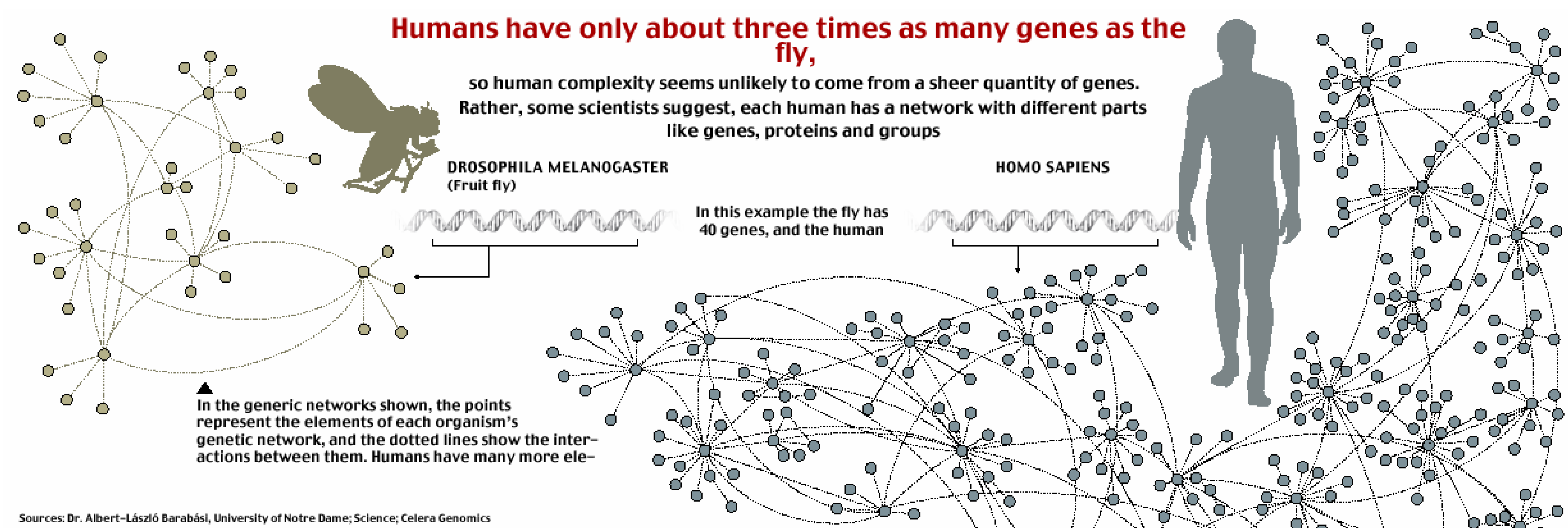
The architecture of complexity: From the topology of the www to the cell's genetic network

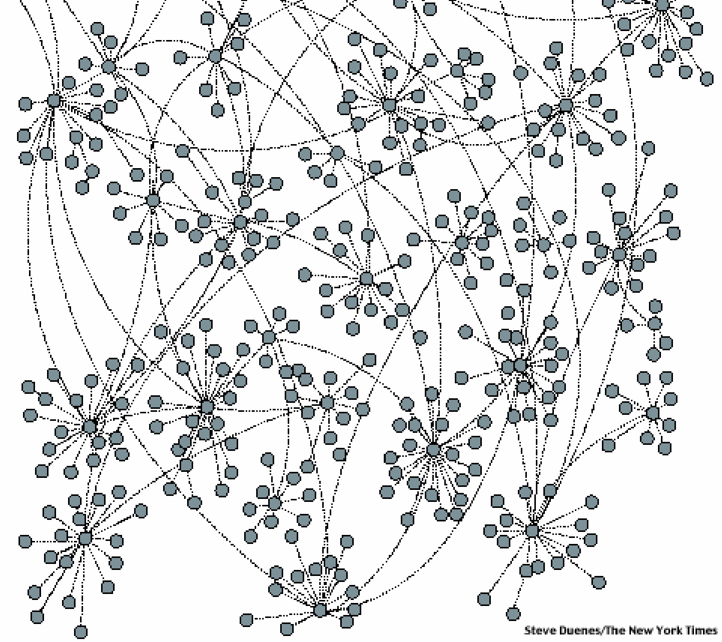
Albert-László Barabási
University of Notre Dame

Zoltán N. Oltvai
Northwestern Univ., Medical School

H. Jeong, R. Albert, E. Ravasz, G. Bianconi, E. Almaas

www.nd.edu/~networks





Complex systems

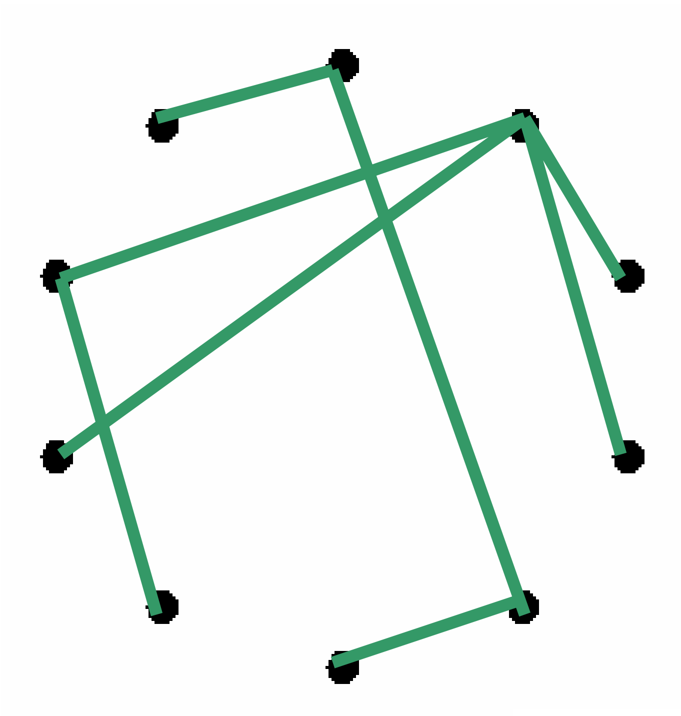
Made of

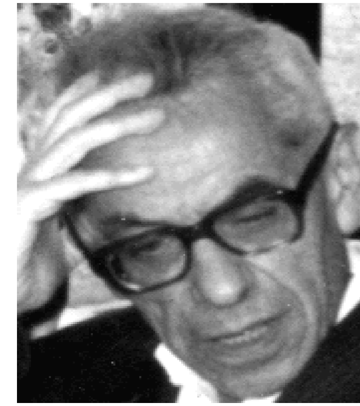
many non-identical **elements**

connected by diverse **interactions**.

NETWORK

Erdös-Rényi model
(1960)



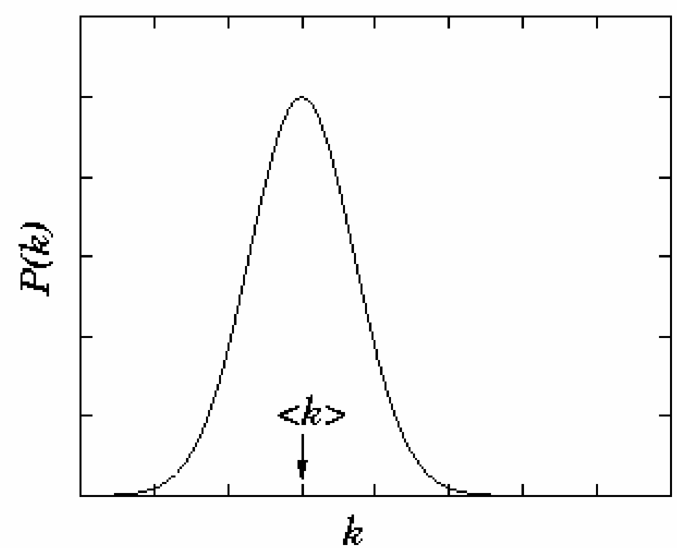


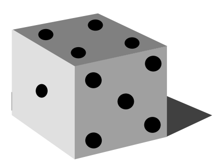
Connect with probability p

p=1/6
N=10
$\langle k \rangle \sim 1.5$

Pál Erdös
(1913-1996)

Poisson distribution

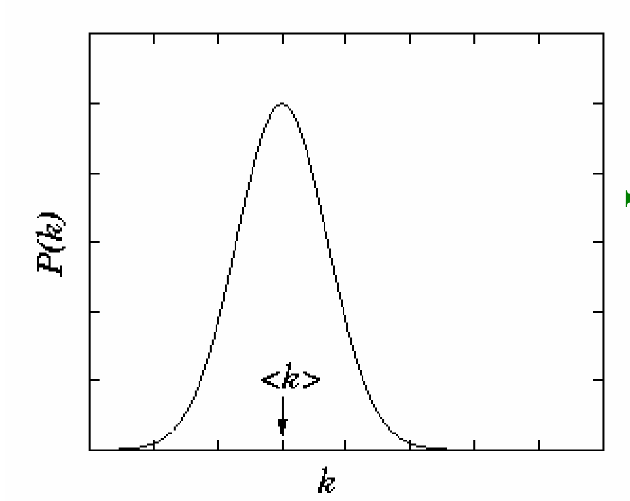




- **Democratic**
- **Random**

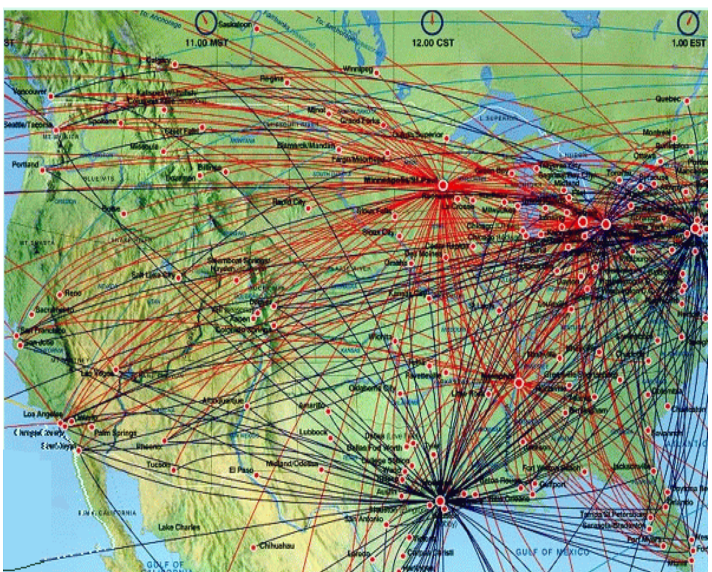
World Wide Web

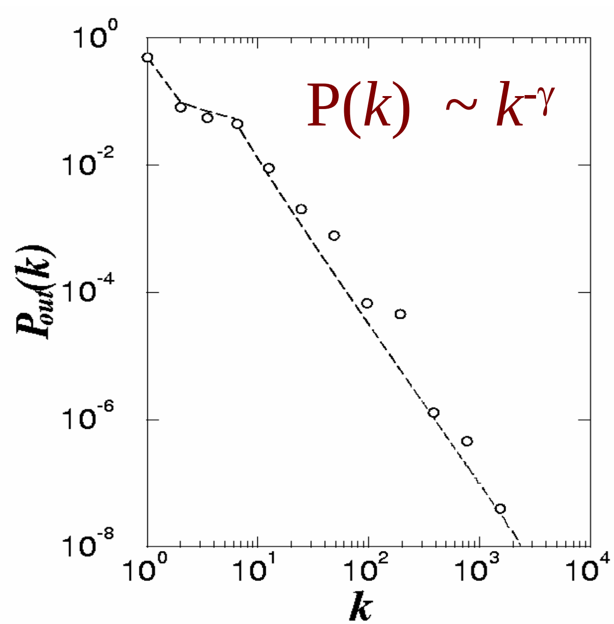
Exponential Network



Expected

Scale-free Network



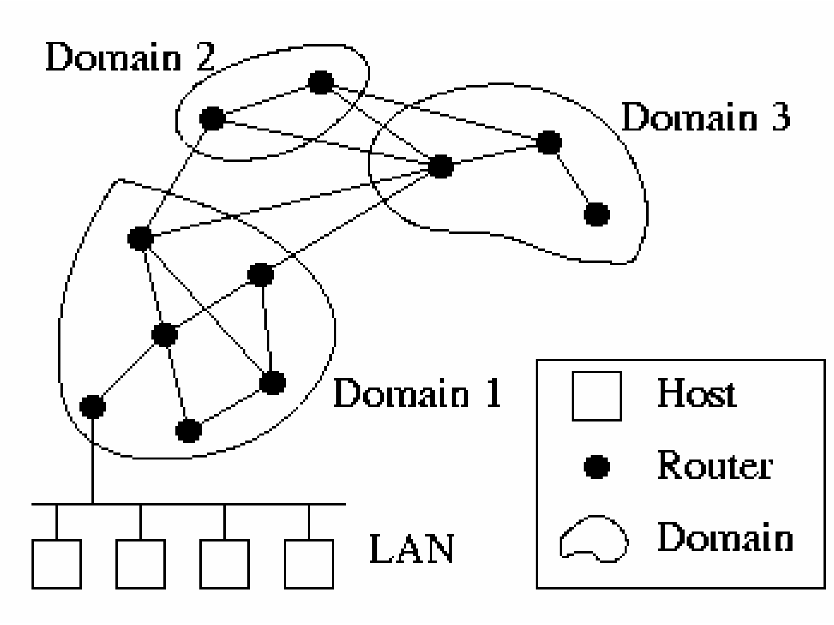


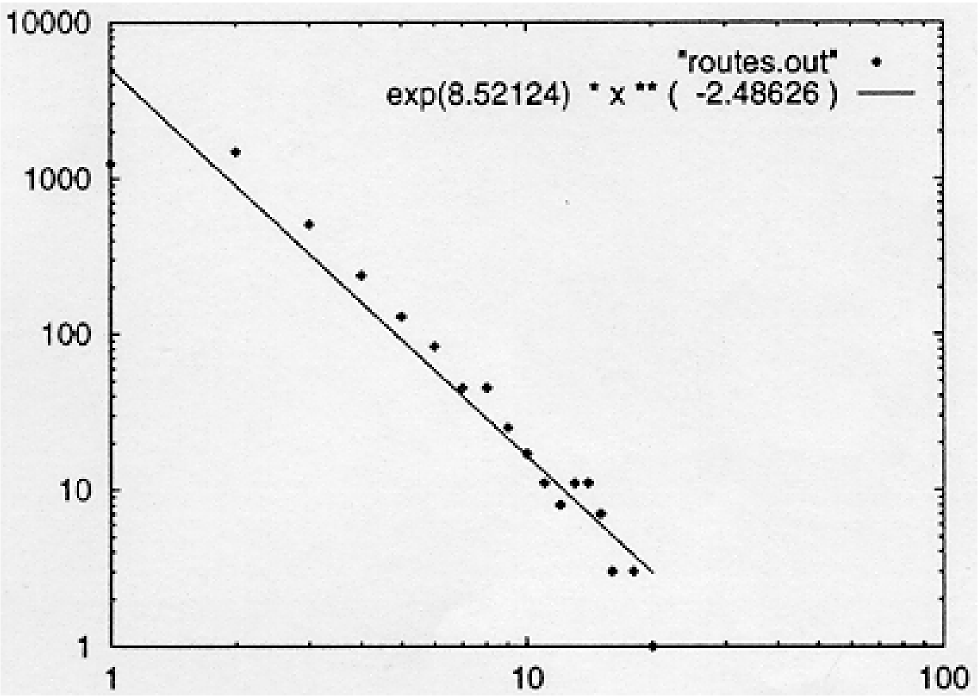
Found

R. Albert, H. Jeong, A-L Barabasi, *Nature*, **401** 130 (1999).

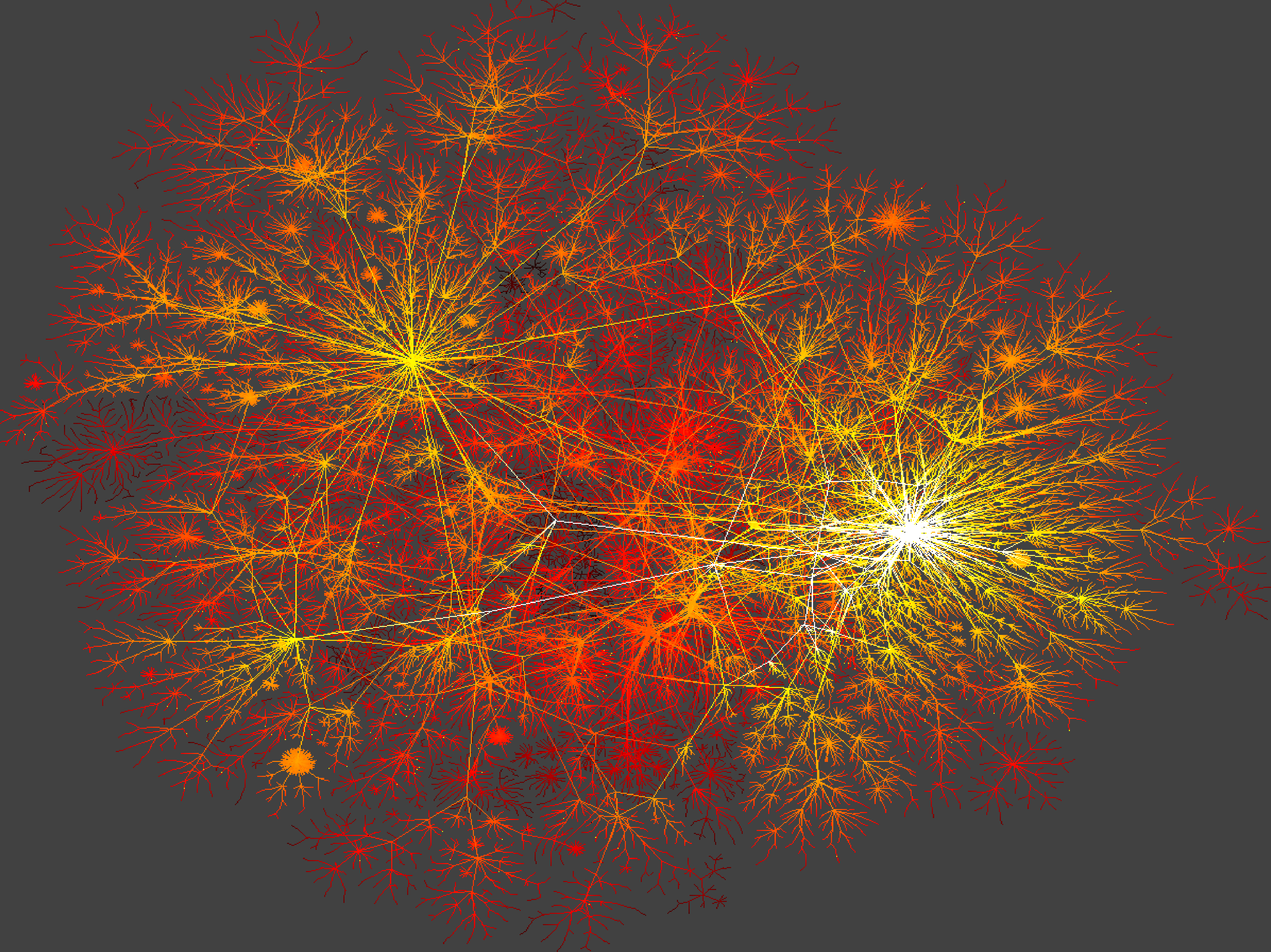
INTERNET BACKBONE

Nodes: computers, routers
Links: physical lines



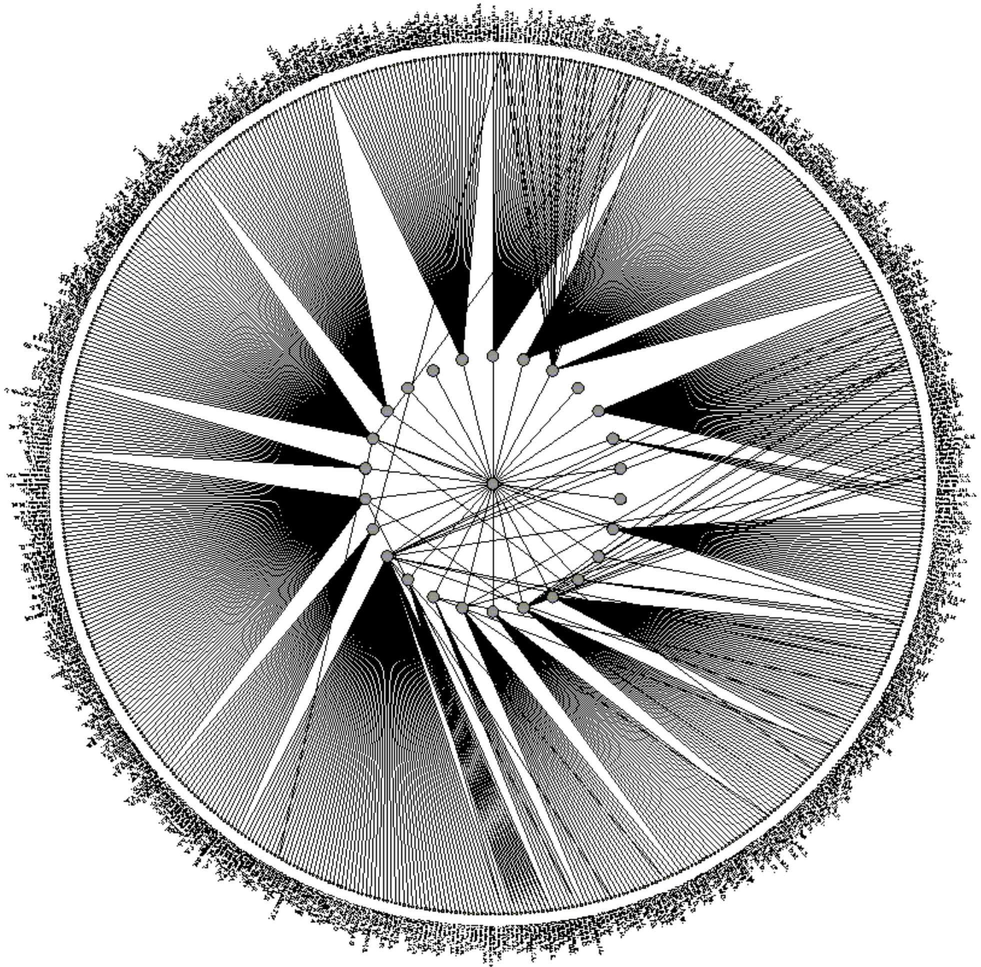


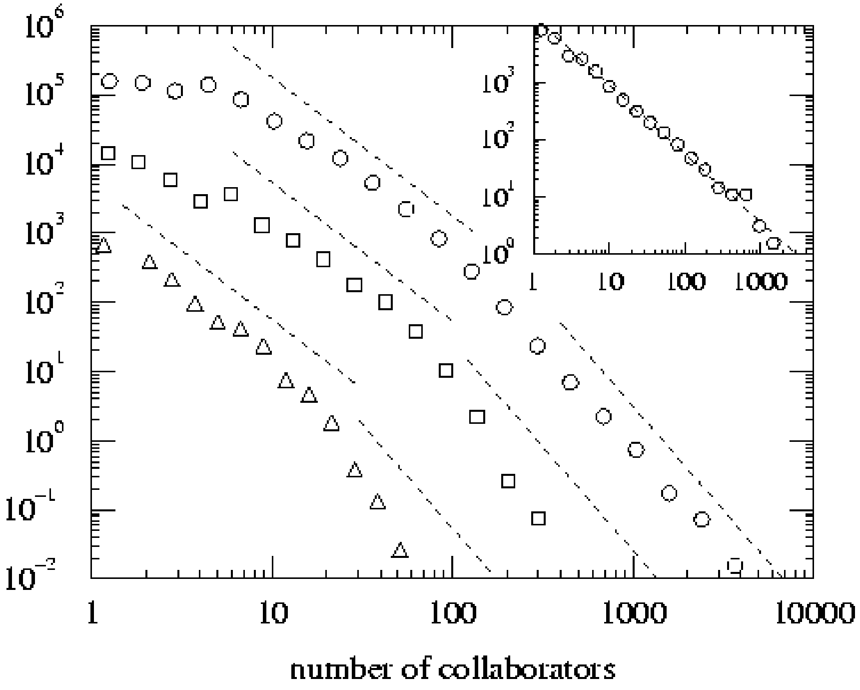
(Faloutsos, Faloutsos and Faloutsos, 1999)



SCIENCE COAUTHORSHIP

Nodes: scientist (authors)
Links: write paper together





(Newman, 2000, A.-L. B. *et al* 2001)

SCIENCE CITATION INDEX

1,000 Most Cited Physicists, 1981-June 1997

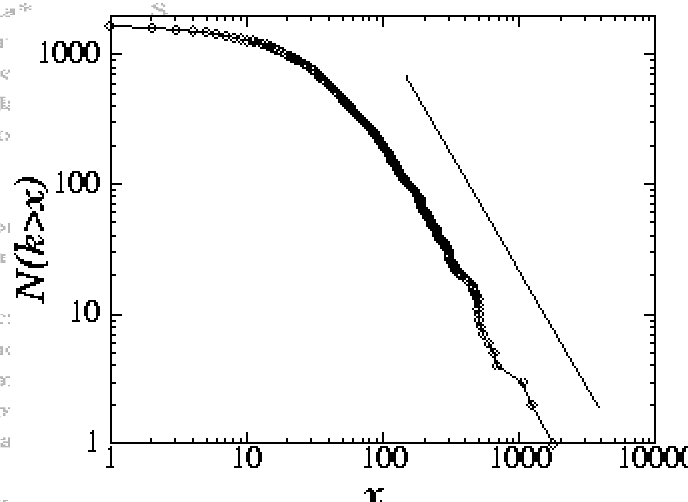
Out of over 500,000 Examined

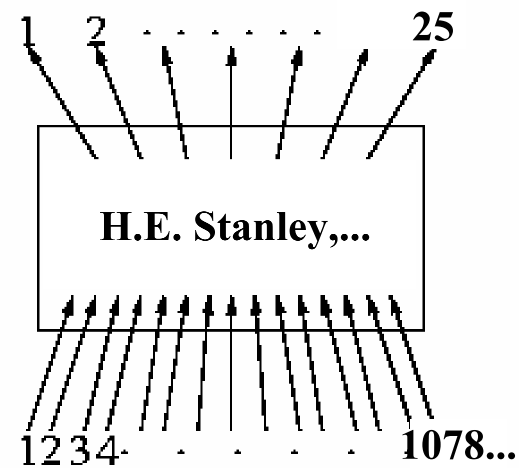
(see http://www.ssl.nrel.gov)

Nodes: papers

Links: citations

1736 PRL papers (1988)





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

(γ = 3)

(S. Redner, 1998)

