33
86
85
76
26
9
75
58
69
97

Conclusions I

- **We investigate the degree distribution of the spanning trees of complex networks.**

- **The spanning trees are obtained in 3 different ways : minimum spanning tree technique, percolation approach, and burning from a randomly selected vertex.**

- **The spanning trees show *scale-free* behavior in their degree distributions, regardless of methods to obtain the spanning trees.**

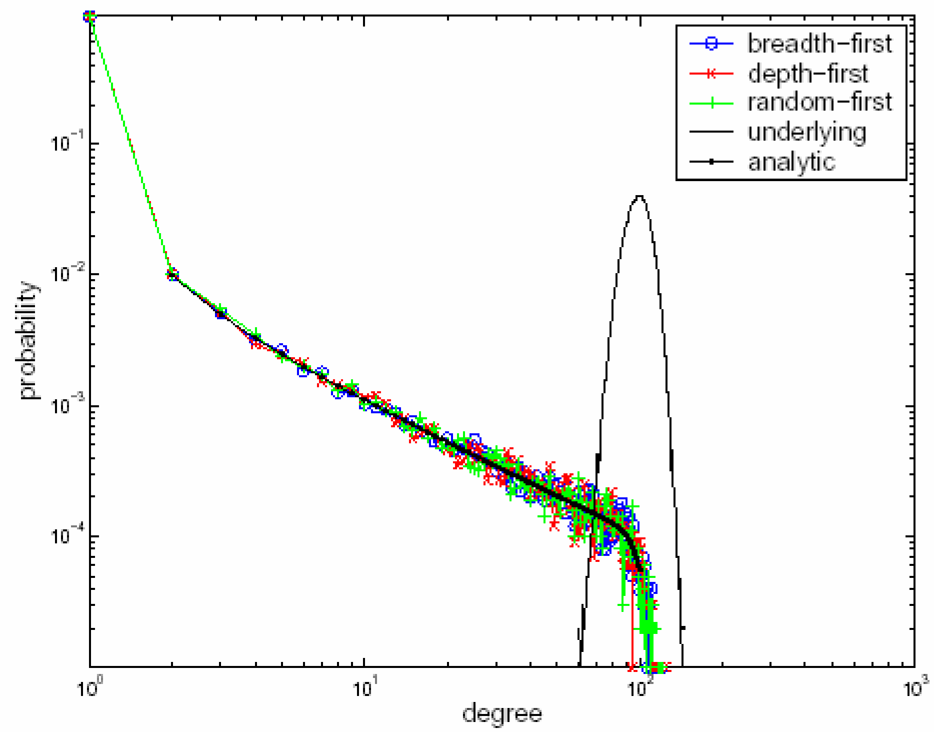
SF Network = SF Tree (Communication Kernel) + Shortcuts

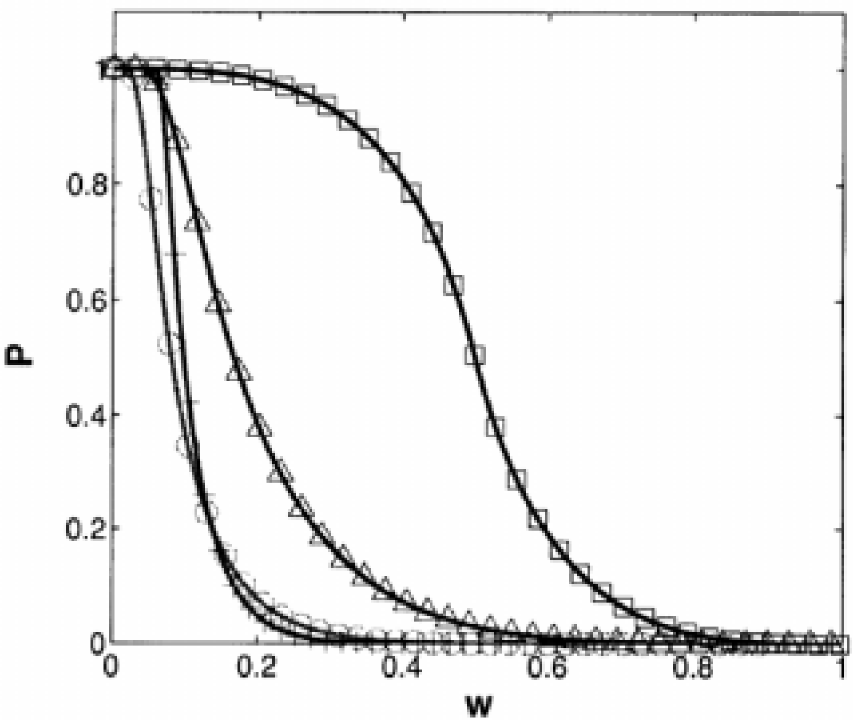
Conclusions II

- From the response (changes in BC) of the network under simple perturbation (node deletion), we construct a secondary network.
- Two secondary networks, MST & percolation network, reproduce the scale-free behavior and its degree of each node is in proportion to degree of primary network. Its degree contains the global information of primary network.
- Similarity measurement shows that the secondary networks reproduce original network quite well. (MST: 91% , percolation: 77%)
- From biological perturbation data (DNA chip), we developed a method to construct unknown primary network (genetic network).

Secondary Network can be found from simple perturbations.

Communication Kernel ≈ Secondary Network


breadth-first
depth-first
random-first
underlying
analytic
probability
degree
10^0
10^1
10^2
10^3
10^-1
10^-2
10^-3
10^-4



P
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
w