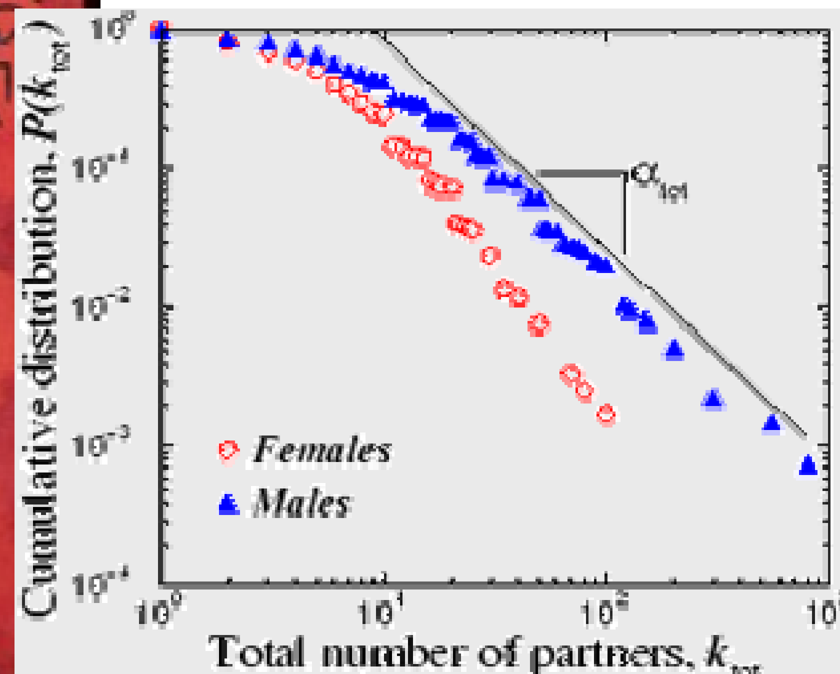
Author name		Institute	Country	Field	avg. cites	total art.	total cites	rank by total cit.
Witten	E	Princeton (U)	USA, NJ	High-energy (T)	168	138	23235	1
Gossard	AC	UCSB (U)	USA, CA	Semi				2
Cava	RJ	Bell Labs (I)	USA, NJ	Supe				3
Batlogg	B		USA, NJ	Supe				4
Ploog	K		Germany	Semi				5
Ellis	J	Euro Nuclear Cent.	Switzerland	Astr				6
Fisk	Z			Solid				7
Cardona	M	Max Planck (NL)	Germany	Semi				8
Nanopoulos	DV	Texas A&M (U)	USA, TX	High				9
Heeger	AJ	UCSB (U)	USA, CA	Poly				10
Lee*	PA							11
Suzuki*	T							12
Anderson			NJ	Solid				13
Suzuki*								14
Freeman				Solid				15
Tanaka*	S							16
Muller			nd	Supe				17
Schnee				Supe				18
Chemk				Optics (E)	60	162	9668	19
Morko				Semiconductors (E)	20	477	9668	19
Miller				Semiconductors (E)	67	144	9652	21
Chu				Supercond			9453	22
Bedno			nd	Supercond		85	9311	23
Cohen				Solid State (T)	33	284	9311	23
Meng				Superconductivity (E)	86	108	9300	25
Waszc				Superconductivity (E)		162	9170	26
Shiran				Supercond		269	8841	27
Wiegm				Semiconductors (E)	85	104	8822	28
Vando				Magnetism (E)	67	129	8686	29
Uchida					28	301	8520	30
Hor				Superconductivity (E)	72	119	8512	31
Murphy				Astronomy (E)	111	76	8439	32
Birgeneau	RJ	MIT (U)	USA, MA	Superconductivity (E)	41	286	8375	33
Jorgensen	JD	Argonne (NL)	USA, IL	Superconductivity (E)				
Hinks	DG	Argonne (NL)	USA, IL	Superconductivity (E)				

* citation total may be skewed because of multiple authors with the same name

Swedish sex-web

Nodes: people (Females; Males)

Links: sexual relationships



4781 Swedes; 18-74;
59% response rate.

Liljeros et al. *Nature* 2001

Many real world networks have a similar architecture:

Scale-free networks

WWW, Internet (routers and domains), electronic circuits, computer software, movie actors, coauthorship networks, sexual web, instant messaging, email web, citations, phone calls, metabolic, protein interaction, protein domains, brain function web, linguistic networks, comic book characters, international trade, bank system, encryption trust net, energy landscapes, earthquakes, astrophysical network…

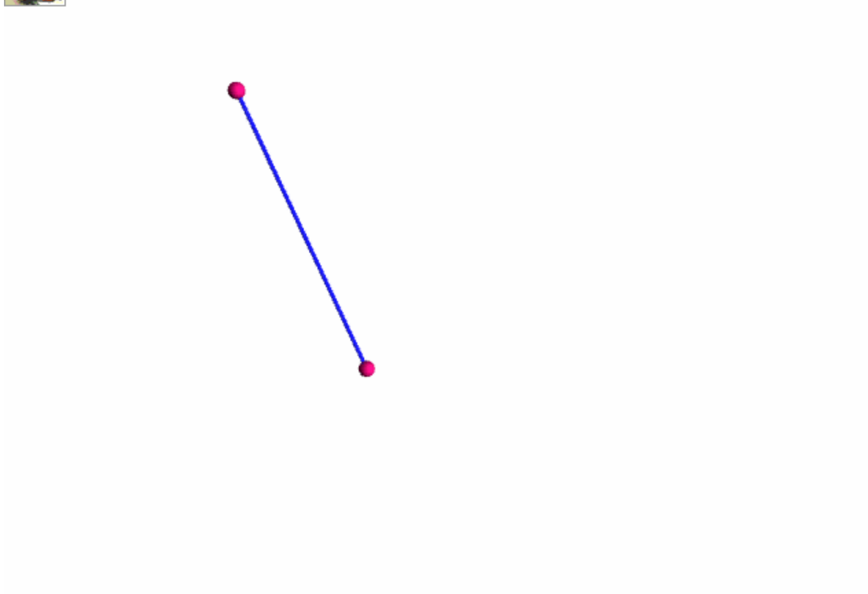
Scale-free model

(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.

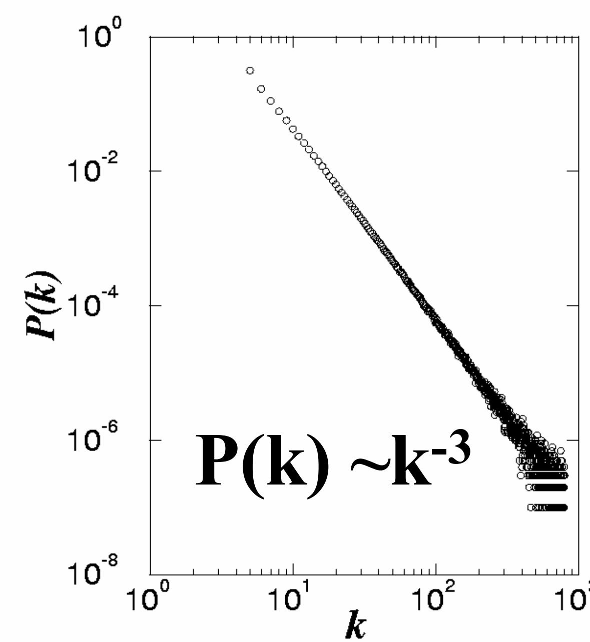
WWW : linking to well known sites
Citation : citing again highly cited papers



GROWTH:
add a new node with m links

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

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

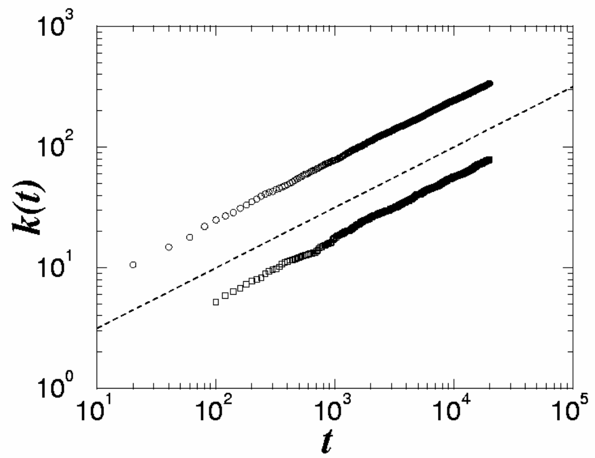


Barabási & Albert, *Science* **286,** 509 (1999)

Mean Field Theory

$$\frac{\partial k_i}{\partial t} \propto \Pi(k_i) = A\frac{k_i}{\sum_j k_j} = \frac{k_i}{2t}$$, with initial condition $k_i(t_i) = m$

$$k_i(t) = m\sqrt{\frac{t}{t_i}}$$



$$P(k_i(t) < k) = P_t(t_i > \frac{m^2 t}{k^2}) = 1 - P_t(t_i \le \frac{m^2 t}{k^2}) = 1 - \frac{m^2 t}{k^2(m_0 + t)}$$

$$\therefore P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^2 t}{m_o + t}\frac{1}{k^3} \sim k^{-3}$$

$$\boldsymbol{\gamma = 3}$$

A.-L.Barabási, R. Albert and H. Jeong, *Physica* A **272,** 173 (1999)

Can Latecomers Make It? *Fitness Model*

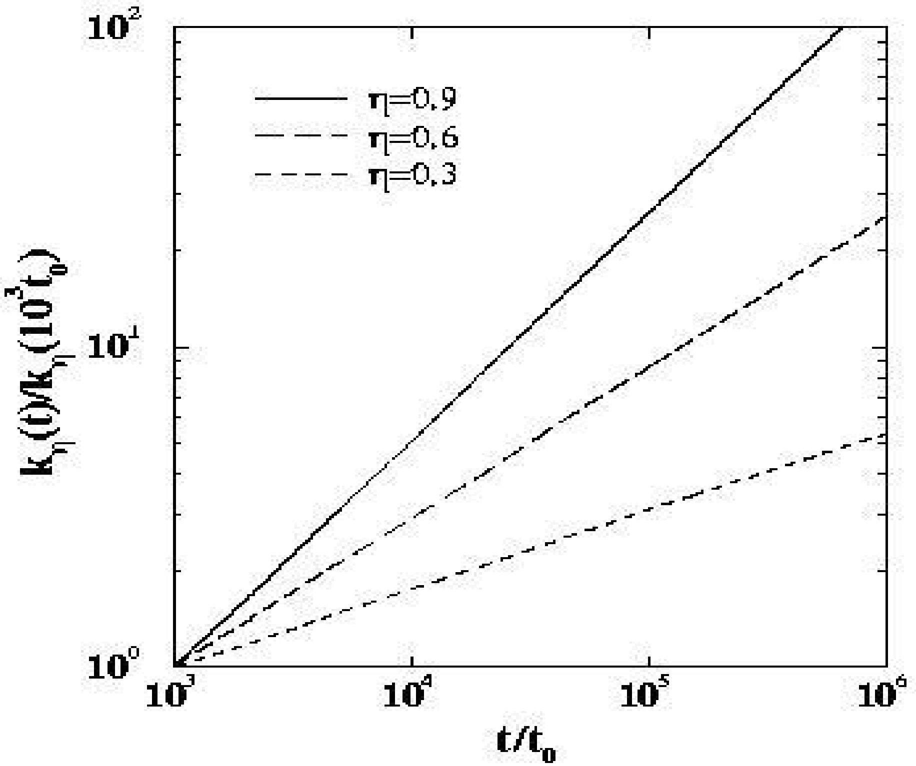
SF model: $k(t) \sim t^{1/2}$ → (first mover advantage)

Real systems: nodes compete for links -- ***fitness***

Fitness Model: fitness (η)

$$\Pi(k_i) \cong \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

$$\boldsymbol{k(\eta,t) \sim t^{\beta(\eta)}}$$

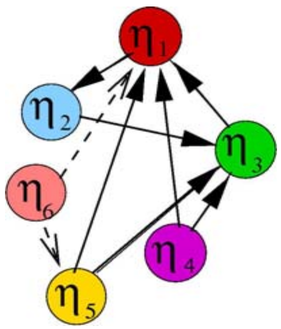


where

$$\beta(\eta) = \eta/C \qquad \int d\eta \rho(\eta) \frac{1}{C/\eta - 1} = 1$$

Bose-Einstein Condensation in Evolving Networks

Network



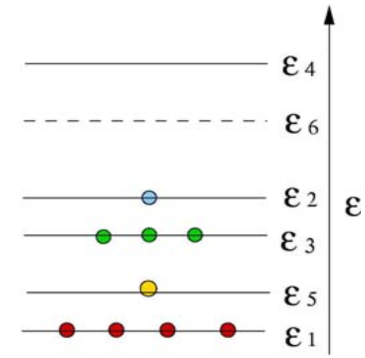
$$\Pi_i = \frac{\eta_i k_i}{\Sigma_j \eta_j k_j}$$

$$\eta \longrightarrow e^{-\beta\varepsilon}$$

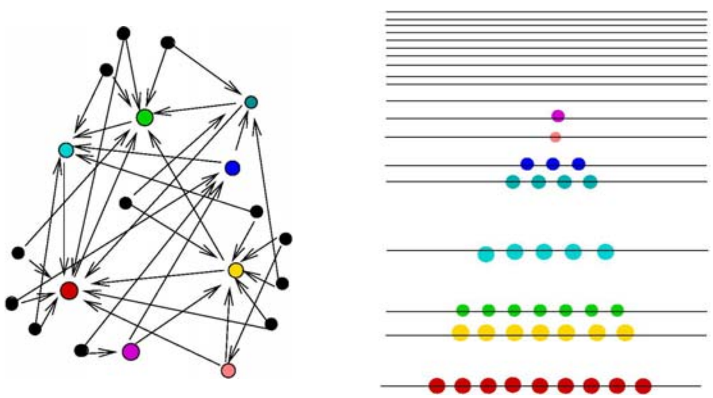
$$k_{in}(\eta) \longrightarrow n(\varepsilon) = \frac{1}{e^{-\beta\varepsilon} - 1}$$

$$\rho(\eta) \longrightarrow g(\varepsilon)$$

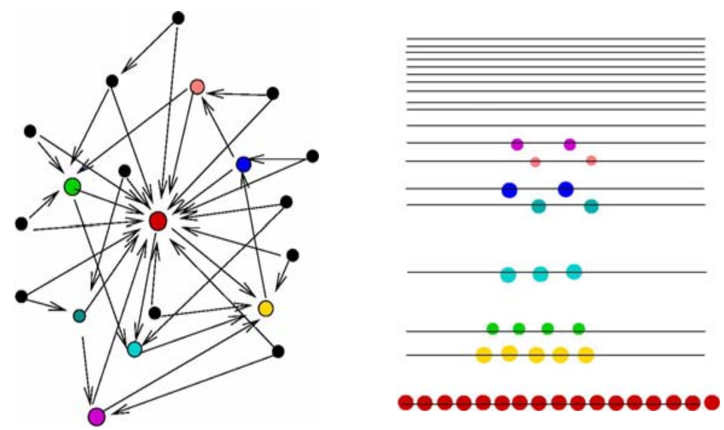
Bose gas



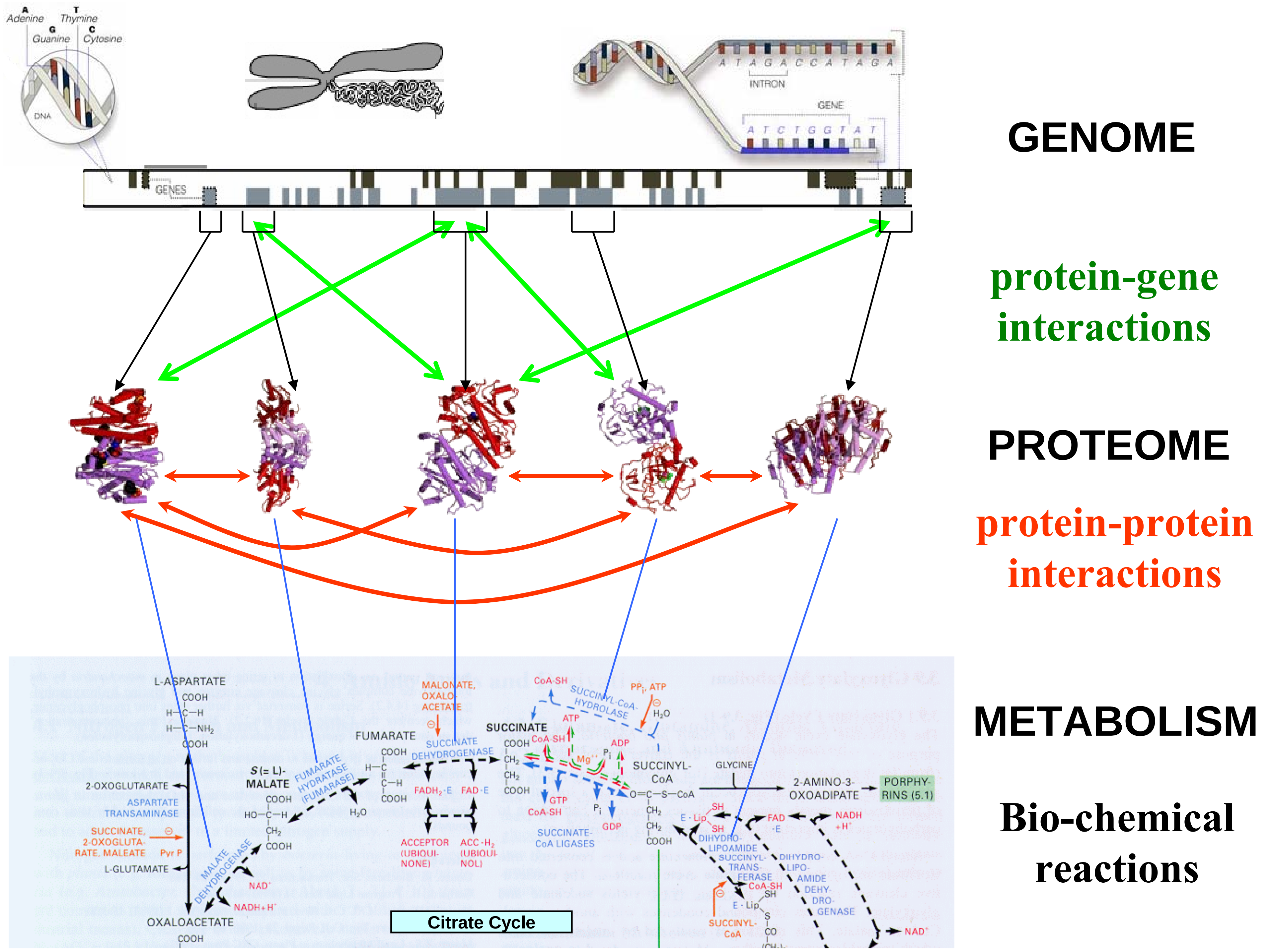
Fit-gets-rich



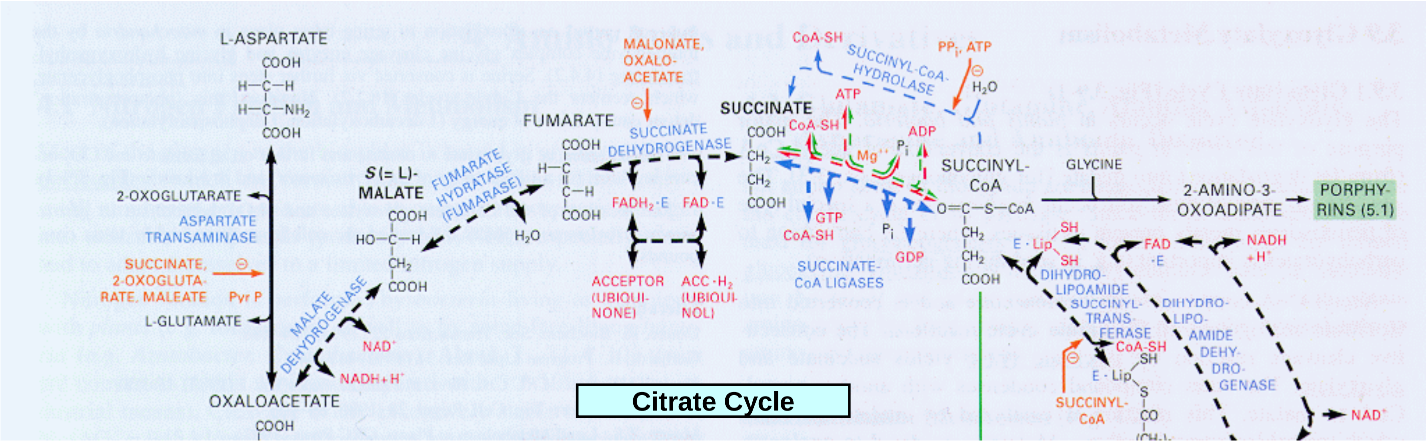
Bose-Einstein condensation



G. Bianconi and A.-L. Barabási, *Physical Review Letters* 2001; *Europhys. Lett.* 2001.



A Adenine
T Thymine
G Guanine
C Cytosine
DNA
INTRON
GENE
GENES
GENOME
protein-gene interactions
PROTEOME
protein-protein interactions
METABOLISM
Bio-chemical reactions
L-ASPARTATE
2-OXOGLUTARATE
ASPARTATE TRANSAMINASE
L-GLUTAMATE
S (= L)-MALATE
MALATE DEHYDROGENASE
NAD+
NADH+H+
OXALOACETATE
FUMARATE HYDRATASE (FUMARASE)
FUMARATE
SUCCINATE DEHYDROGENASE
MALONATE, OXALOACETATE
FADH2-E
FAD-E
ACCEPTOR (UBIQUINONE)
ACC-H2 (UBIQUINOL)
SUCCINATE
SUCCINYL-CoA HYDROLASE
SUCCINATE-CoA LIGASES
SUCCINYL-CoA
GLYCINE
2-AMINO-3-OXOADIPATE
PORPHYRINS (5.1)
DIHYDROLIPOAMIDE SUCCINYLTRANSFERASE
DIHYDROLIPOAMIDE DEHYDROGENASE
Citrate Cycle

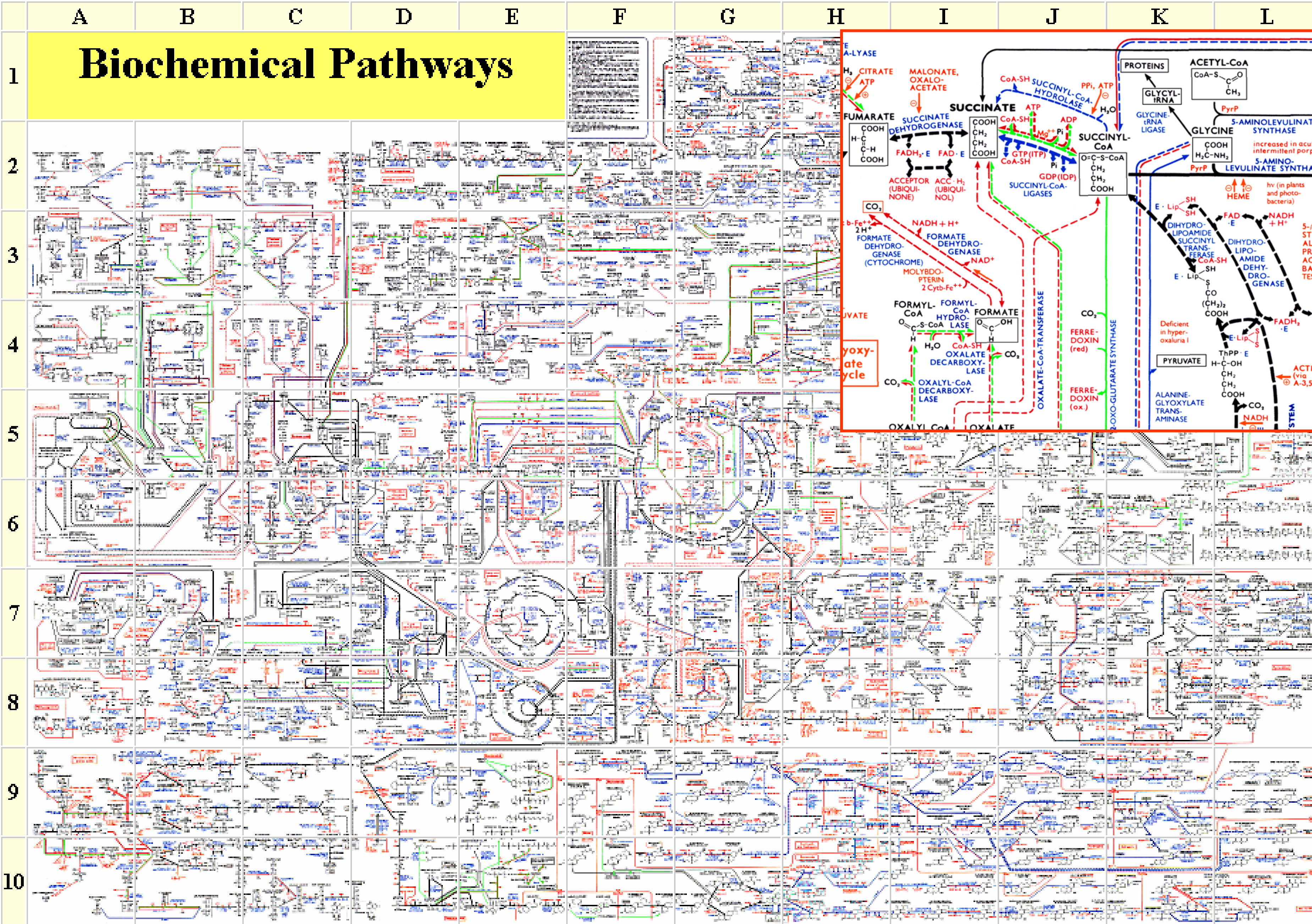


Citrate Cycle

METABOLISM

Bio-chemical reactions

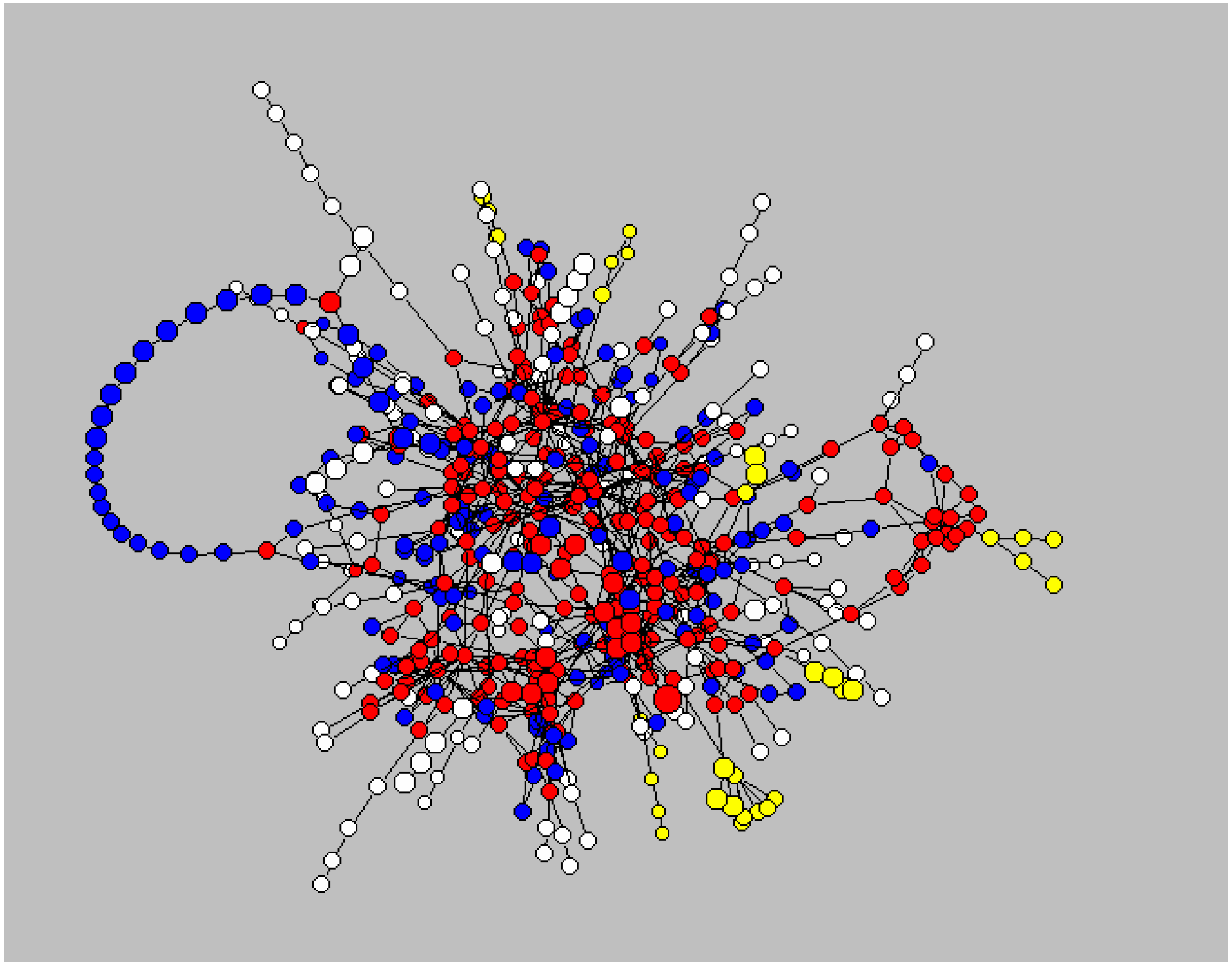
Biochemical Pathways



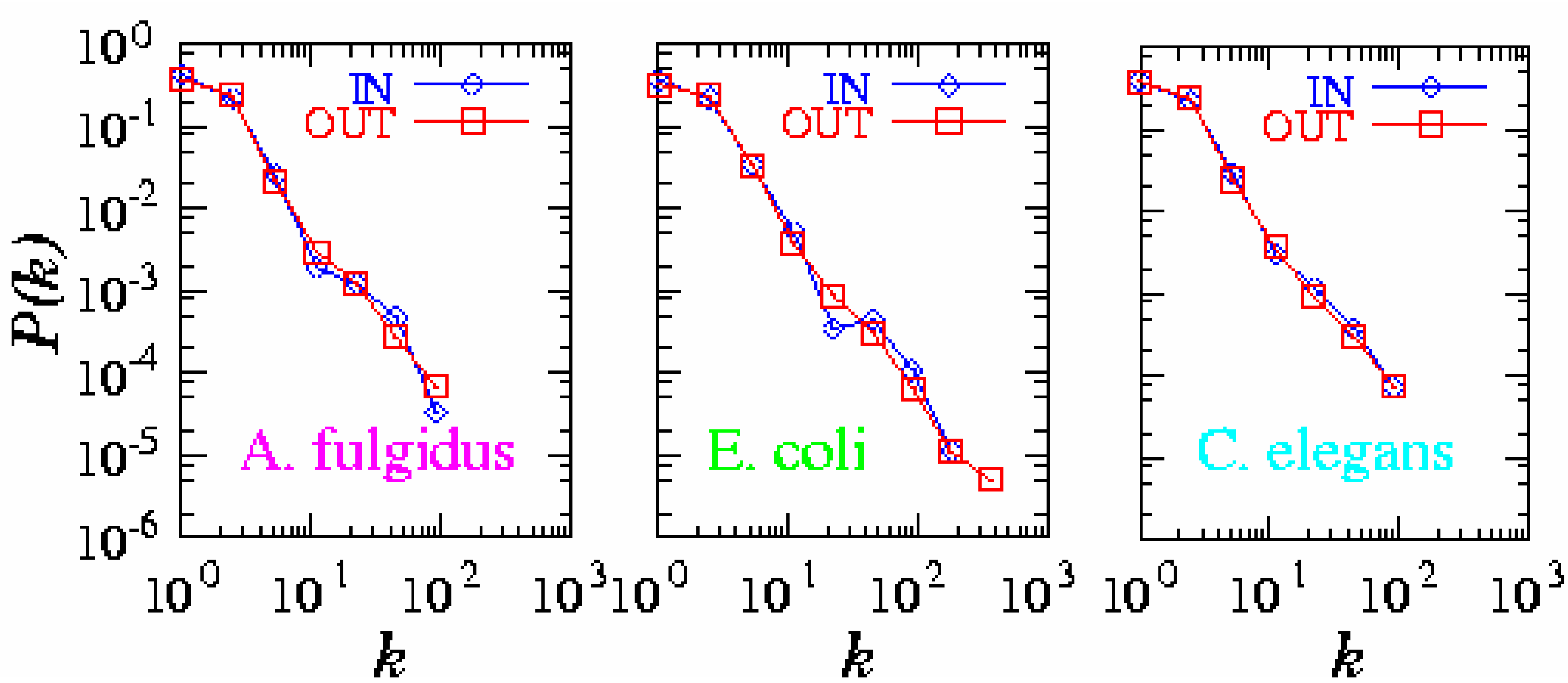
Metabolic Network

Nodes: chemicals (substrates)

Links: bio-chemical reactions



Metabolic network



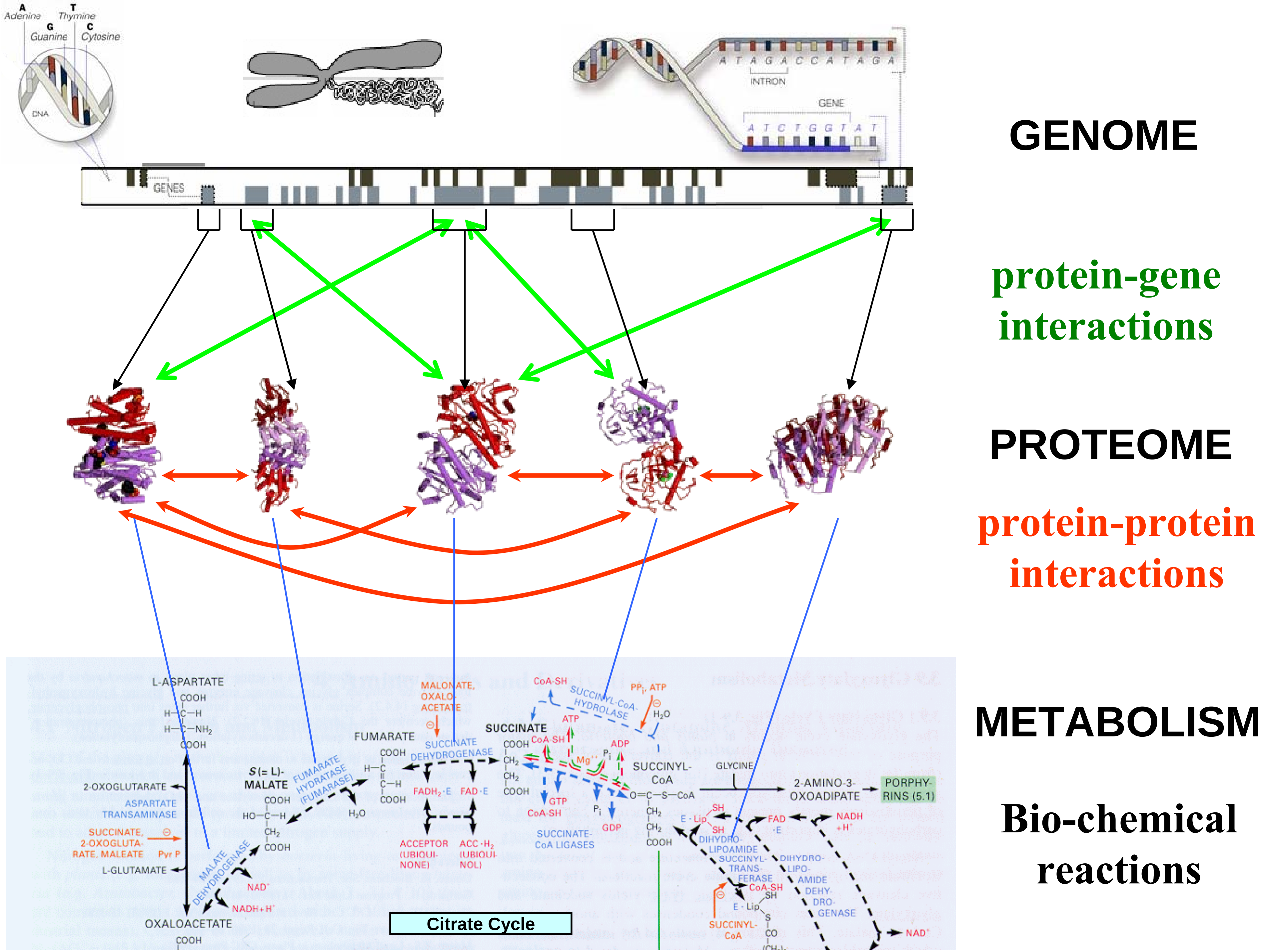
Archaea

Bacteria

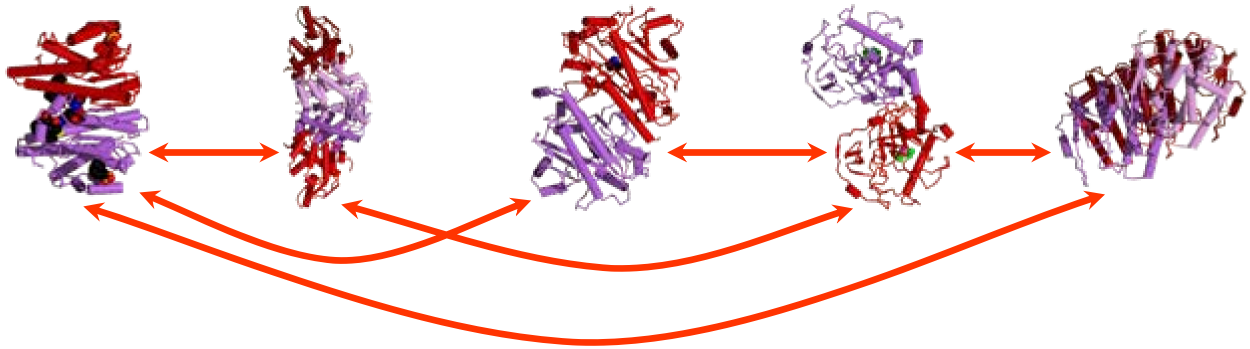
Eukaryotes

Organisms from all three domains of life are **scale-free** networks!

H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, *Nature*, **407** 651 (2000)



A Adenine
T Thymine
G Guanine
C Cytosine
DNA
A T A G A C C A T A G A
INTRON
GENE
A T C T G G T A T
GENES
GENOME
protein-gene interactions
PROTEOME
protein-protein interactions
METABOLISM
Bio-chemical reactions
L-ASPARTATE
2-OXOGLUTARATE
ASPARTATE TRANSAMINASE
L-GLUTAMATE
OXALOACETATE
MALATE DEHYDROGENASE
NAD+
NADH+H+
S(= L)-MALATE
FUMARATE HYDRATASE (FUMARASE)
H2O
FUMARATE
MALONATE, OXALO-ACETATE
SUCCINATE DEHYDROGENASE
FADH2-E
FAD-E
ACCEPTOR (UBIQUINONE)
ACC-H2 (UBIQUINOL)
SUCCINATE
CoA-SH
SUCCINYL-CoA-HYDROLASE
PPi, ATP
H2O
ATP
ADP
GTP
GDP
SUCCINATE-CoA LIGASES
SUCCINYL-CoA
GLYCINE
2-AMINO-3-OXOADIPATE
PORPHYRINS (5.1)
DIHYDROLIPOAMIDE SUCCINYL-TRANSFERASE
DIHYDROLIPOAMIDE DEHYDROGENASE
FAD
NADH +H+
NAD+
SUCCINYL-CoA
Citrate Cycle



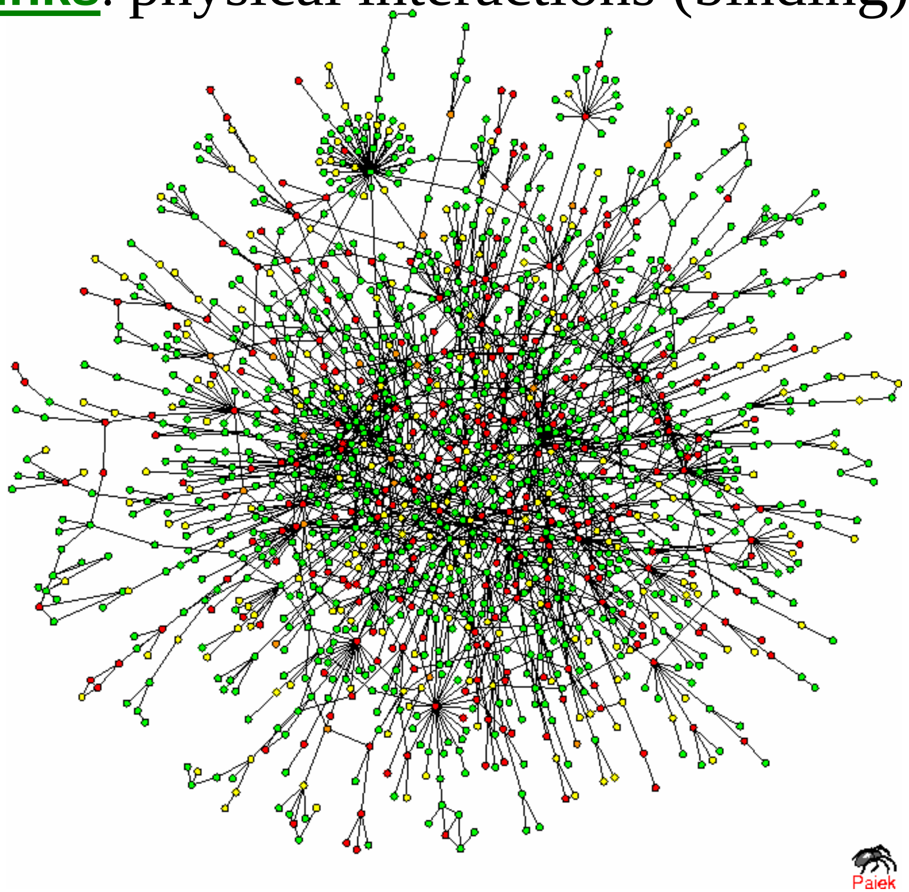
PROTEOME

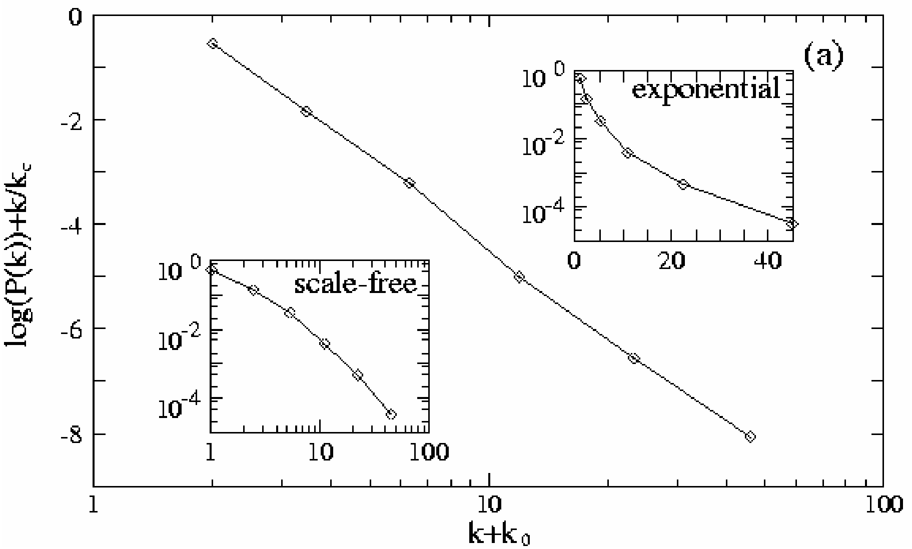
protein-protein
interactions

Topology of the protein network

Nodes: proteins

Links: physical interactions (binding)



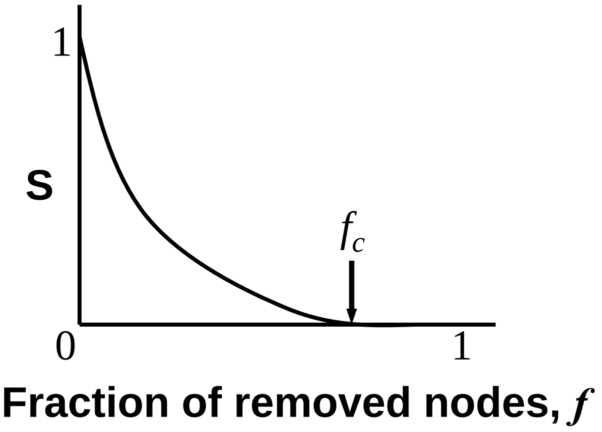


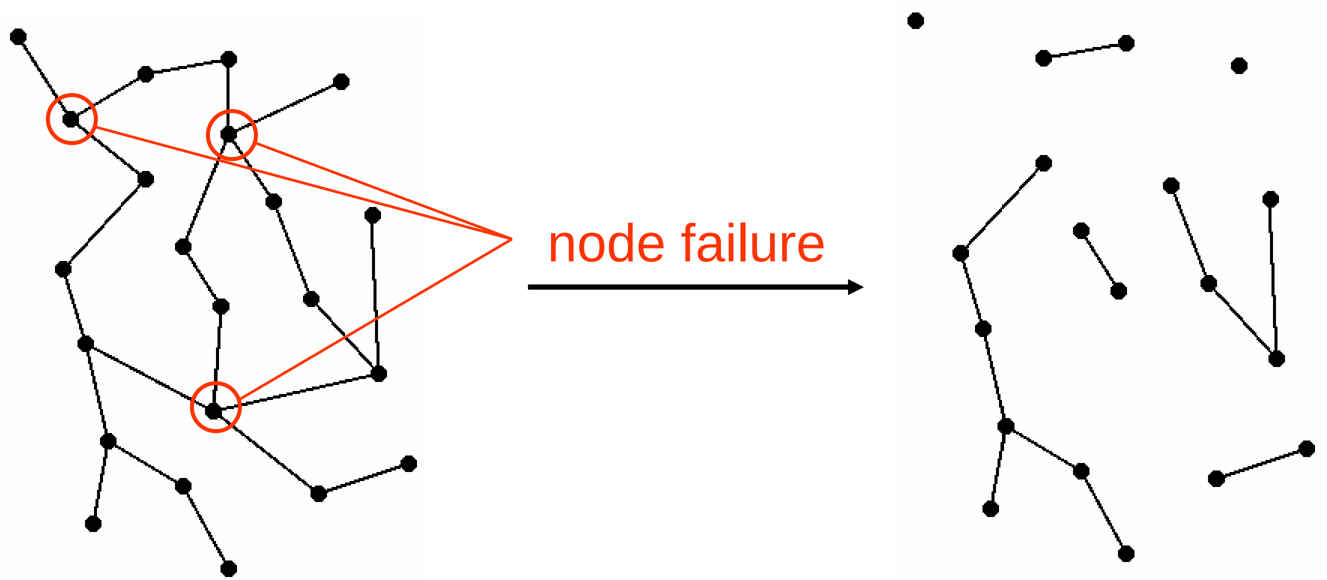
$$P(k) \sim (k + k_0)^{-\gamma} \exp(-\frac{k + k_0}{k_\tau})$$

H. Jeong, S.P. Mason, A.-L. Barabasi, Z.N. Oltvai, *Nature* 411, 41-42 (2001)

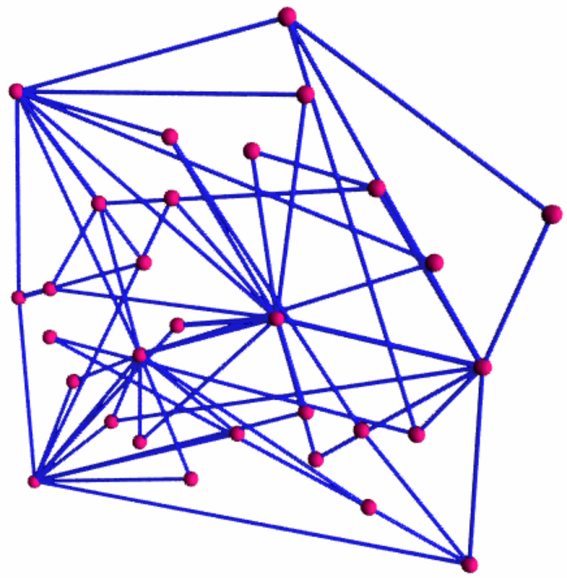
Robustness

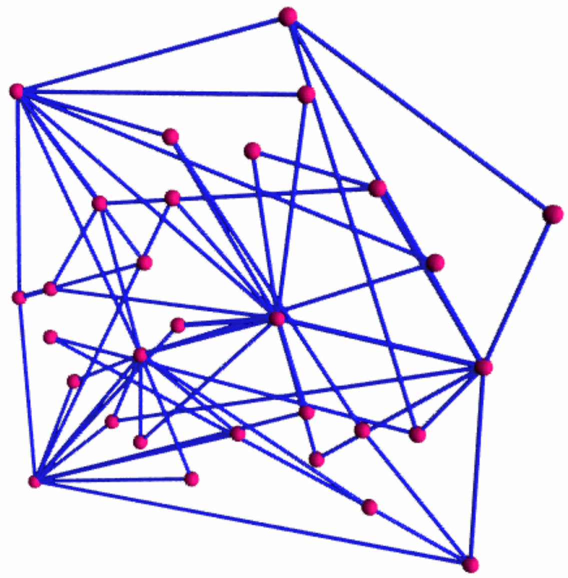
Complex systems maintain their basic functions
even under errors and failures
(cell → mutations; Internet → router breakdowns)



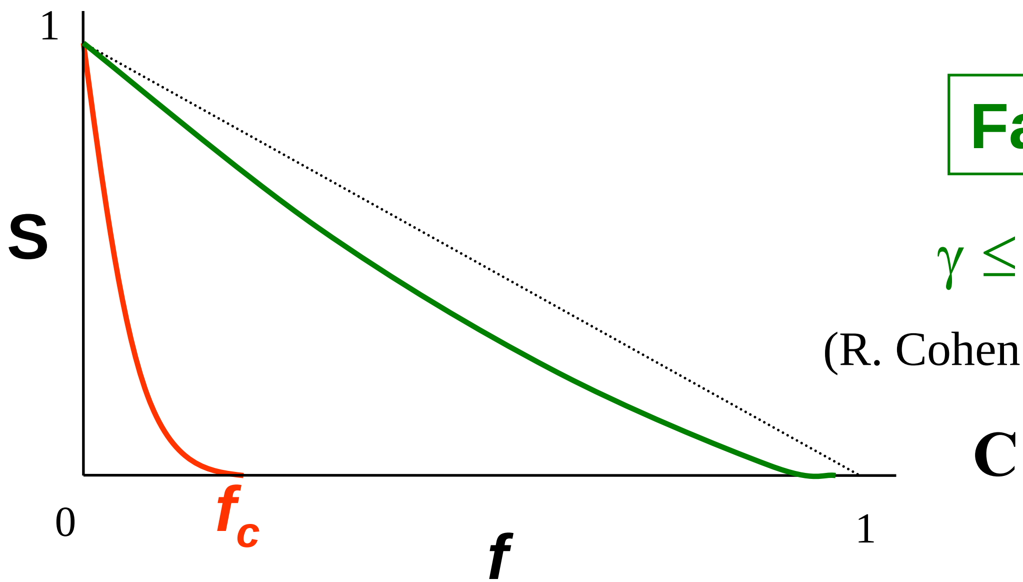


Robustness of scale-free networks





Attacks



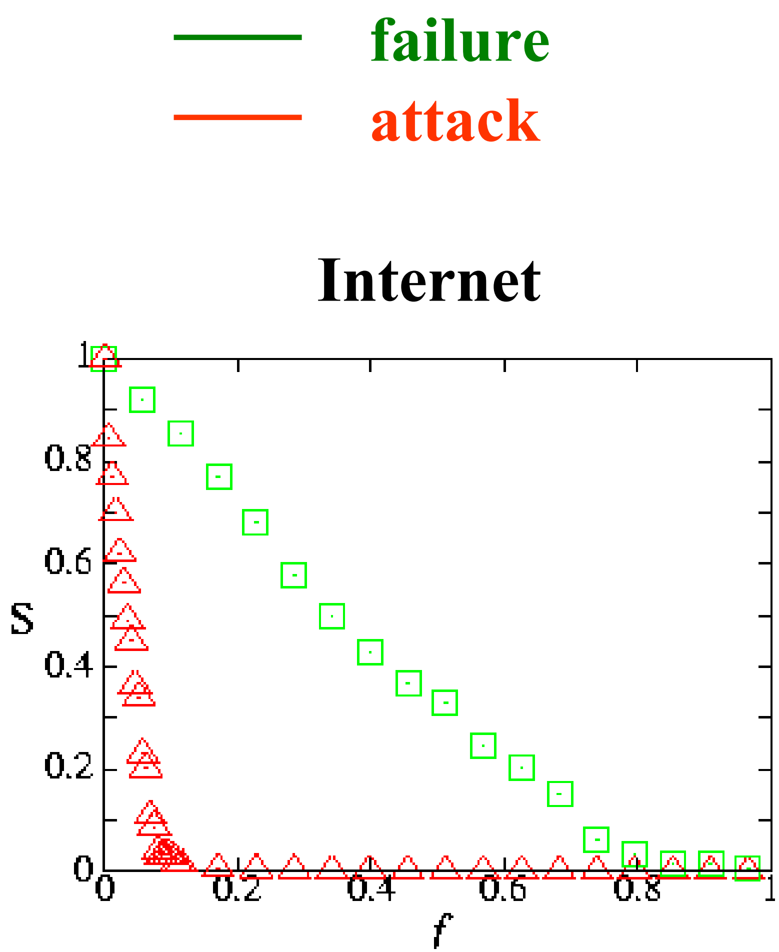
Failures

$\gamma \leq 3 : f_c = 1$

(R. Cohen et al PRL, 2000)

Albert, Jeong, Barabasi, *Nature* **406** 378 (2000)

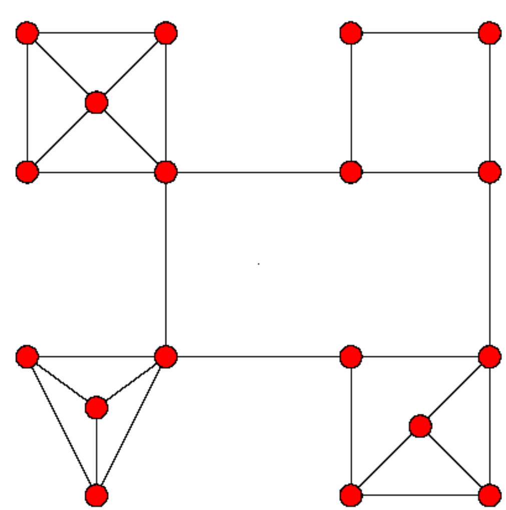
Achilles' Heel of complex networks

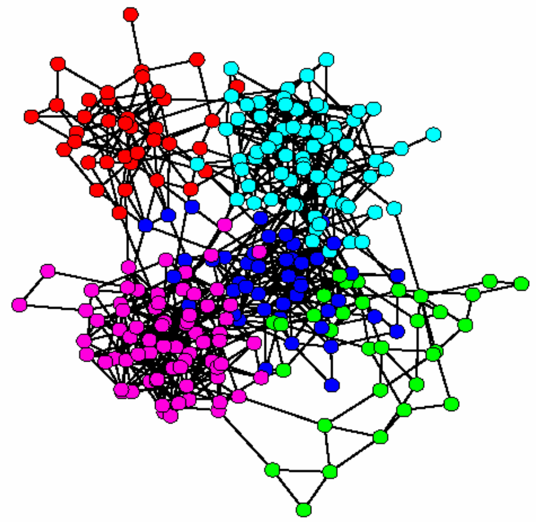


R. Albert, H. Jeong, A.L. Barabasi, *Nature* **406** 378 (2000)

Modularity

- Real networks are fragmented into group or modules
  - 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).
  - Internet: Vasquez, Pastor-Satorras, Vespignani(2001).
- Traditional view of modularity:

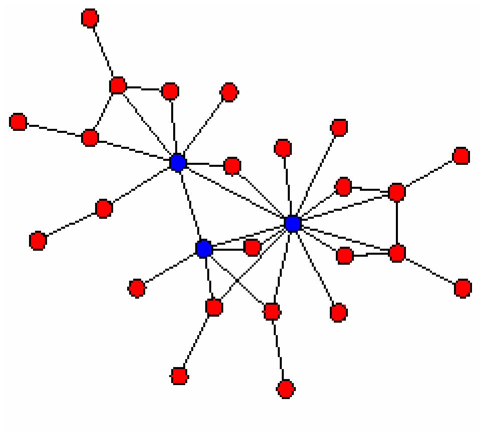




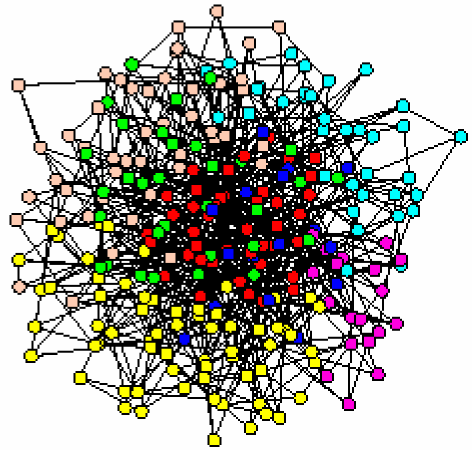
Ravasz, Somera, Mongru, Oltvai, A-L. B, ***Science* 297,** 1551 (2002).

Modular vs. Scale-free Topology

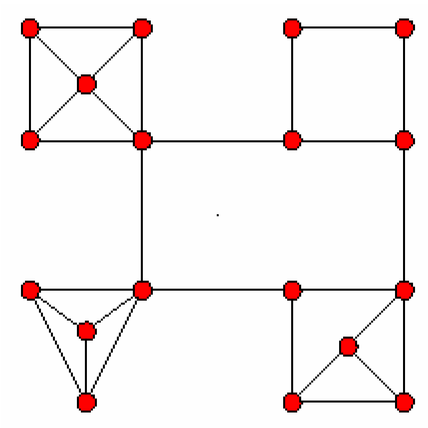
(a)



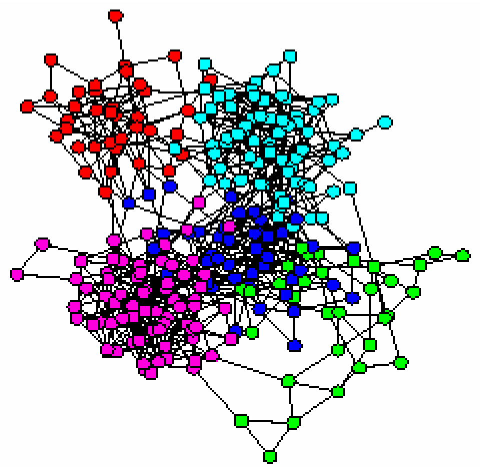
Scale-free



(b)



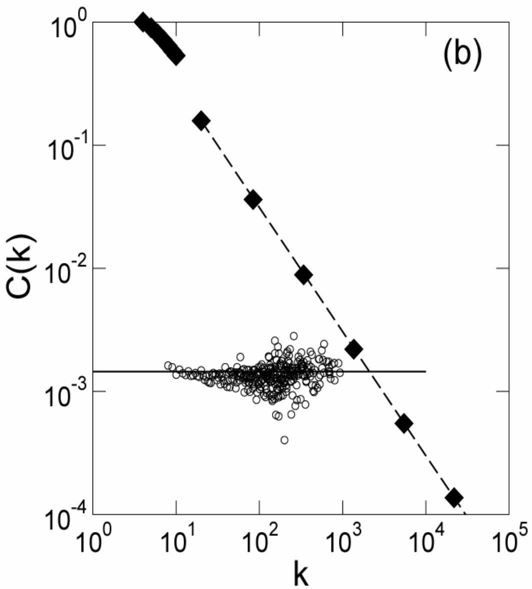
Modular

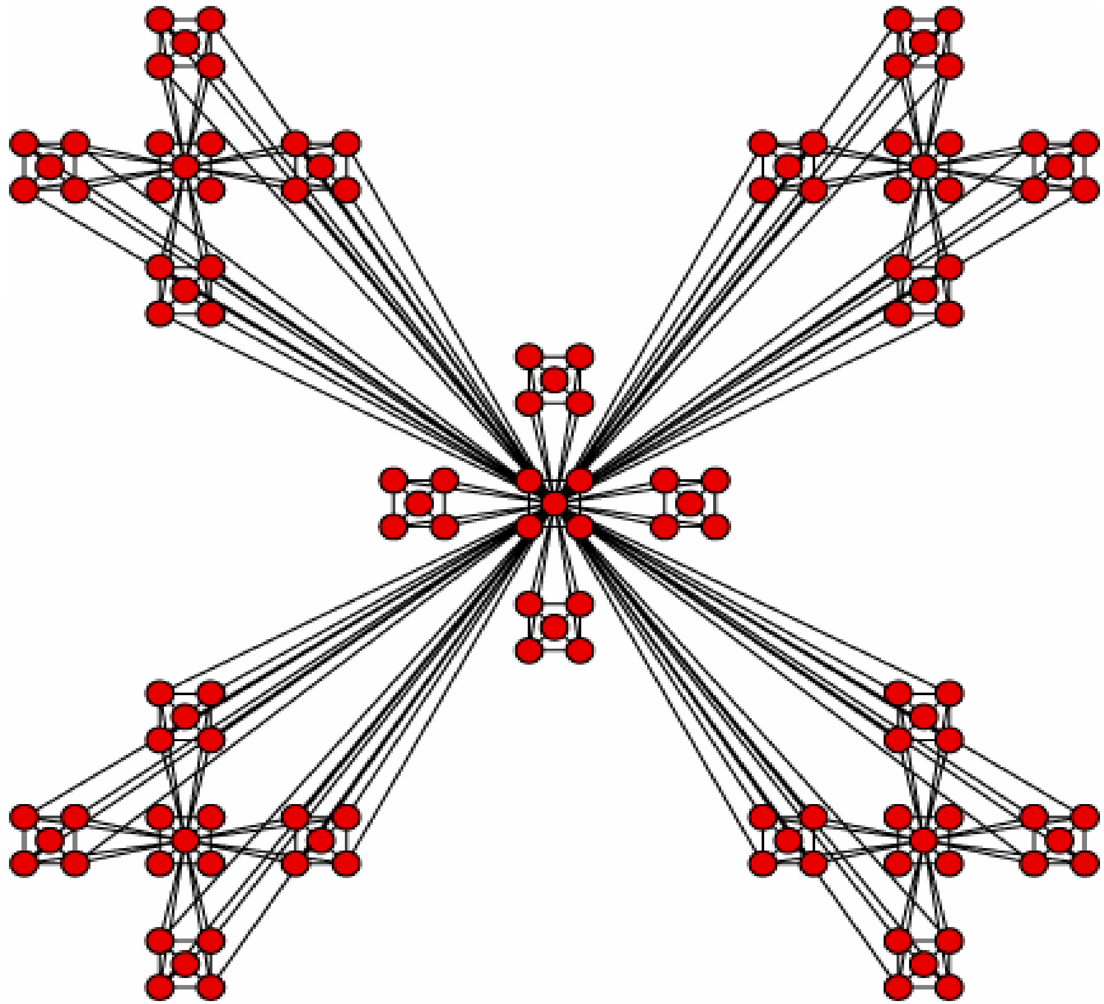


Hierarchical Networks

3. Clustering coefficient scales

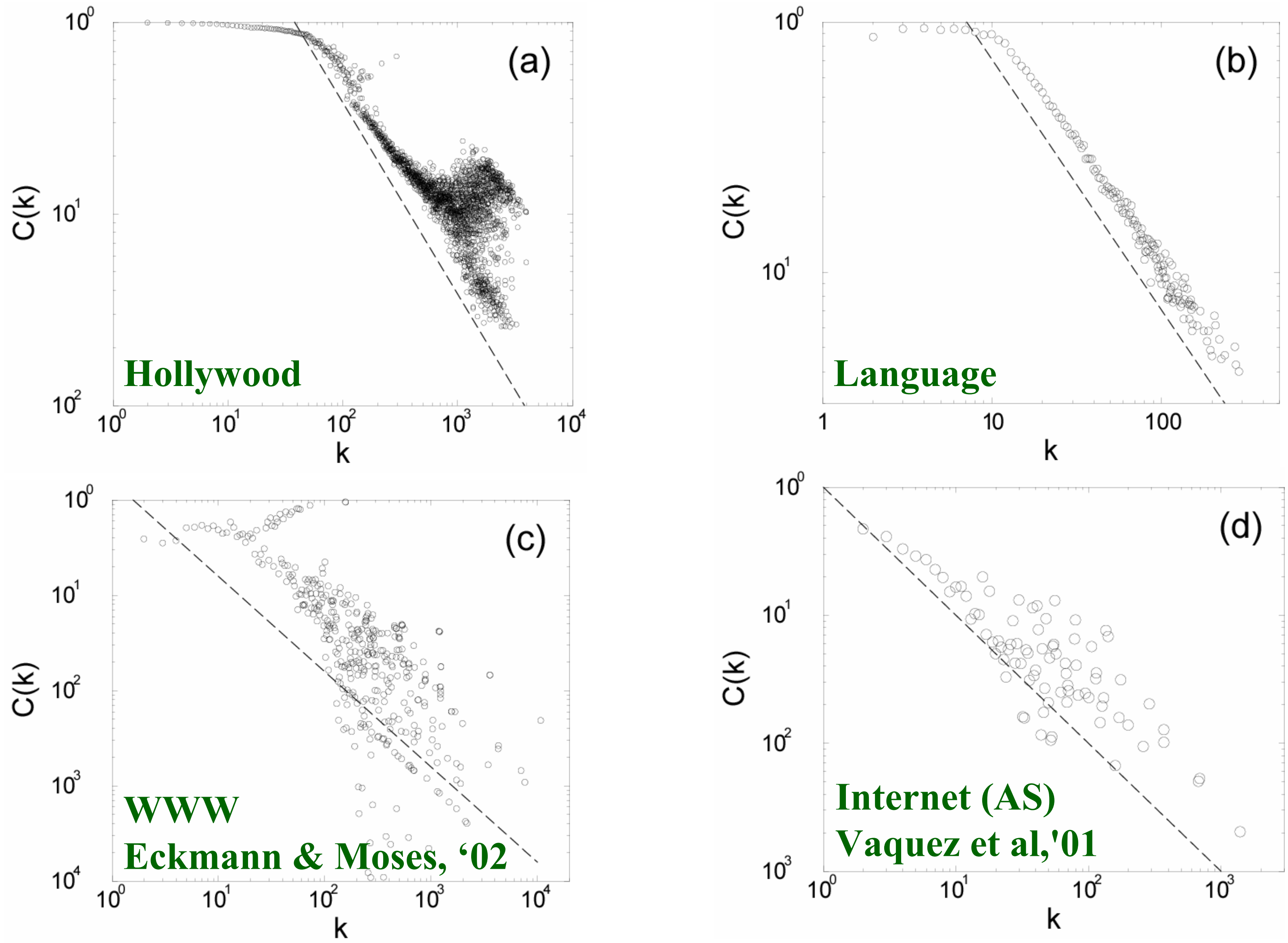
$$C(k) \sim k^{-1}$$



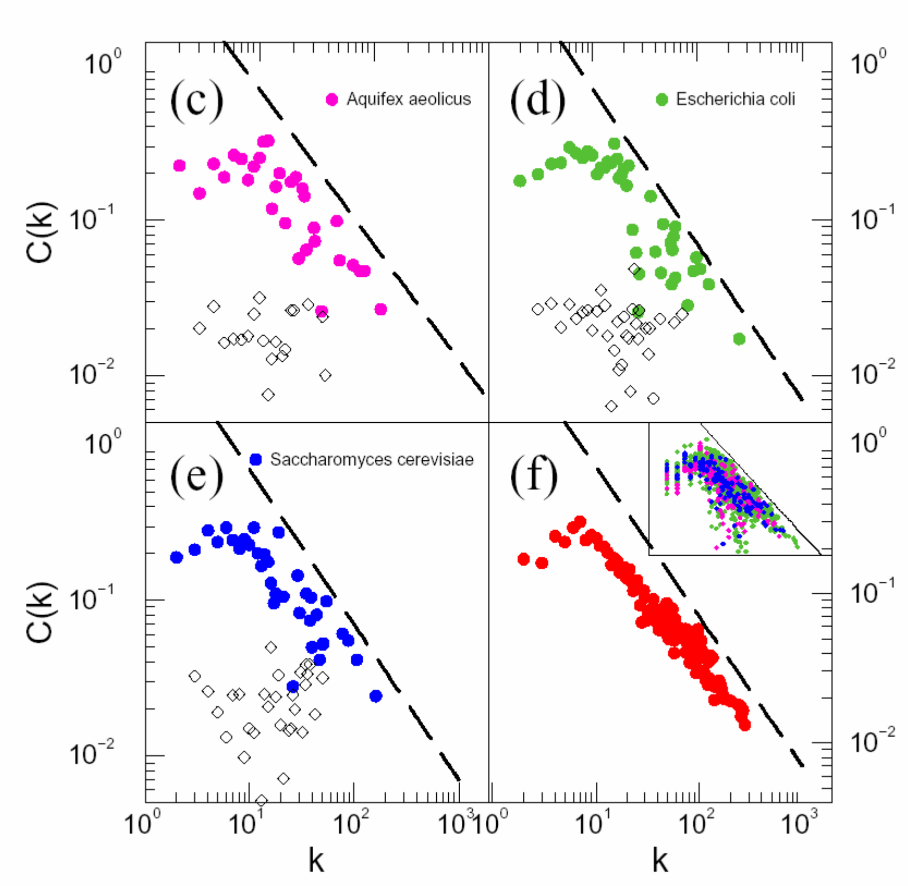


$$C(k) = \frac{\#\text{ links between k neighbors}}{k(k-1)/2}$$

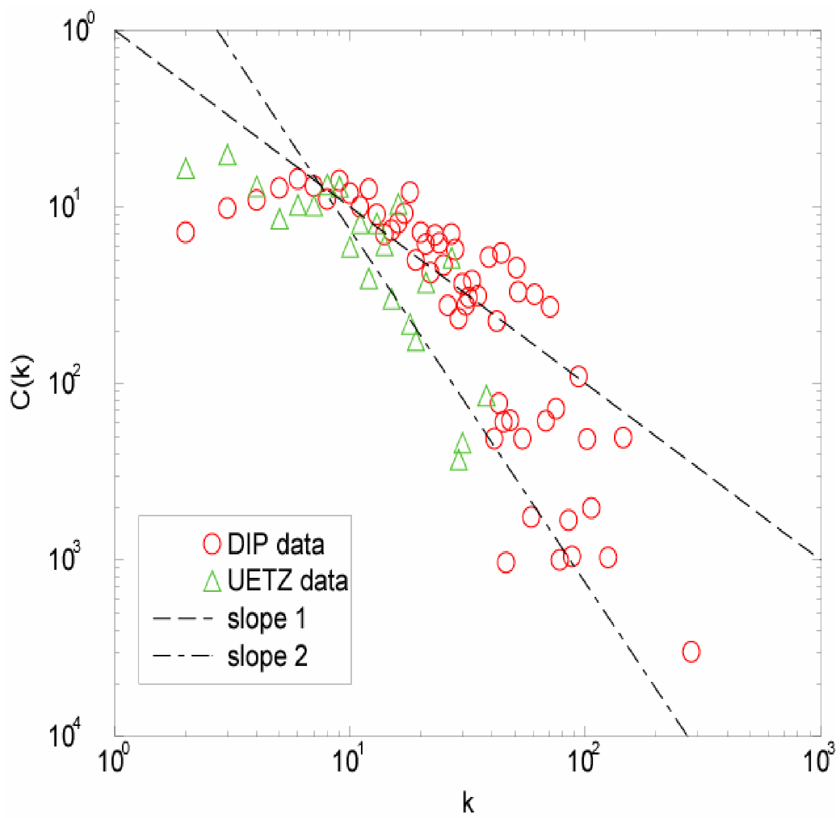
Real Networks



Hierarchy in biological systems

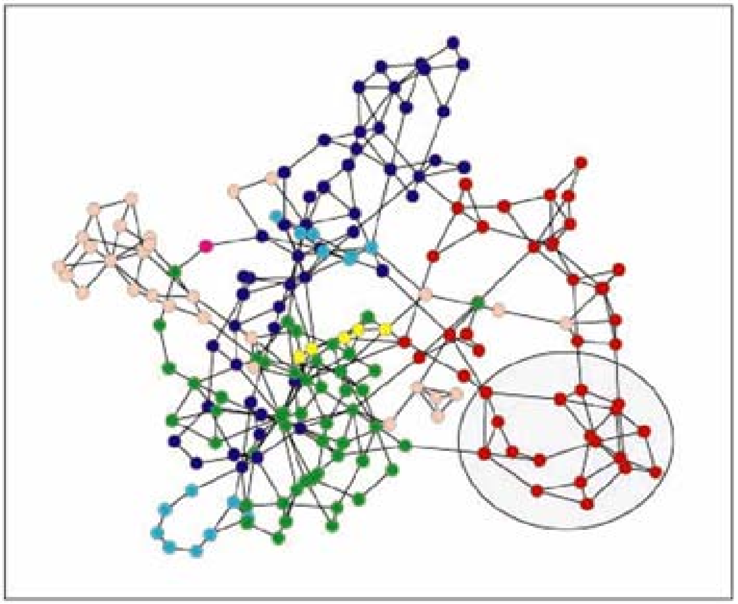


Metabolic networks

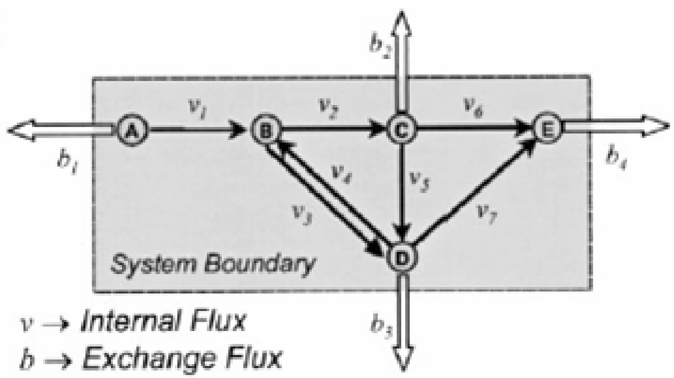


Protein networks

Characterizing the links



Metabolism:
Flux Balance Analysis (Palsson)
Metabolic flux for each reaction



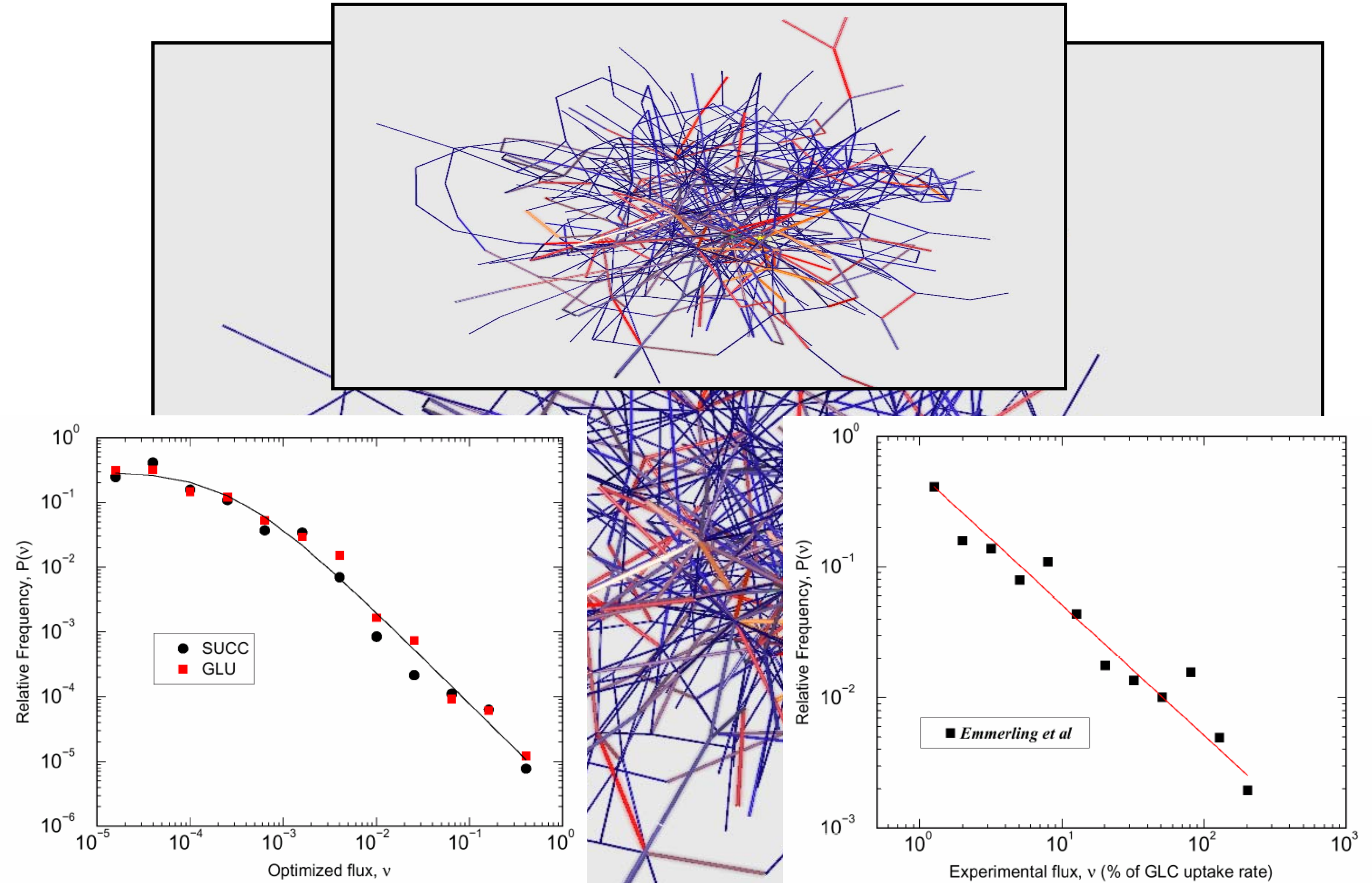
Balance Equations:

A: $-v_1 - b_1 = 0$
B: $v_1 + v_4 - v_2 - v_3 = 0$
C: $v_2 - v_5 - v_6 - b_2 = 0$
D: $v_3 + v_5 - v_4 - v_7 - b_3 = 0$
E: $v_6 + v_7 - b_4 = 0$

Matrix Notation
$\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$

Edwards, J. S. & Palsson, B. O, *PNAS* **97**, 5528 (2000).
Edwards, J. S., Ibarra, R. U. & Palsson, B. O. *Nat Biotechnol* **19**, 125 (2001).
Ibarra, R. U., Edwards, J. S. & Palsson, B. O. *Nature* **420**, 186 (2002).

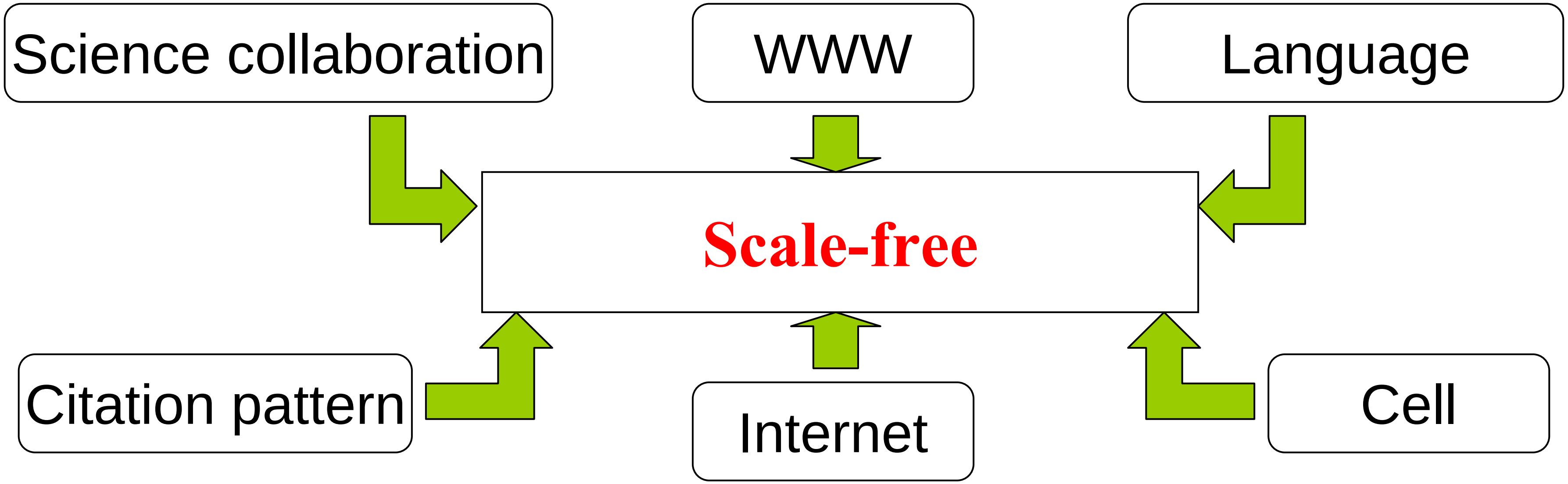
Global flux organization in the *E. coli* metabolic network



SUCC: Succinate uptake
GLU : Glutamate uptake

Central Metabolism,
Emmerling et. al, *J Bacteriol* **184**, 152 (2002)

E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai, A.-L. B. Nature, 2004; Goh et al, PRL 2002.



Hierarchical Networks

Where do we go from here?…

⇨ **How topology affects function?**

⇨ **Dynamics on networks: Are there universal properties?**

http://www.nd.edu/~networks

http://www.nd.edu/~networks

There may be a postdoctoral position open in my research group.
For more details see www.nd.edu/~networks

Traditional modeling: Network as a static graph

Given a network with N nodes and L links

$\Downarrow$

Create a graph with statistically identical topology

RESULT: model the static network topology

PROBLEM: **Real networks are dynamical systems!**

Evolving networks

OBJECTIVE: capture the network dynamics

METHOD :
- identify the processes that contribute to the network topology
- develop dynamical models that capture these processes

$\Downarrow$

BONUS: **get the topology correctly**.

Bonus: Why Kevin Bacon?

Measure the average distance between Kevin Bacon and all other actors.

Kevin Bacon

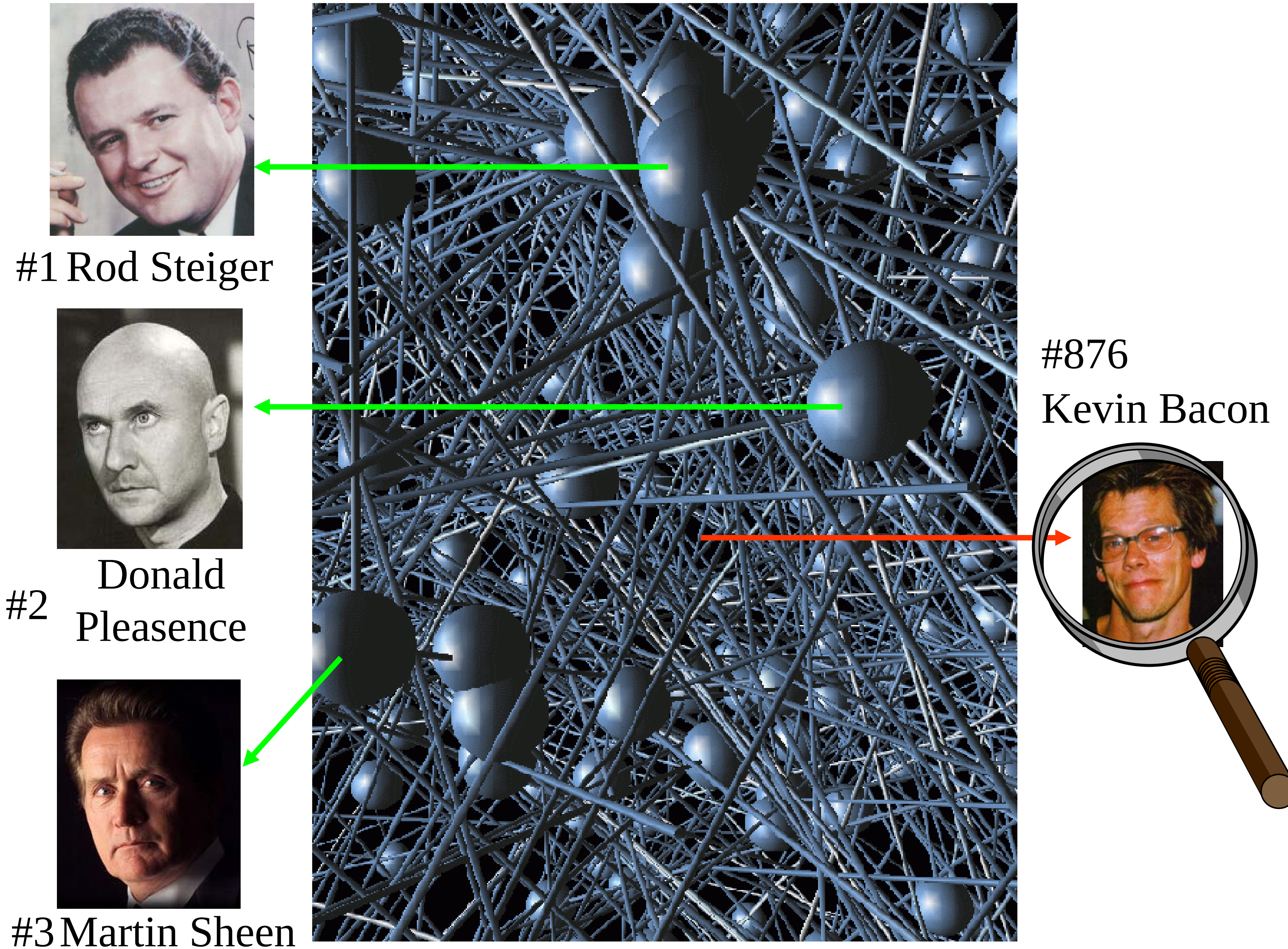
No. of movies : 46   No. of actors : 1811

Average separation: 2.79

Is Kevin Bacon the most connected actor?

NO!

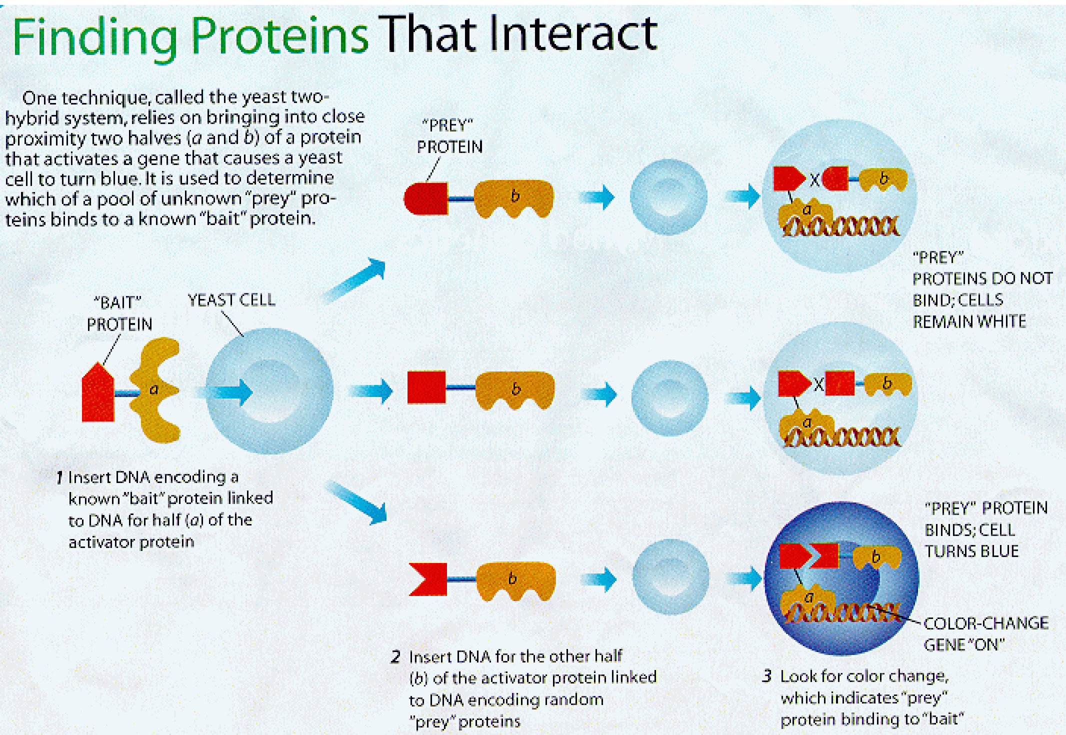
Rank	Name	Average distance	# of movies	# of links
1	Rod Steiger	2.537527	112	2562
2	Donald Pleasence	2.542376	180	2874
3	Martin Sheen	2.551210	136	3501
4	Christopher Lee	2.552497	201	2993
5	Robert Mitchum	2.557181	136	2905
6	Charlton Heston	2.566284	104	2552
7	Eddie Albert	2.567036	112	3333
8	Robert Vaughn	2.570193	126	2761
9	Donald Sutherland	2.577880	107	2865
10	John Gielgud	2.578980	122	2942
11	Anthony Quinn	2.579750	146	2978
12	James Earl Jones	2.584440	112	3787
…				
876	**Kevin Bacon**	**2.786981**	**46**	**1811**
…				

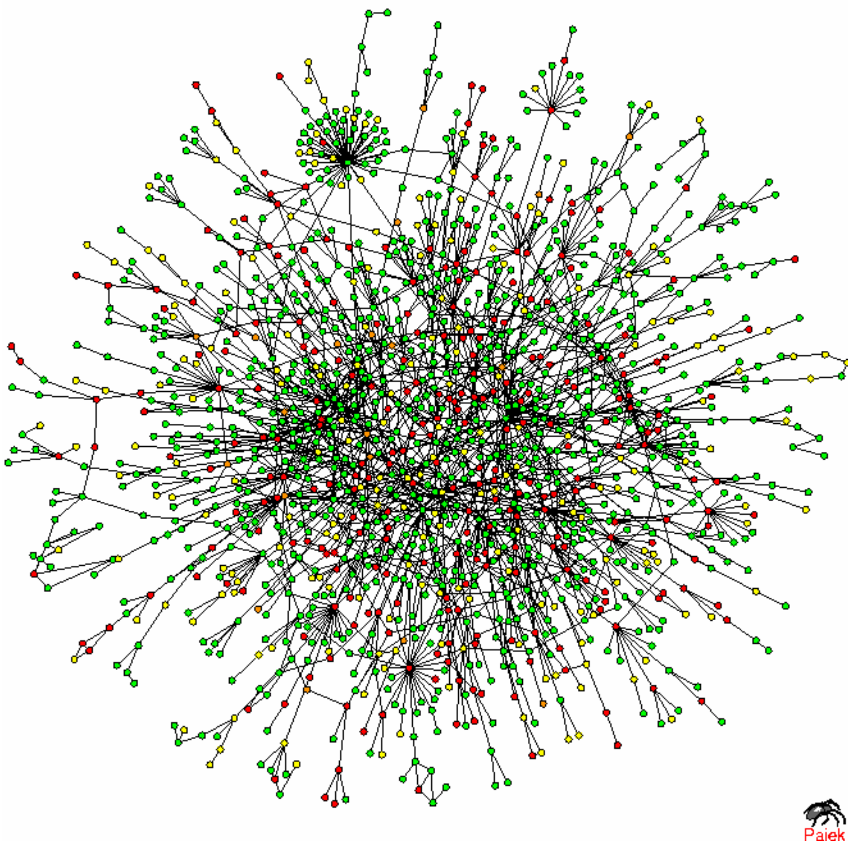

#1 Rod Steiger
#2 Donald Pleasence
#3 Martin Sheen
#876 Kevin Bacon

Protein network

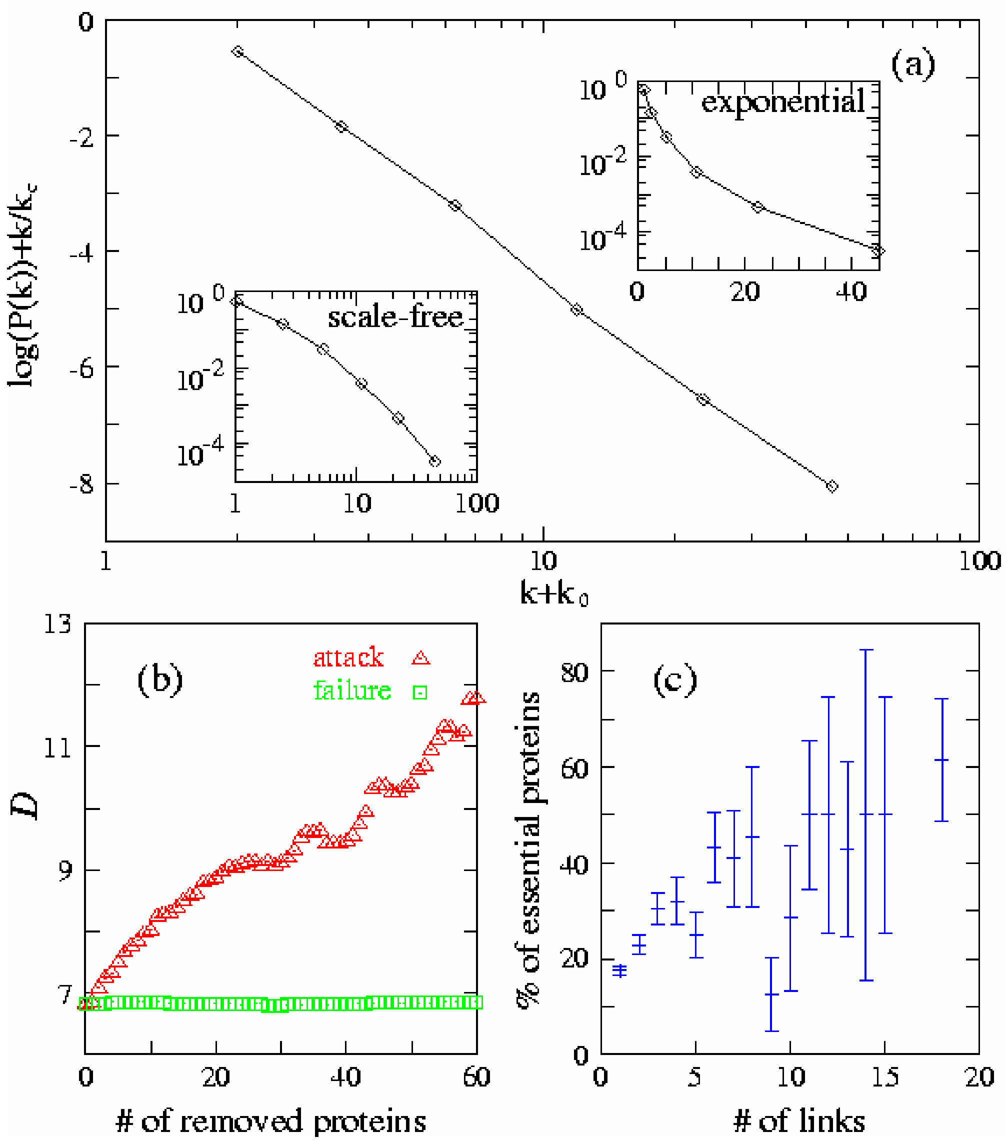
Nodes: proteins
Links: physical interaction (binding)

Proteomics : identify and determine the properties of the proteins. (related to structure of proteins)





Properties of the protein network



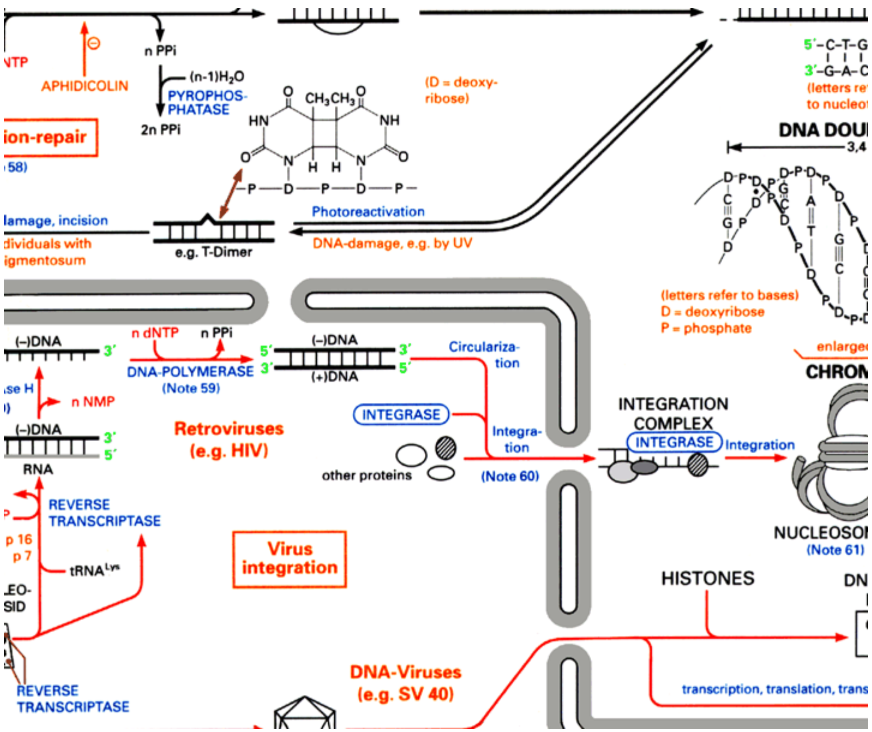
$$P(k) \sim (k+k_0)^{-\gamma} \exp(-\frac{k+k_0}{k_\tau})$$

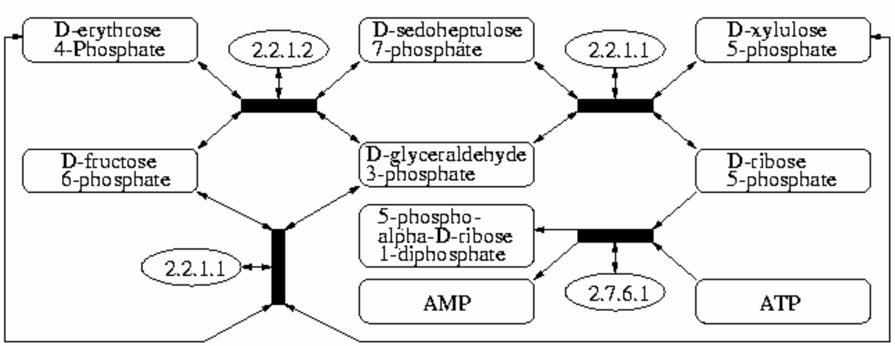
Highly connected proteins are more **essential (lethal)** than less connected proteins.

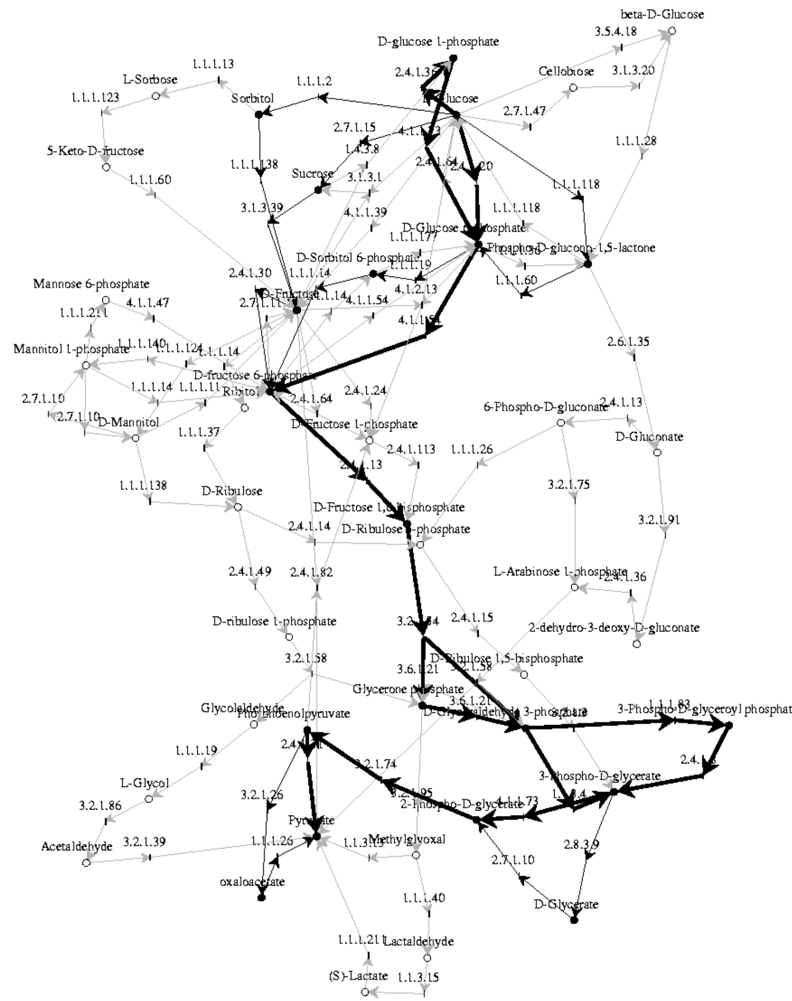
Metabolic Network

Nodes: chemicals (substrates)

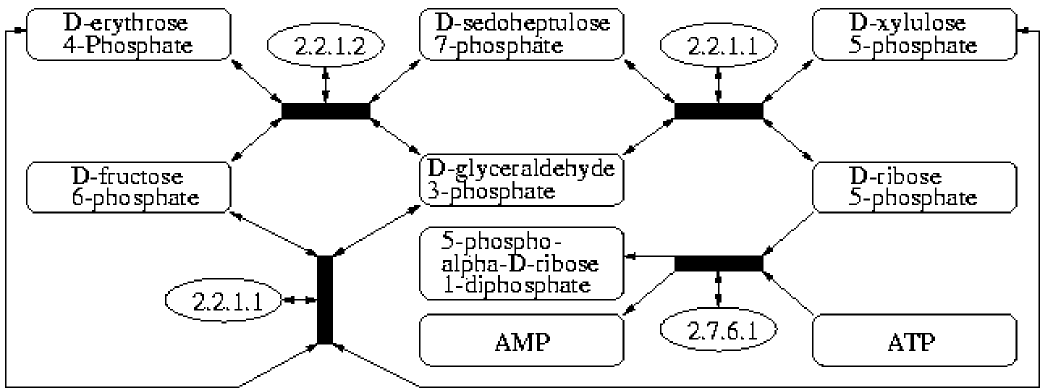
Links: chem. reaction

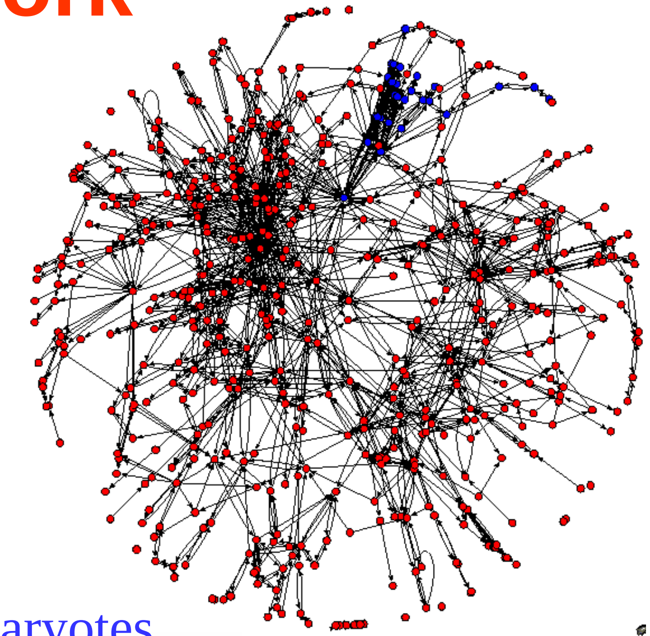


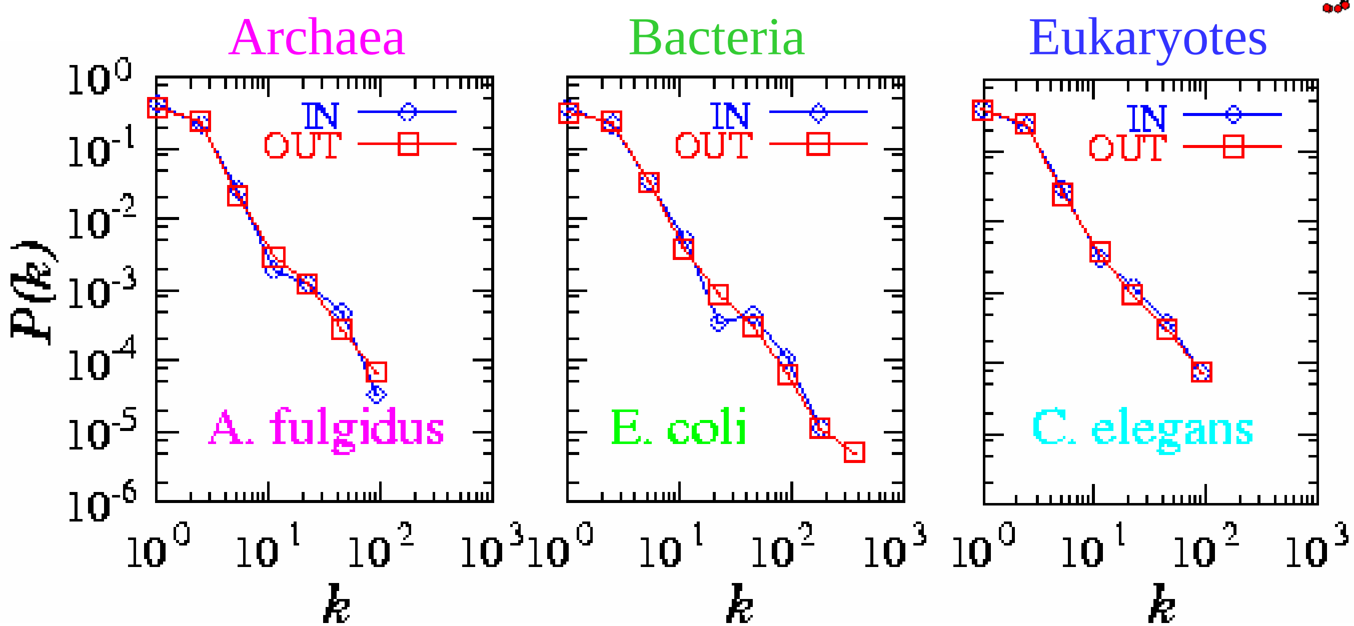




Metabolic network



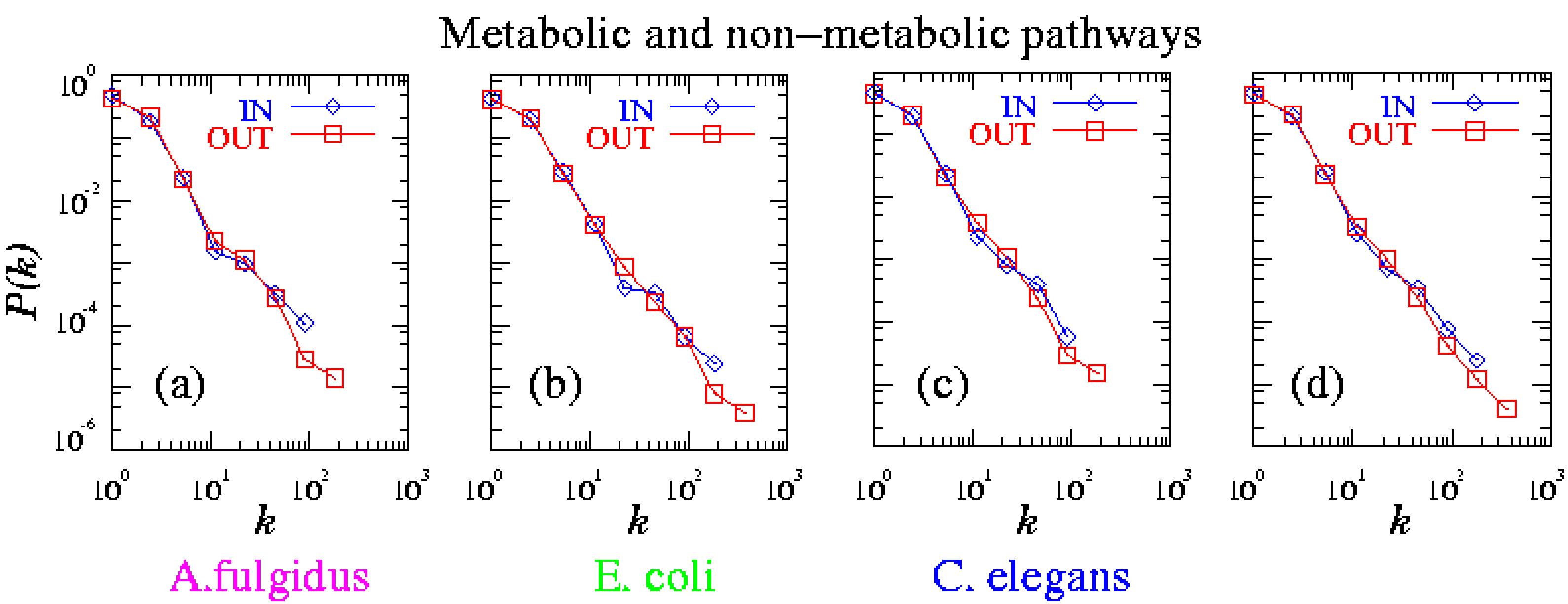




Organisms from all three domains of life are **scale-free** networks!

H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, Nature, **407** 651 (2000)

Whole cellular network



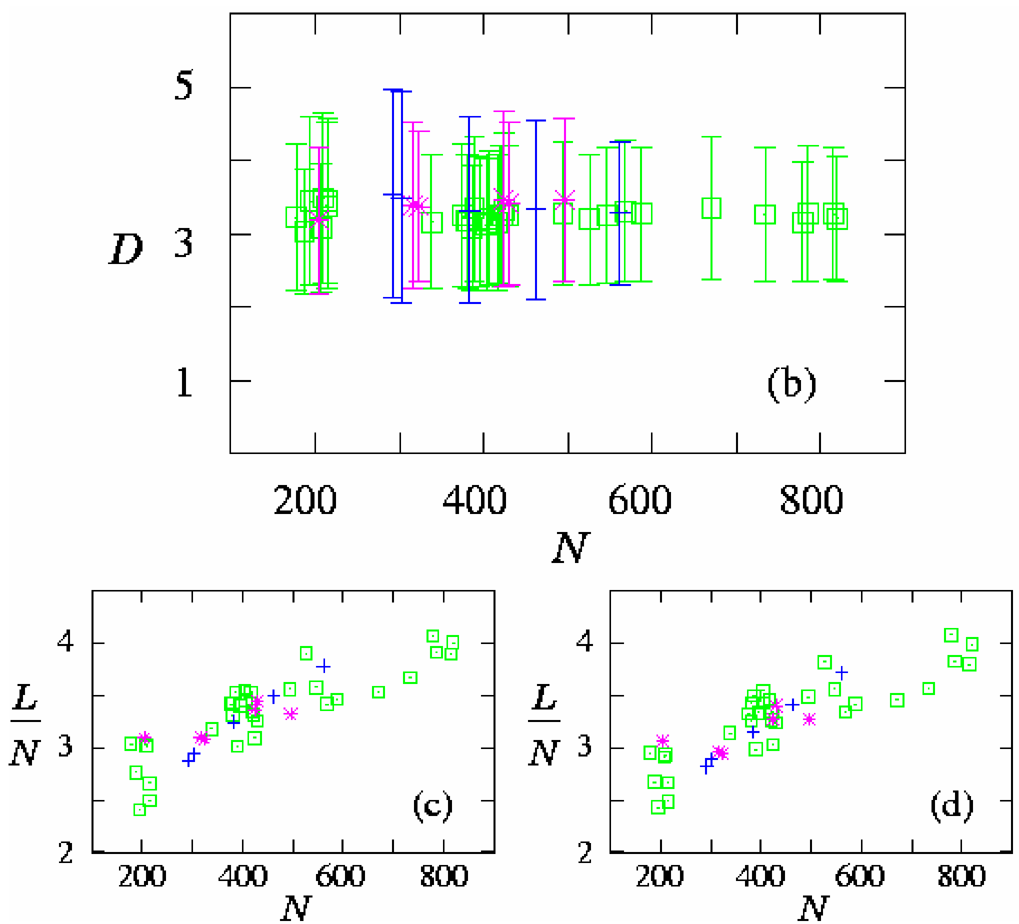
Metabolic pathways

intermediate metabolism
bioenergetics

Non–metabolic pathways

information pathways
electron, transmembrane transport
signal transduction
structure and function of the cell

Properties of metabolic networks



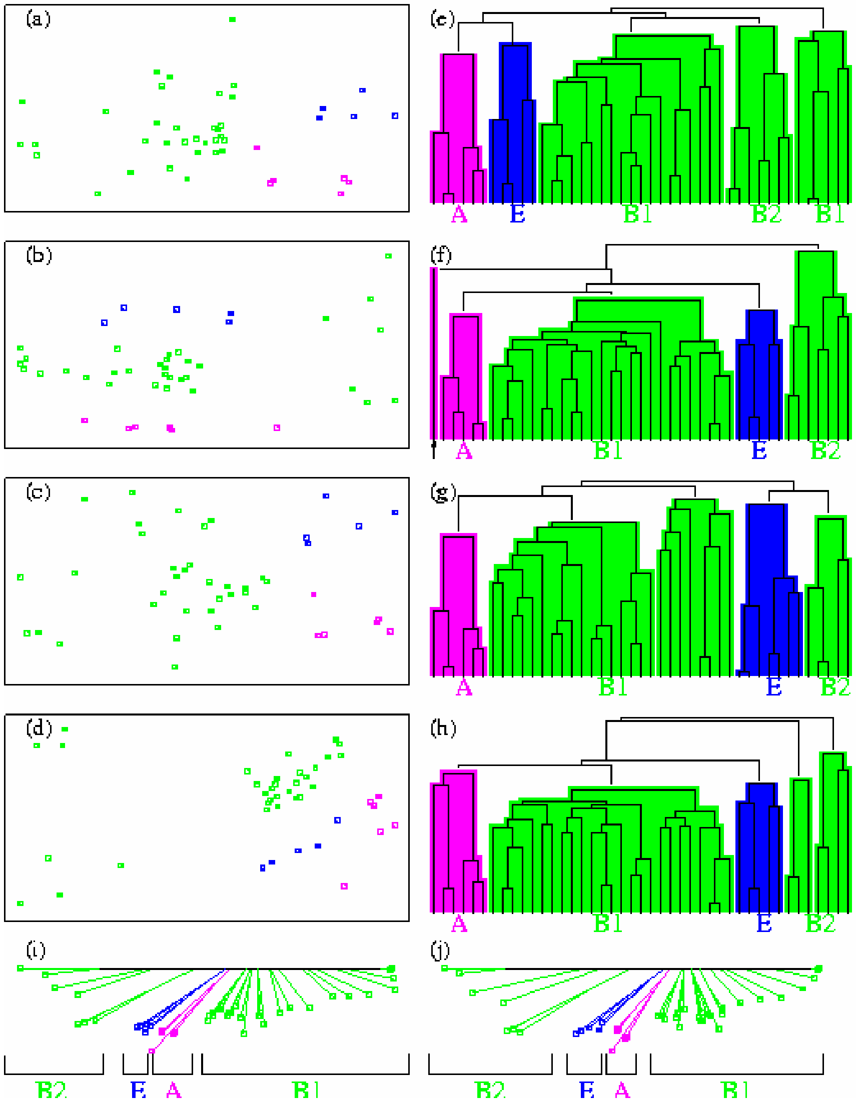
Average distances are independent of organisms!

⇐ by making more links between nodes.

⇐ based on **“design principles”** of the cell through **evolution.**

cf. Other scale-free network: ***D~log(N)***

Taxonomy using networks



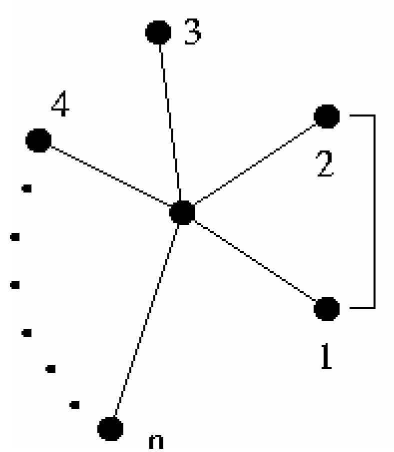
A: Archaea

B: Bacteria

E: Eukaryotes

Watts-Strogatz

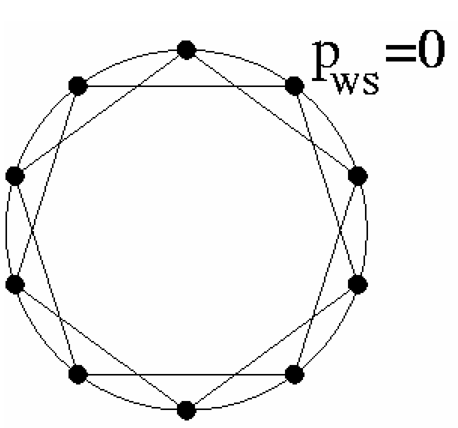
Clustering: My friends will know each other with high probability!

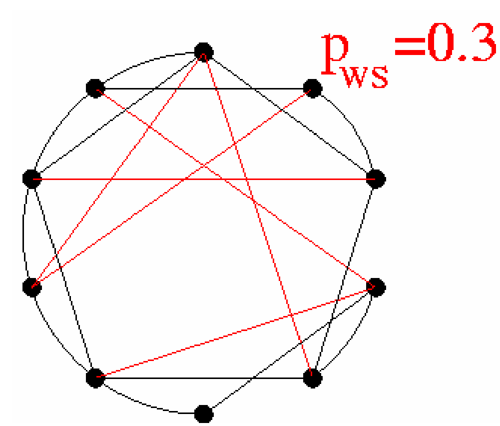


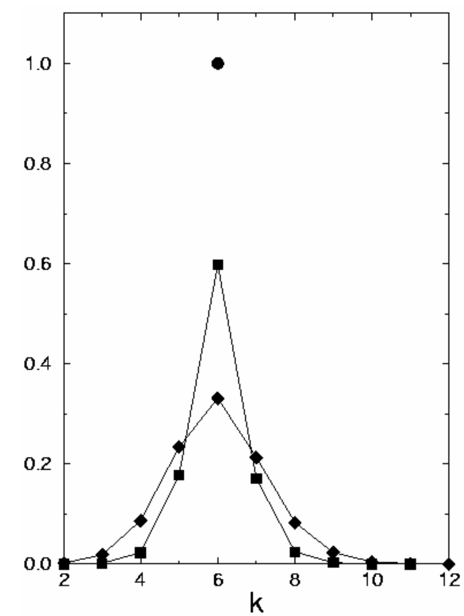
Probability to be connected $C \gg p$

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

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

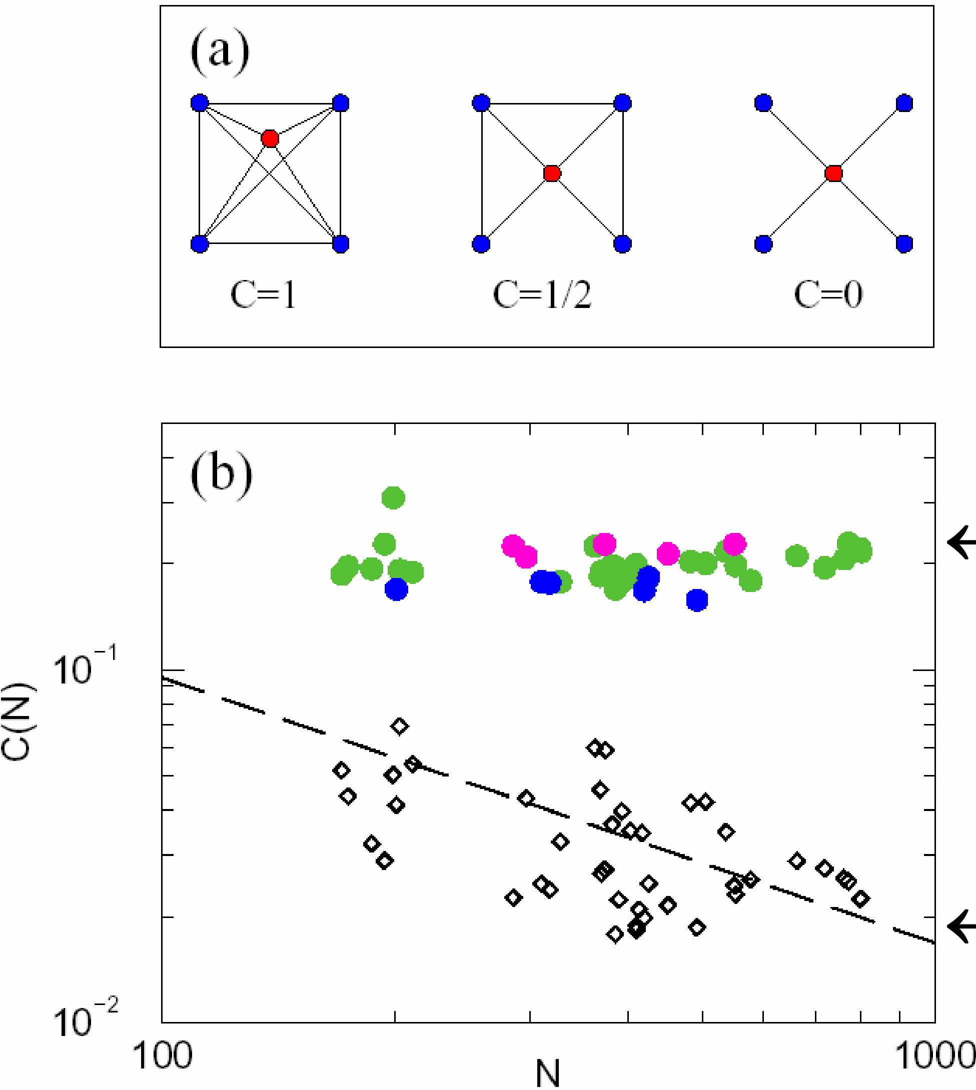






(Nature **393,** 440 (1998))

Modularity in the metabolism

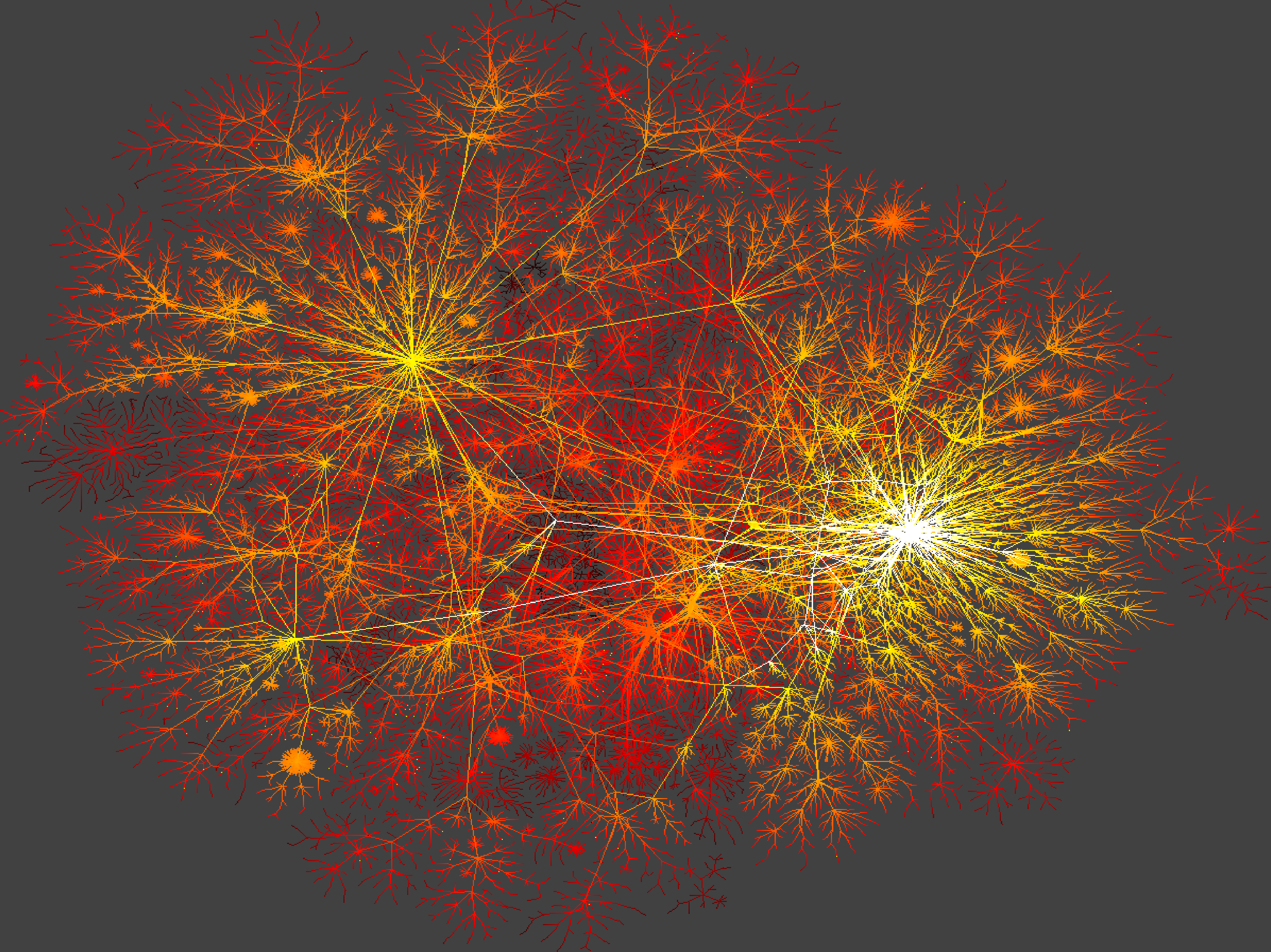


Clustering Coefficient:

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

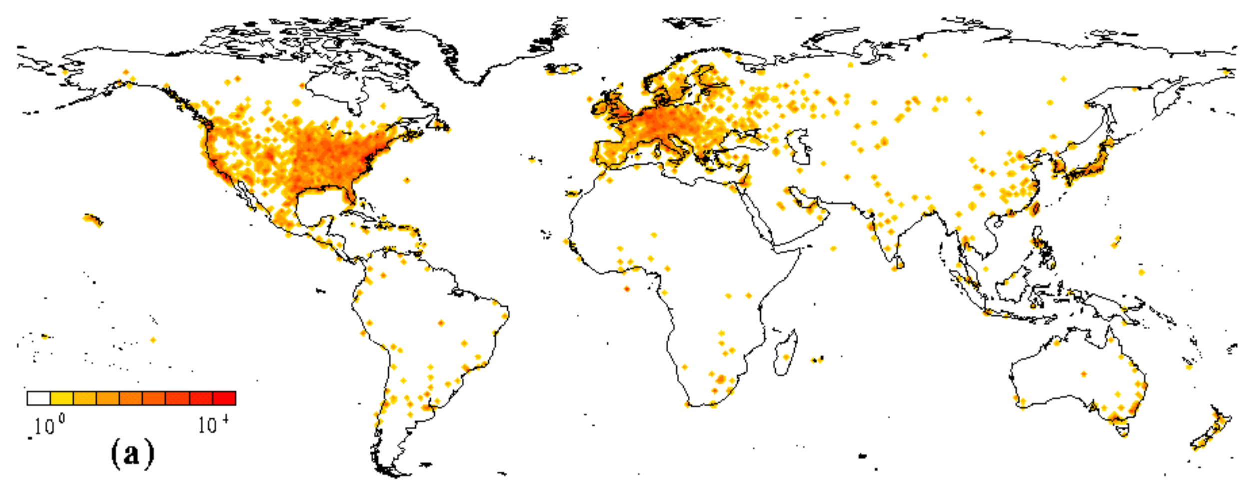
←Metabolic network (43 organisms)

← Scale-free model

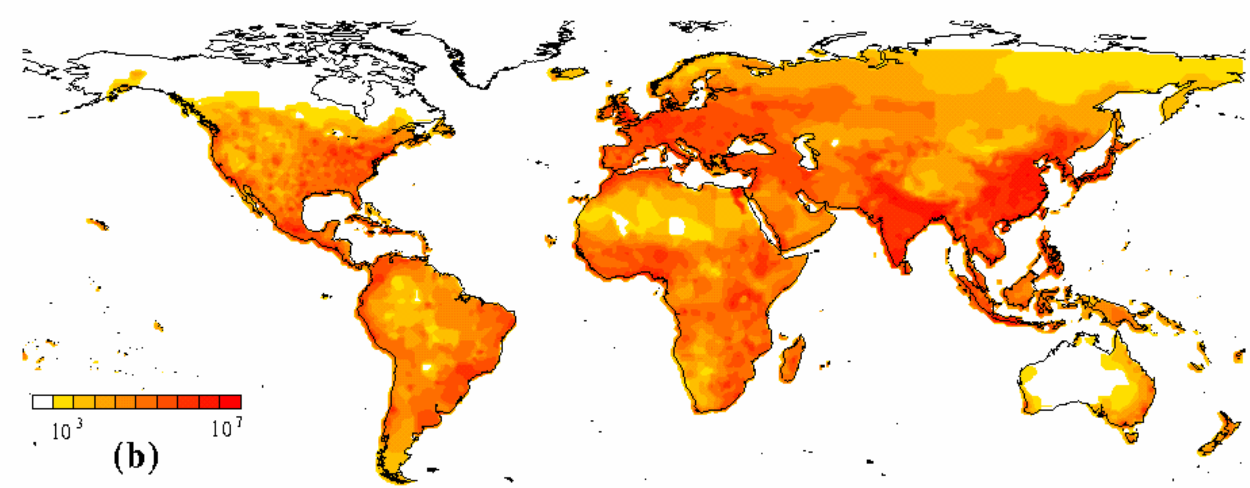


Spatial Distributions

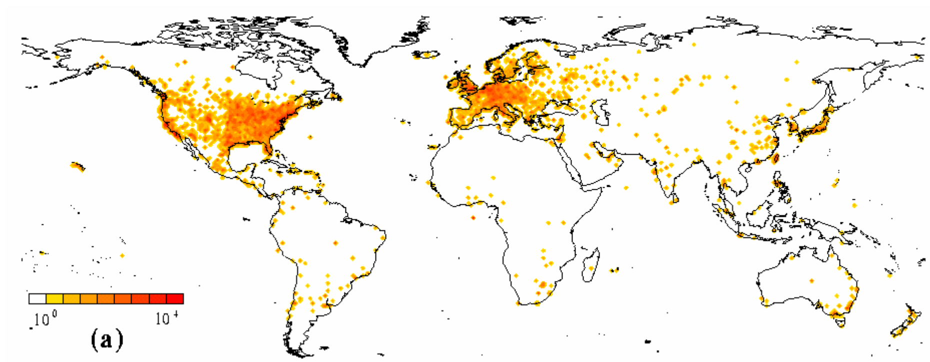
Router density



Population density



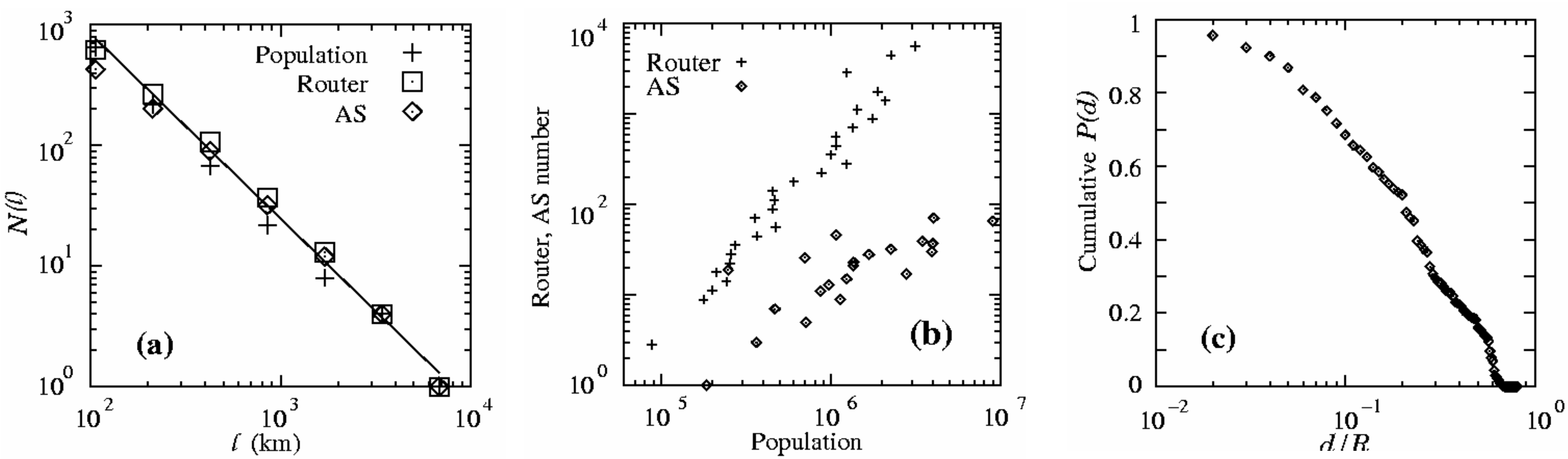
Spatial Distribution of Routers



Fractal set

Box counting: $N(\ell) \equiv$ No. of boxes of size ℓ that contain routers

$N(\ell) \sim \ell^{-D_f}$ $D_f=1.5$

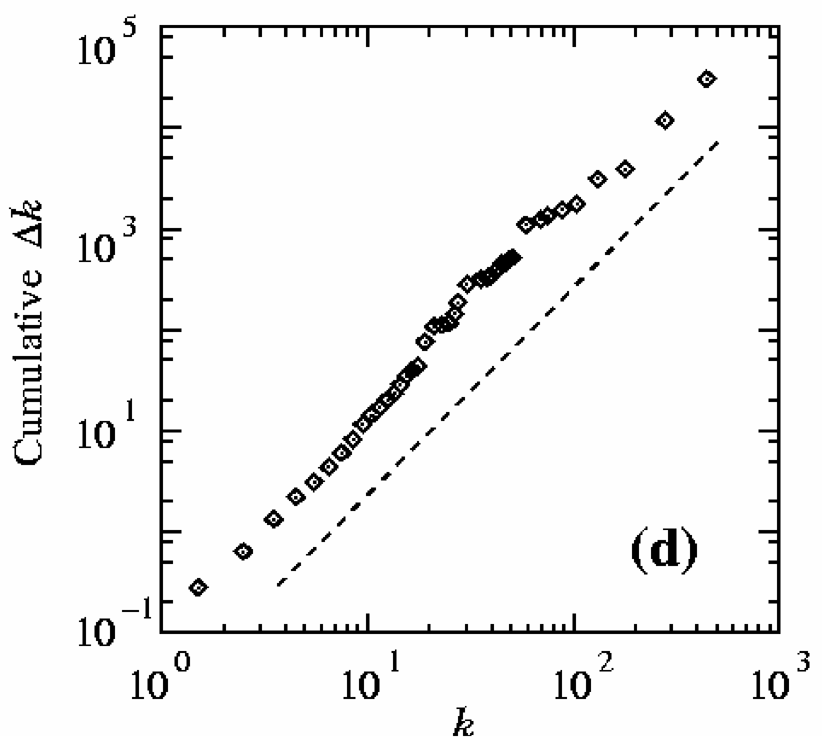


Preferential Attachment

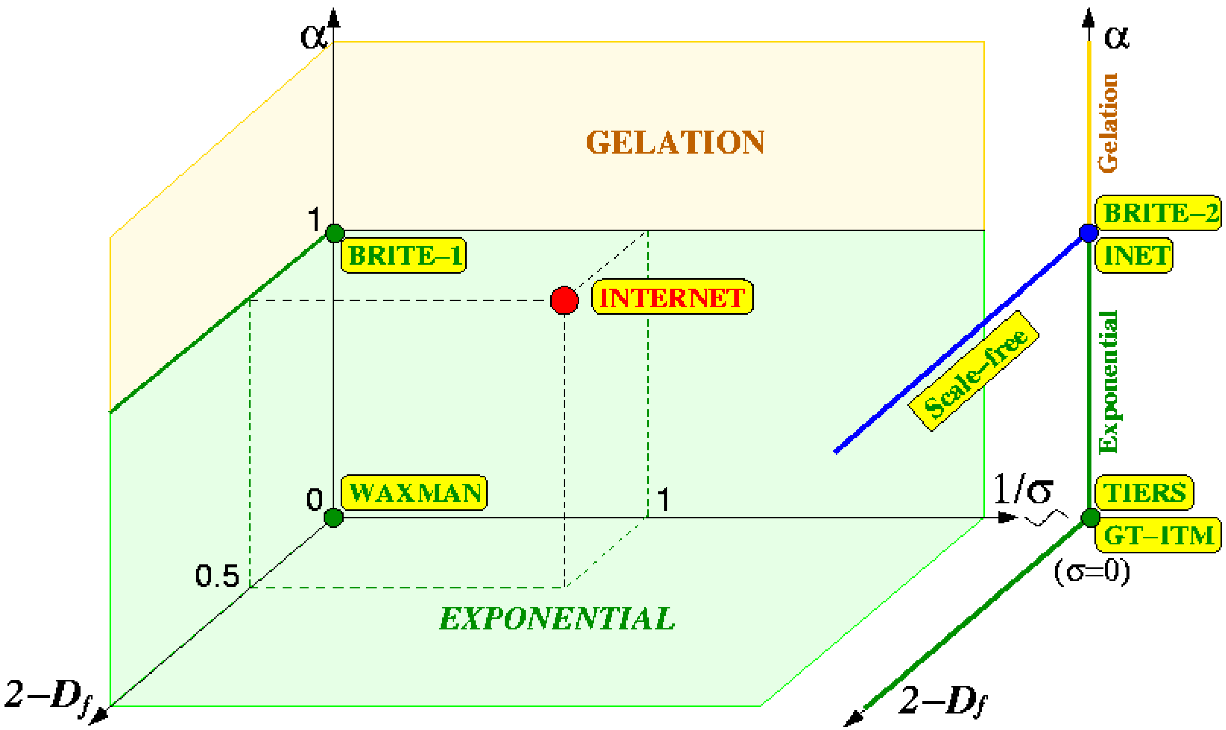
- Compare maps taken at different times (Δt = 6 months)
- Measure $\Delta k(k)$, increase in No. of links for a node with k links

Preferential Attachment:

$$\Delta k(k) \sim k$$



INTERNET



$N(\ell) \sim \ell^{-D_f}$ D_f=1.5

$\Delta k(k) \sim k^{\alpha}$ α=1

$P(d) \sim d^{-\sigma}$ σ=1

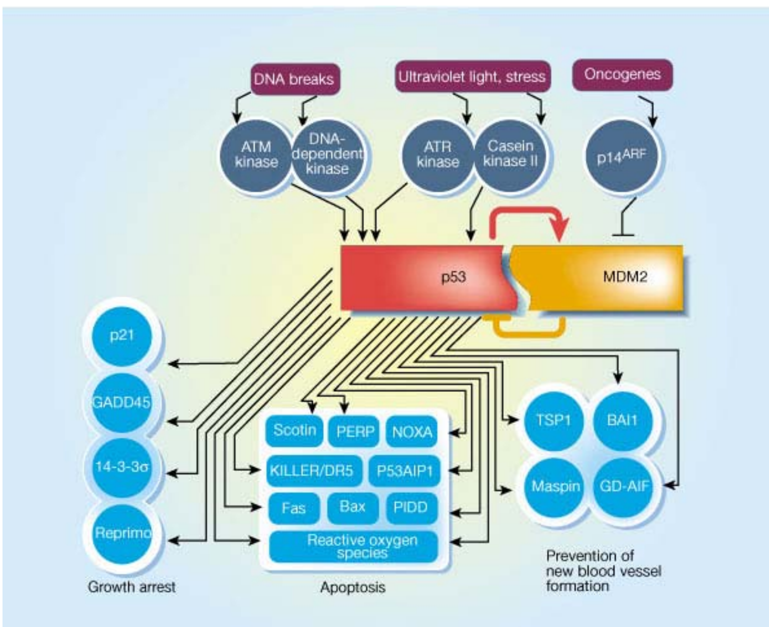
Nature **408** 307 (2000)

news and views feature

Surfing the p53 network

Bert Vogelstein, David Lane and Arnold J. Levine

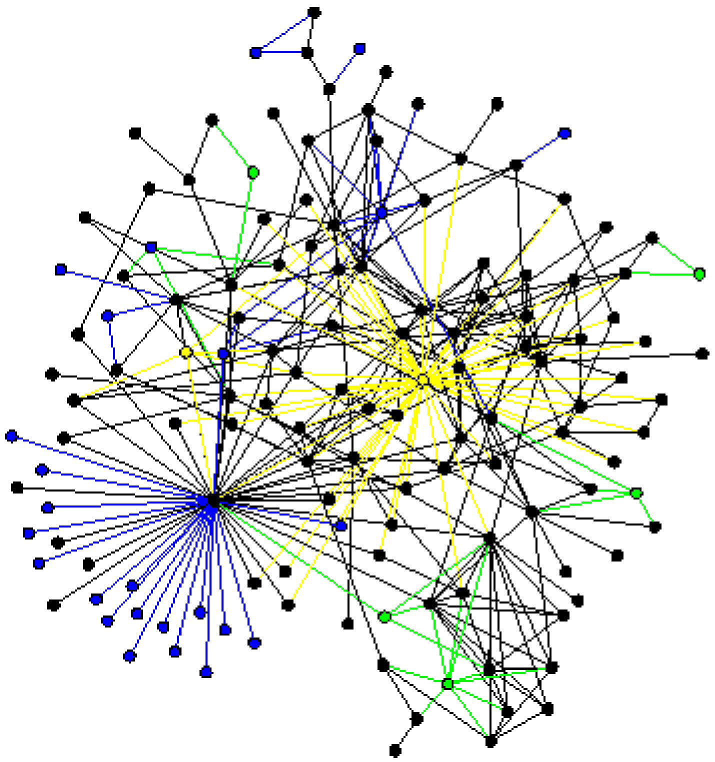
The p53 tumour-suppressor gene integrates numerous signals that control cell life and death. As when a highly connected node in the Internet breaks down, the disruption of p53 has severe consequences.

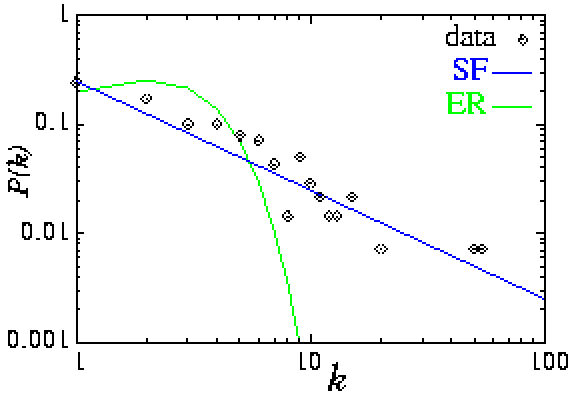


…

“One way to understand the p53 network is to compare it to the Internet.
The cell, like the Internet, appears to be a ‘**scale-free network**’.”

p53 network (mammals)

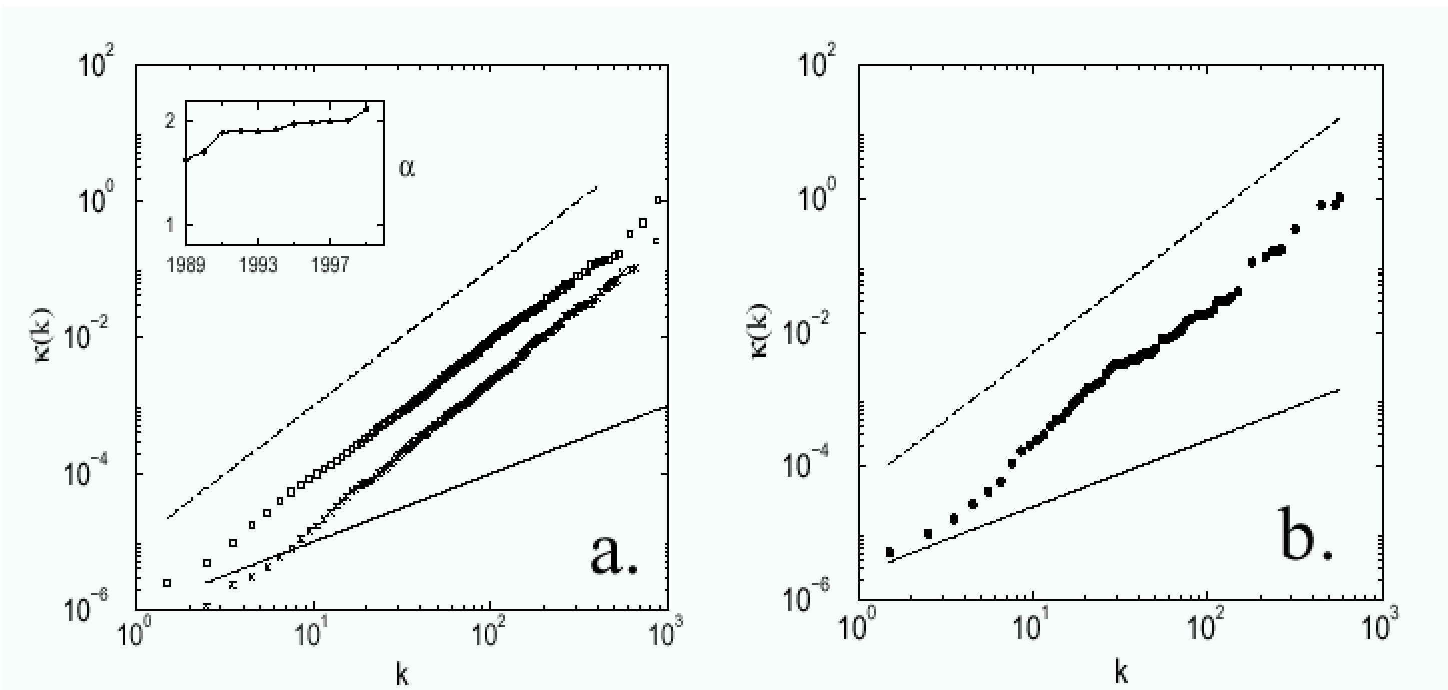




Preferential Attachment

$$\frac{\partial k_i}{\partial t} \propto \Pi(k_i) \sim \frac{\Delta k_i}{\Delta t} \qquad \text{For given } \Delta t, \therefore \Delta k \propto \Pi(k)$$

k vs. Δk : increase in the No. of links in a unit time

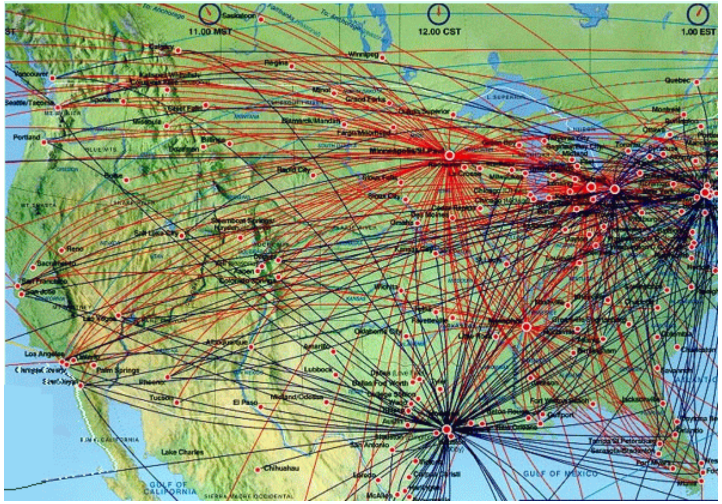


Citation network

Internet

(cond-mat/0104131)

What is the topology of cellular networks?



Argument 1:
Cellular networks are scale-free!

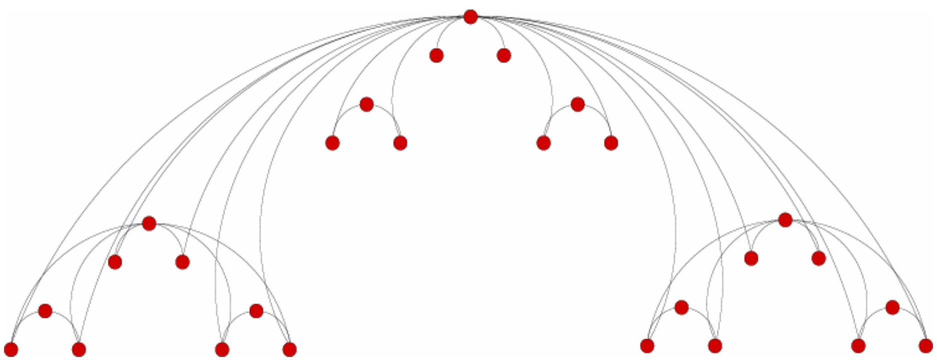
Reason:
They formed one node at a time…

Argument 2:
Cellular networks are exponential!

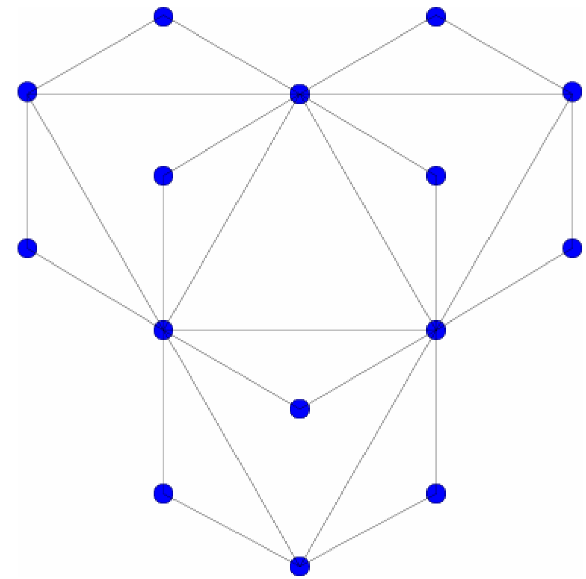
Reason:
They have been streamlined by evolution...

Combining Modularity and the Scale-free Property

Deterministic Scale-Free Networks



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

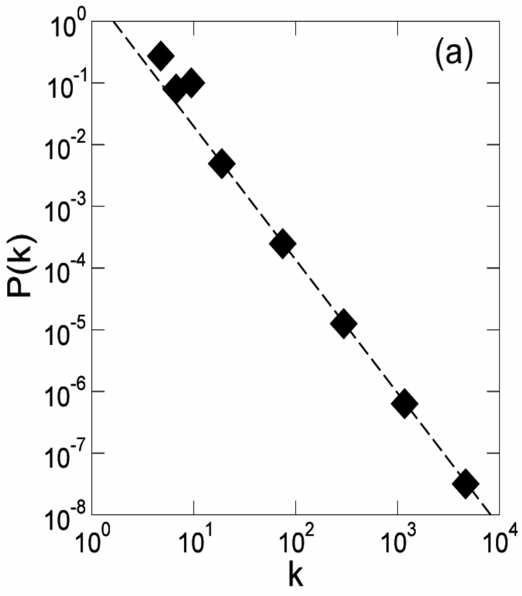


Dorogovtsev, S. N., Goltsev, A. V., & Mendes, J. F. F. (2001) cond-mat/0112143. (DGM)

Properties of hierarchical networks

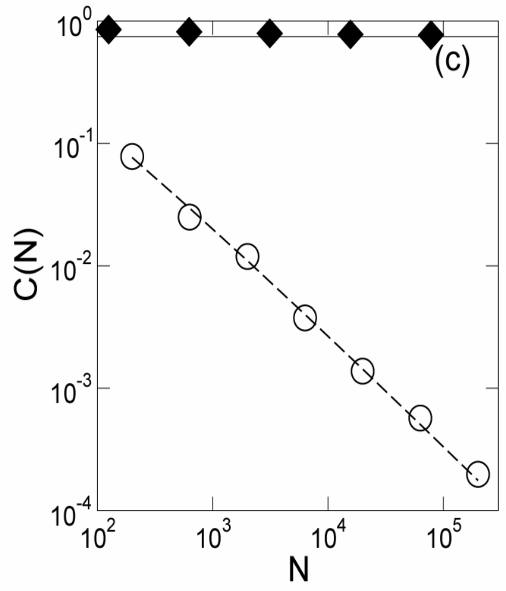
1. Scale-free

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



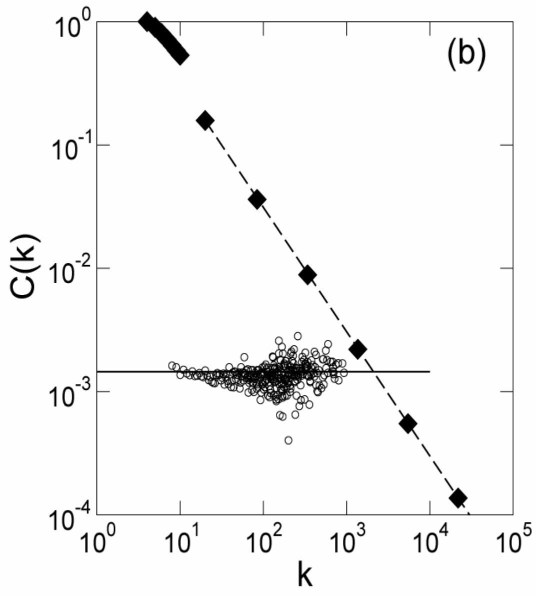
2. Clustering coefficient independent of N

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

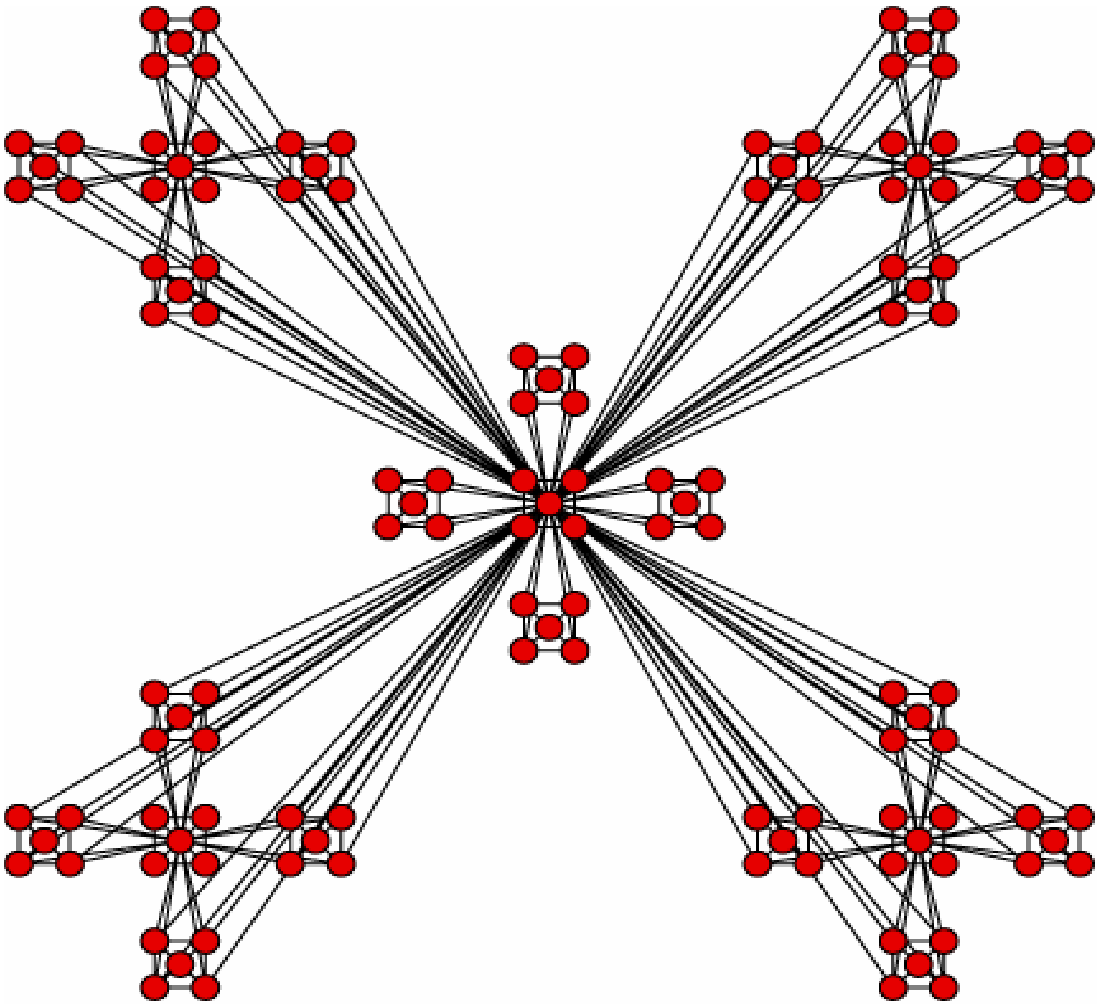


3. Scaling clustering coefficient (DGM)

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



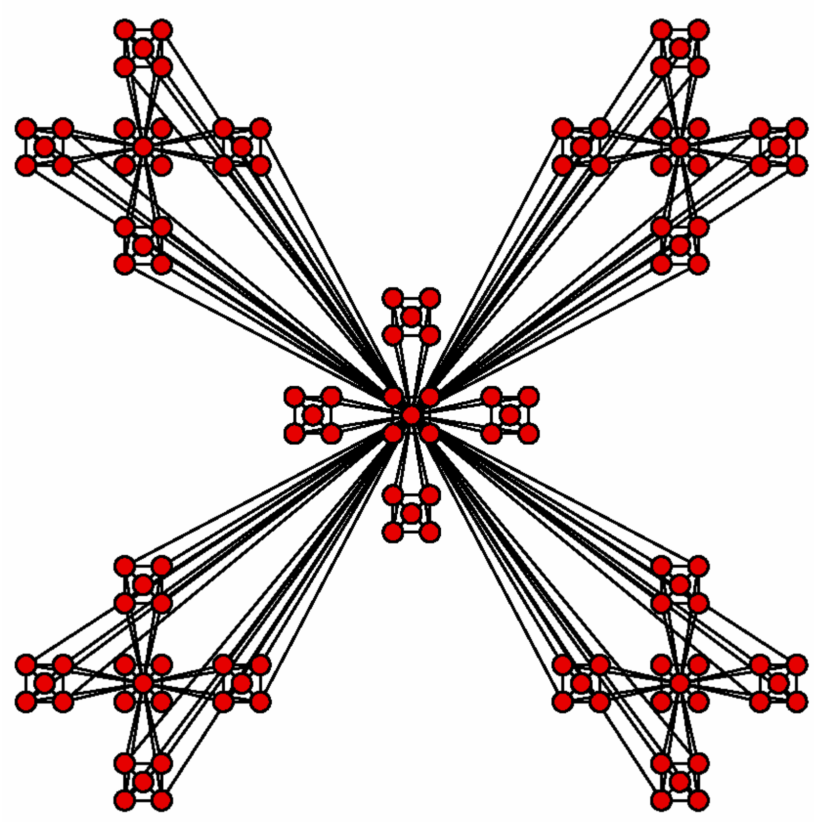
Hierarchical Networks



Real Networks Have a Hierarchical Topology

What does it mean?

Many highly connected small clusters
combine into
few larger but less connected clusters
combine into
even larger and even less connected clusters

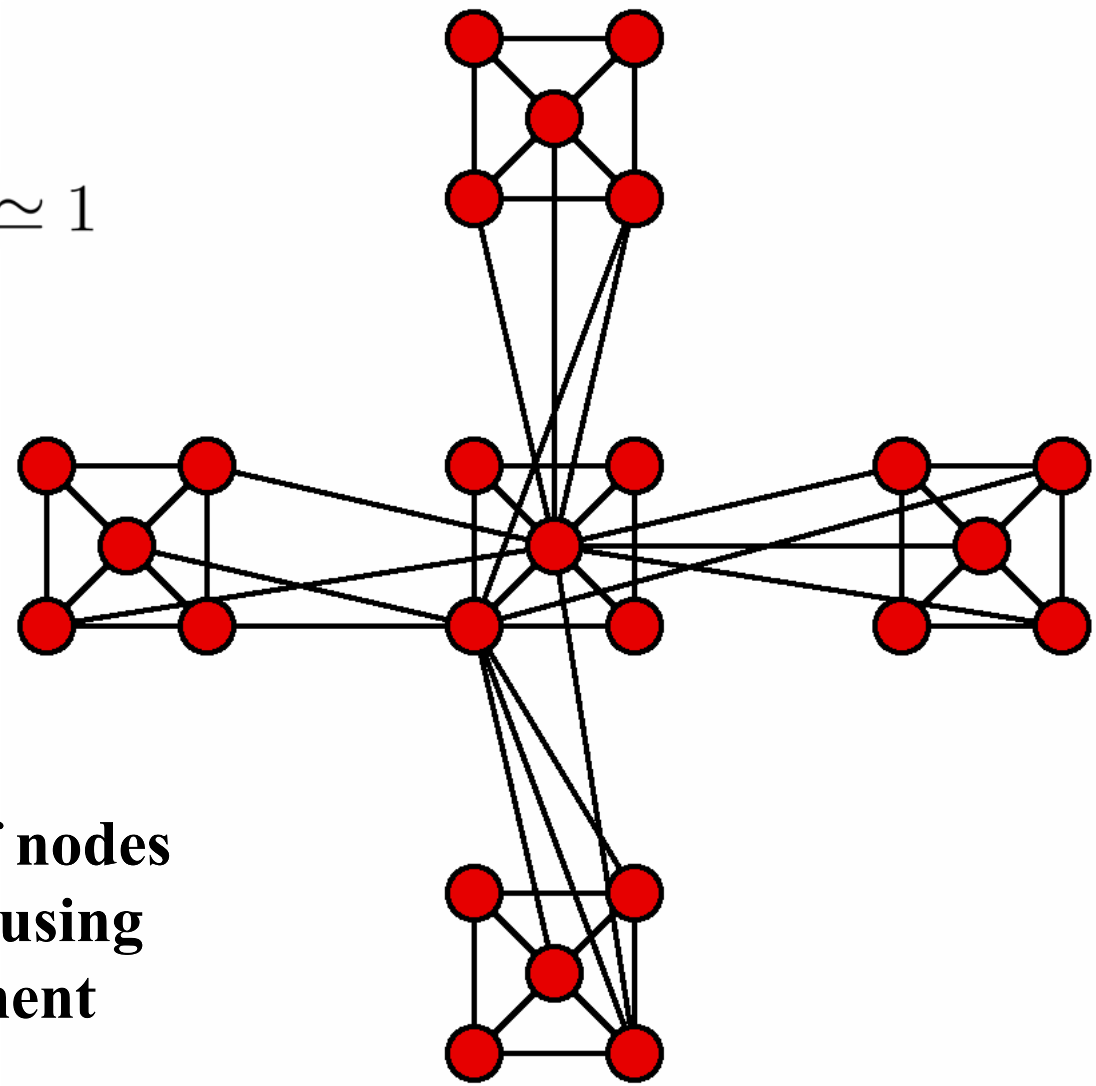


- The degree of clustering follows:

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

Is the hierarchical exponent β universal?

- $C(k) \sim k^{-\beta}$
- For most systems: $\beta \simeq 1$

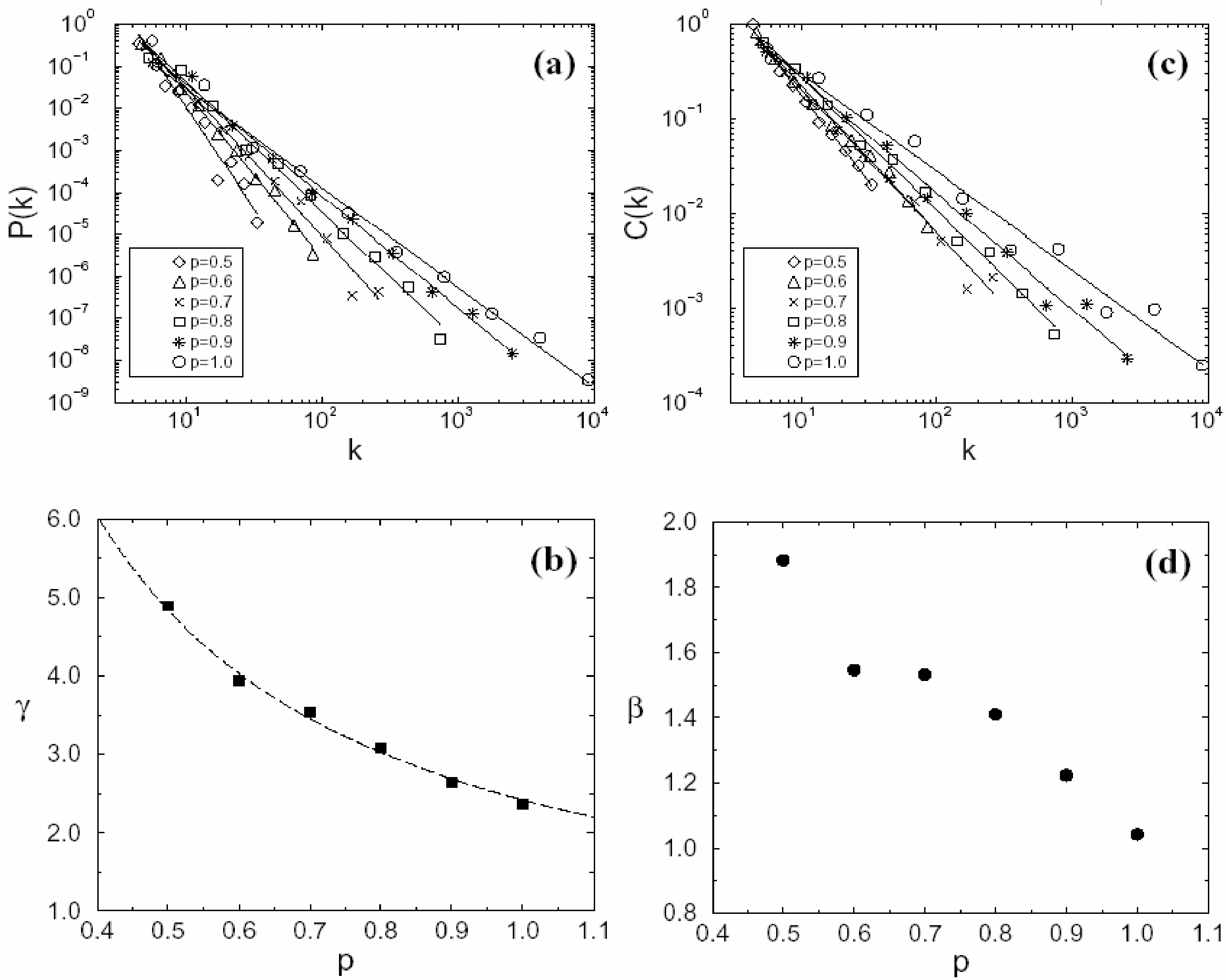


Connect a *p* fraction of nodes to the central module using preferential attachment

Stochastic Hierarchical Model

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

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



Is hierarchy present in network models?

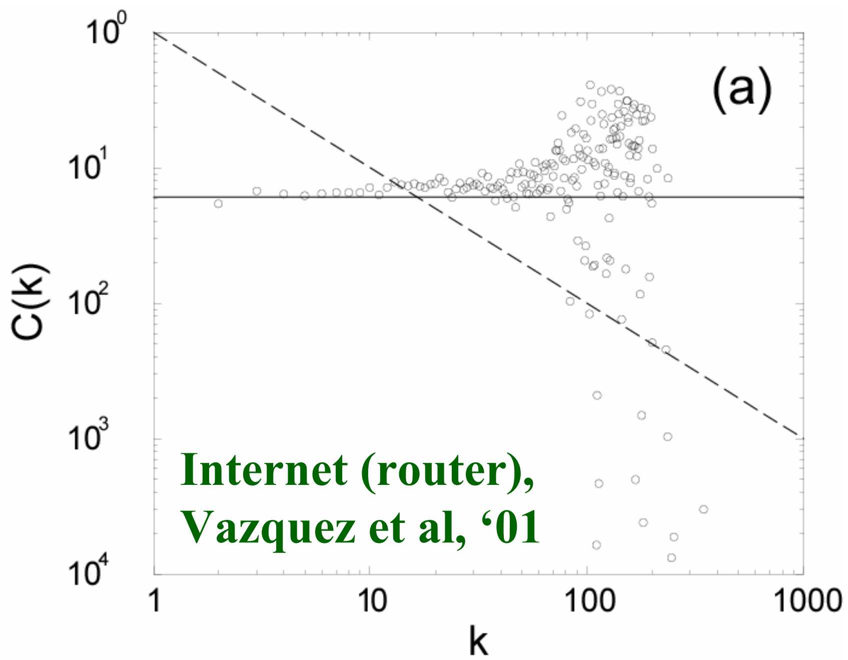
NO:

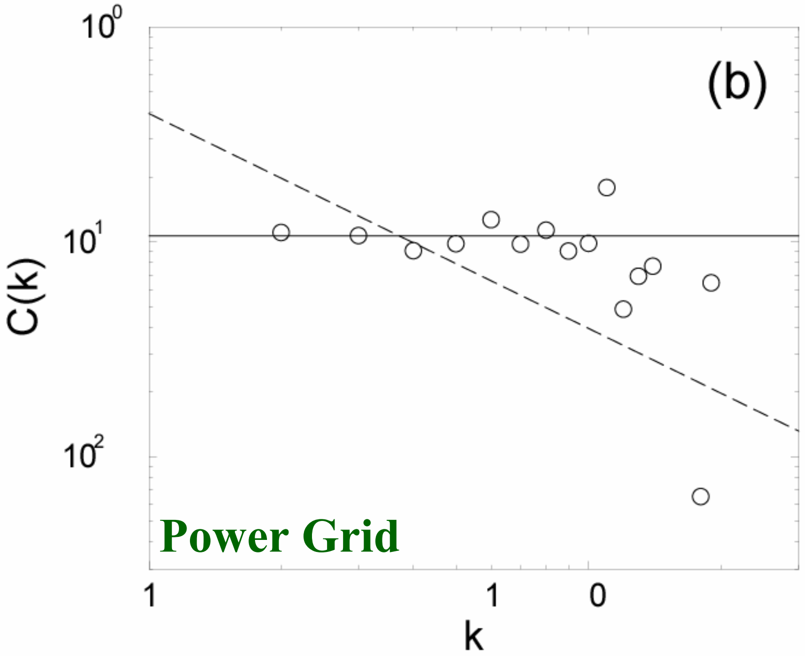
-Scale-free model (alb& Albert,1999)
-Erdos-Renyi model (1959)
-Watts-Strogatz (1998)

YES:

Dorogovtsev, Goltsev, Mendes, 2001 (determ.)
-Klemm and Eguiluz, 2002
-Vasquez, Pastor-Satorras,Vespignani (2001)*
$\Rightarrow$ Bianconi & alb (fitnesss model) (2001)

Exceptions:
Geographically Organized Networks:





Common feature:

economic pressures towards shorter links

Traditional modeling:
Network as a static graph

Given a network with N nodes and L links

⇓

Create a graph with statistically identical topology

RESULT: model the static network topology

PROBLEM: **Real networks are dynamical systems!**

Evolving networks

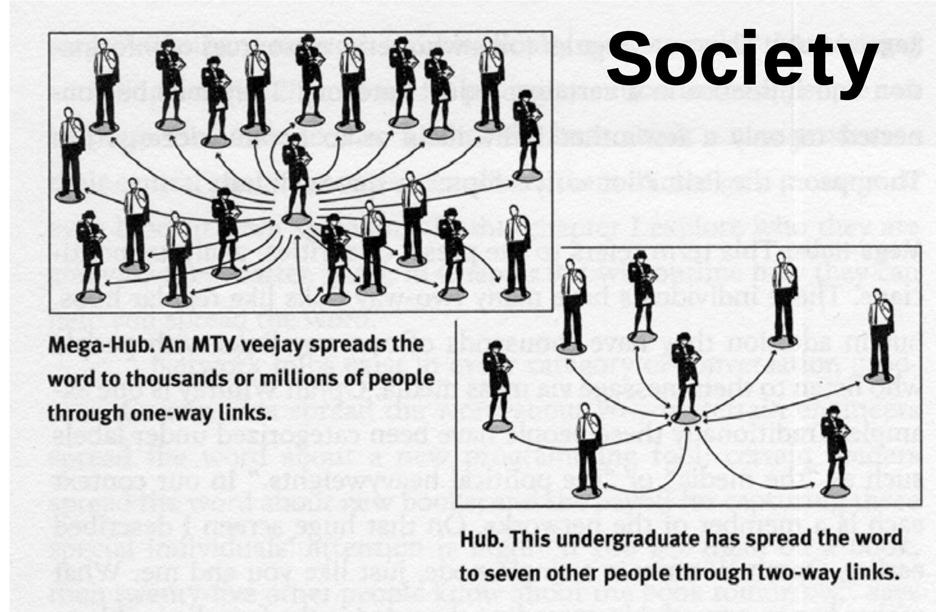
OBJECTIVE: capture the network dynamics

METHOD :
- identify the processes that contribute to the network topology
- develop dynamical models that capture these processes

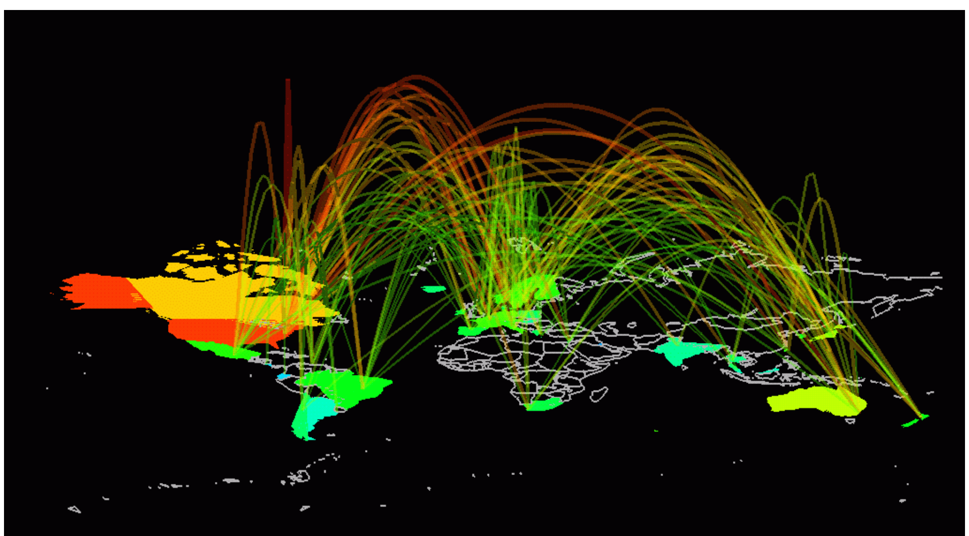
⇓

BONUS: **get the topology correctly**.

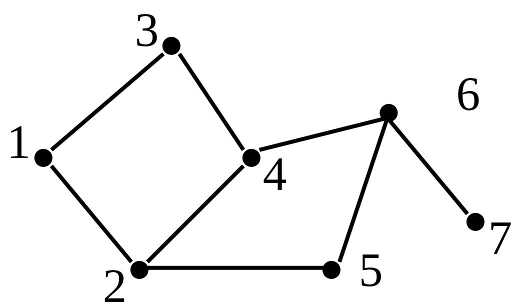
Society



Internet



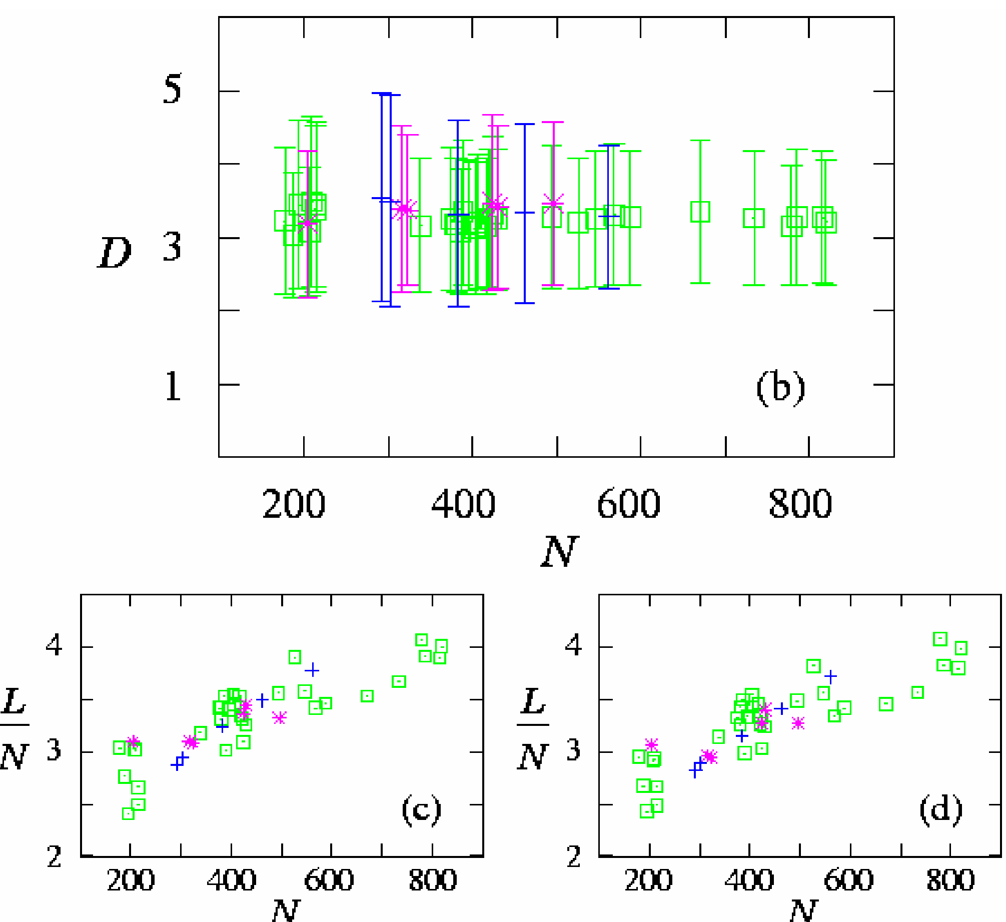
Node-node distance in metabolic networks



$D_{15}=2$ [1→2→5]

$D_{17}=4$ [1→3→4 → 6→7]

… $\boldsymbol{D = ??}$



Scale-free networks:

D~log(N)

Larger organisms are expected to have a larger diameter!

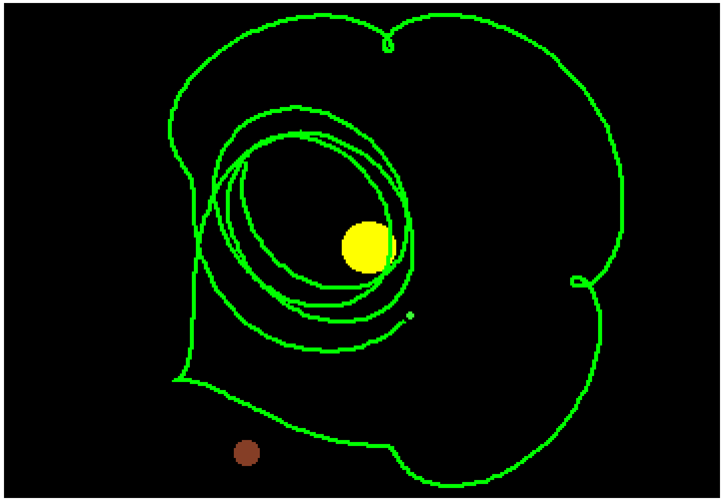
What is Complexity?

A popular paradigm: Simple systems display complex behavior

- non-linear systems
- chaos
- fractals

3 Body Problem

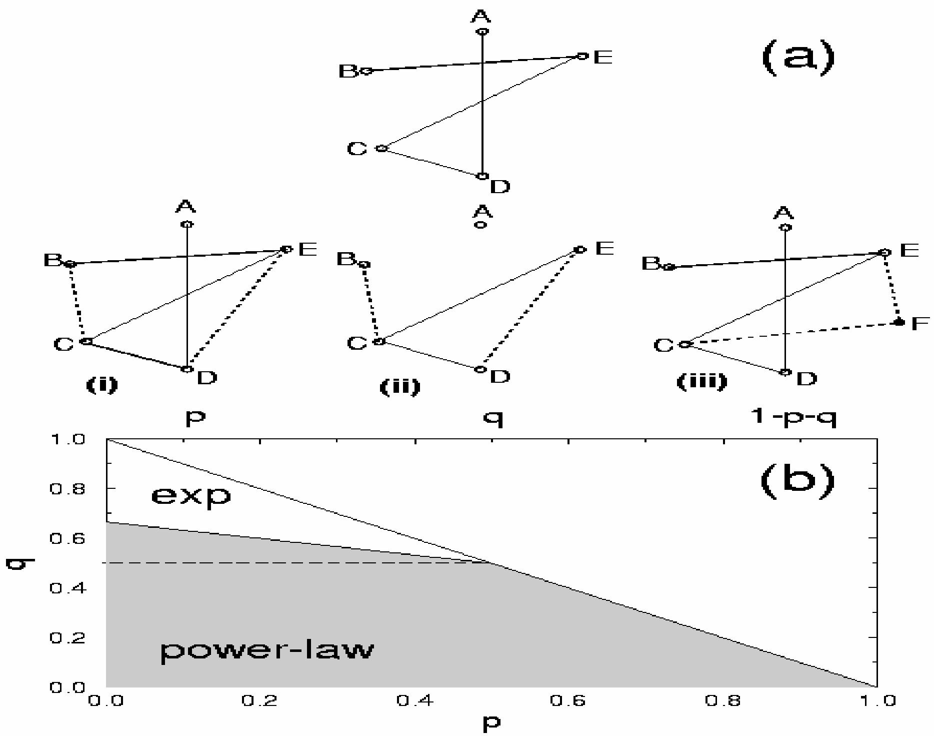
Earth( ) Jupiter ( ) Sun ( )

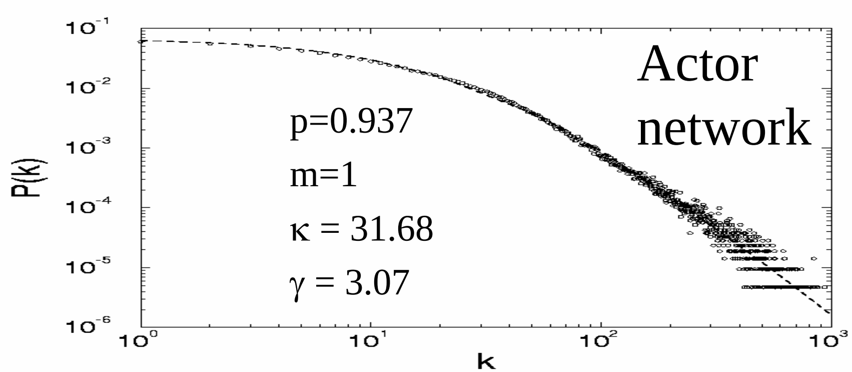


Main Entry: [1]**com·plex**
Function: *noun*
Etymology: Late Latin *complexus* totality, from Latin, embrace, from *complecti*
Date: 1643
1 : a whole made up of complicated or interrelated parts

Universality?

WWW (in)	Internet	Actor	Citation index	Sex Web	Cellular network	Phone call network	linguistics
$\gamma = 2.1$	$\gamma = 2.5$	$\gamma = 2.3$	$\gamma = 3$	$\gamma = 3.5$	$\gamma = 2.1$	$\gamma = 2.1$	$\gamma = 2.8$





Extended Model

- prob. p : internal links
- prob. q : link deletion
- prob. 1-p-q : add node

$$P(k) \sim (k+\kappa(p,q,m))^{-\gamma(p,q,m)}$$

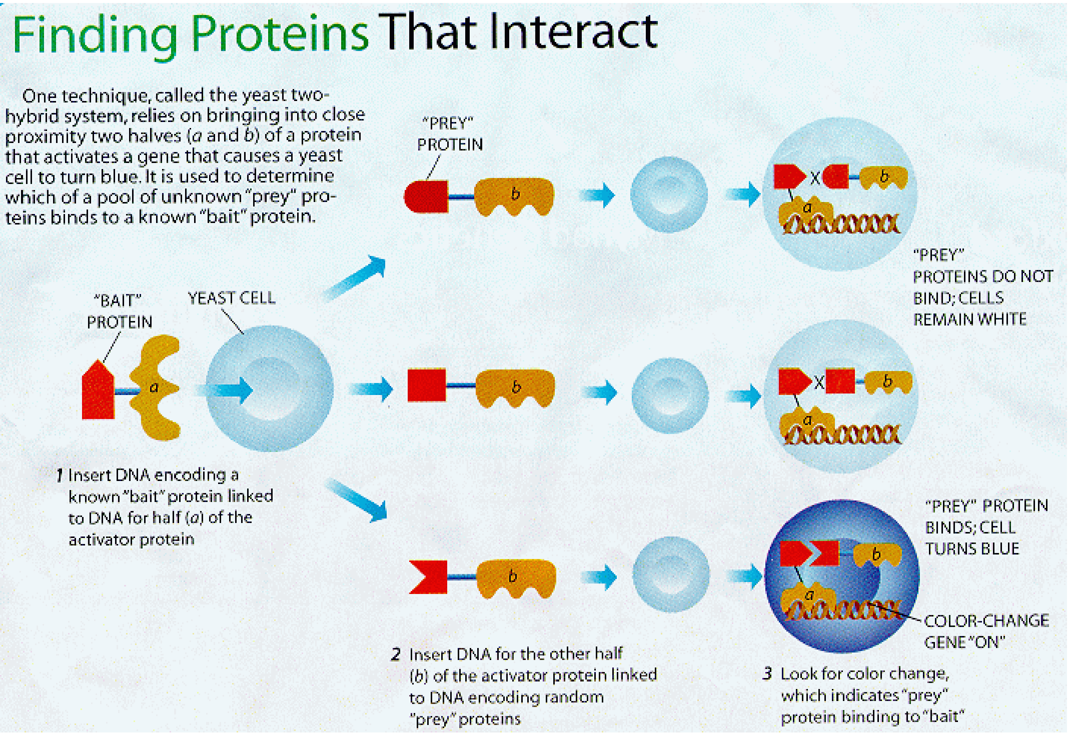
$$\gamma \in [1,\infty)$$

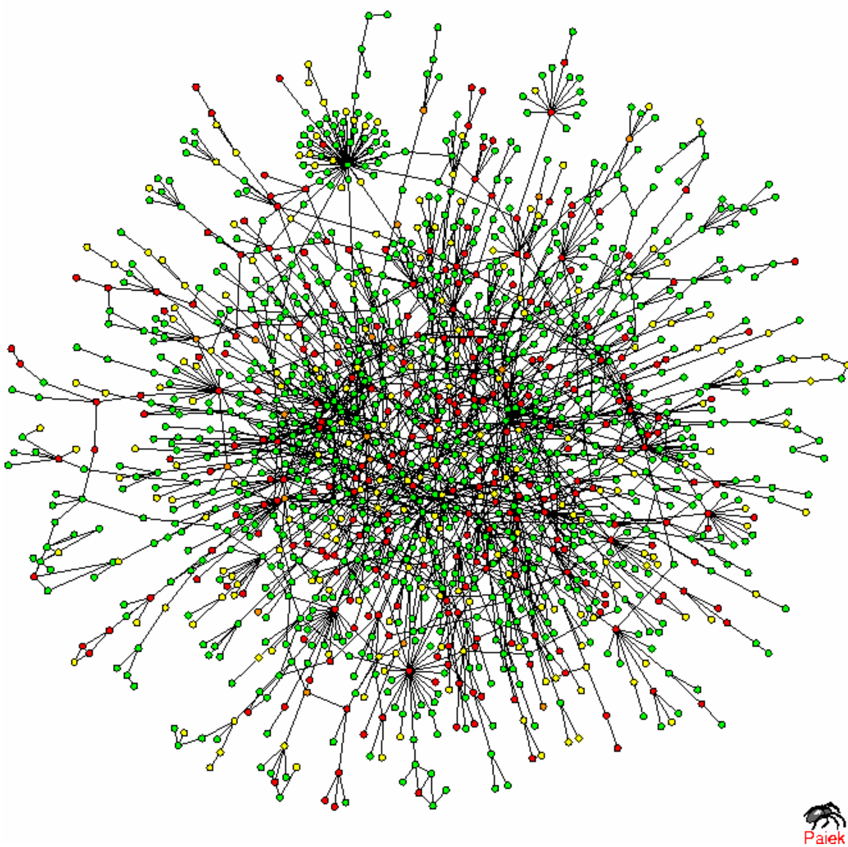
- Predict the network topology from microscopic processes with parameters (p,q,m)
- Scaling but no universality

Yeast protein network

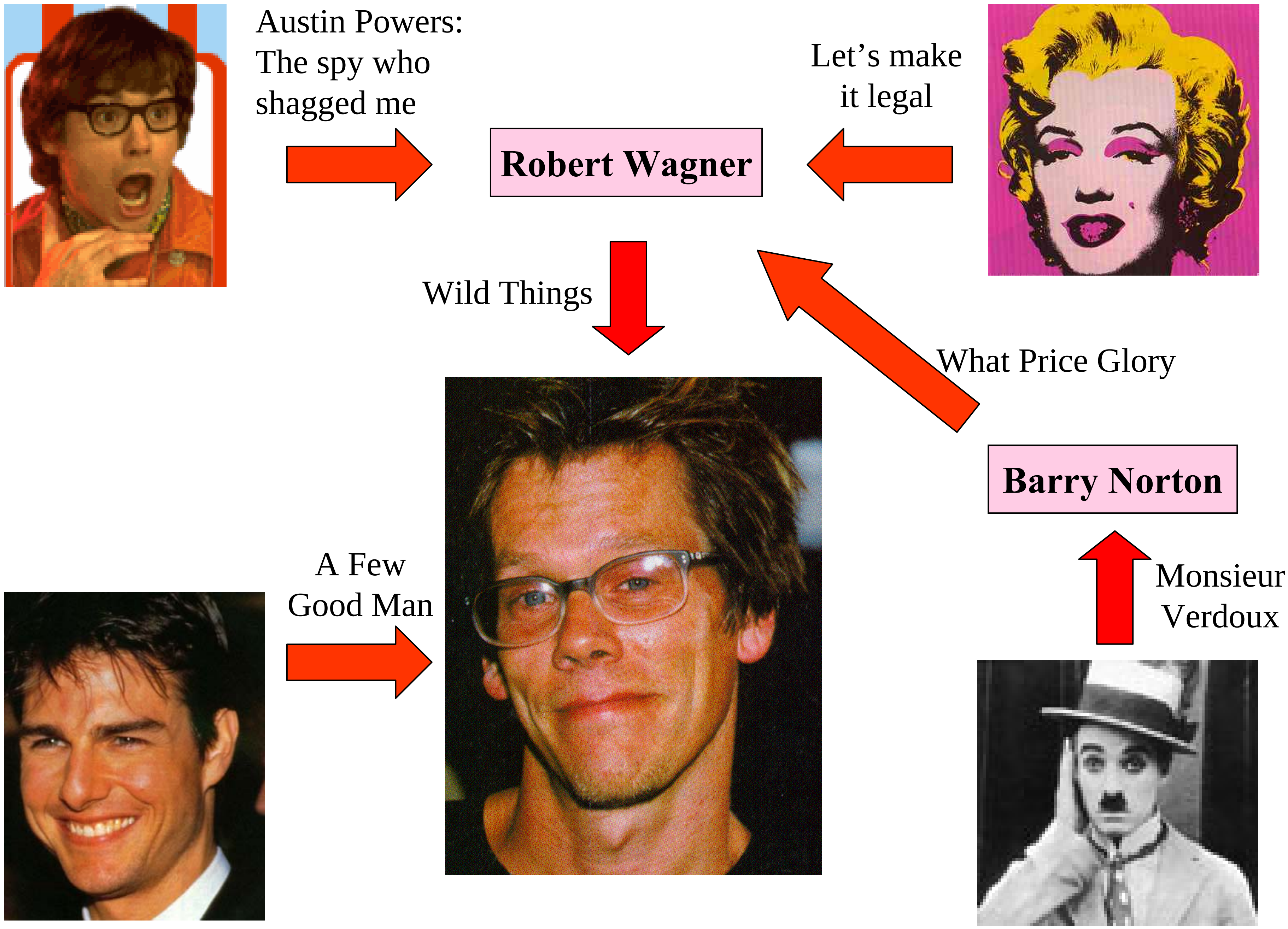
Nodes: proteins

Links: physical interactions (binding)





P. Uetz, *et al.* *Nature* **403**, 623-7 (2000).

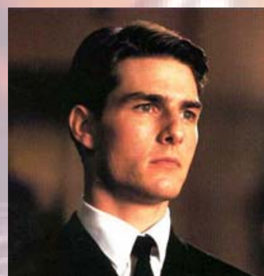

Austin Powers:
The spy who
shagged me
Robert Wagner
Let's make
it legal
Wild Things
What Price Glory
Barry Norton
A Few
Good Man
Monsieur
Verdoux

ARE COMPLEX NETWORKS REALLY RANDOM?

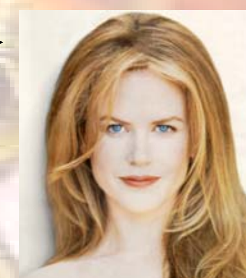
NO!

ACTOR CONNECTIVITIES

Nodes: actors
Links: cast jointly



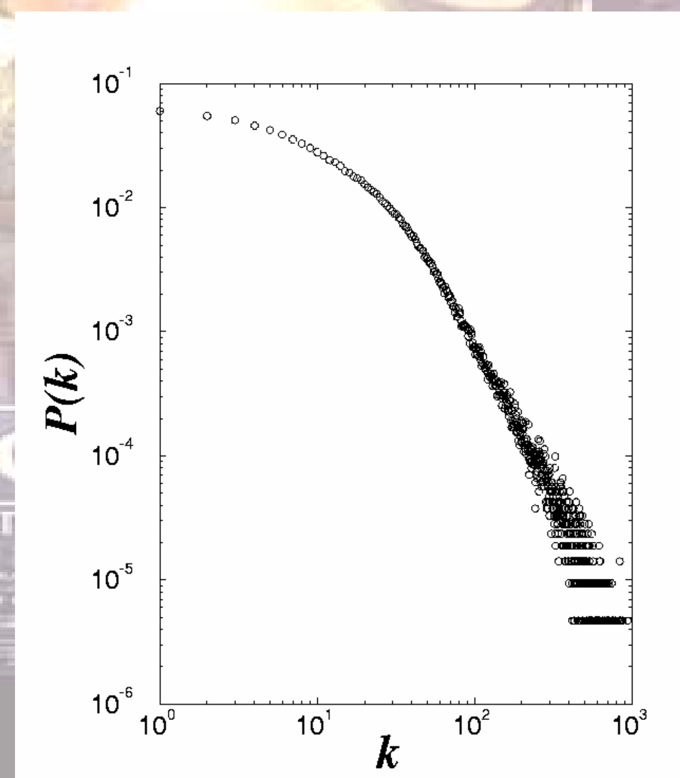
Days of Thunder (1990)
Far and Away (1992)
Eyes Wide Shut (1999)



N = 212,250 actors
$\langle k \rangle = 28.78$

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

$\gamma=2.3$



Society

Nodes: individuals

Links: social relationship (family/work/friendship/etc.)

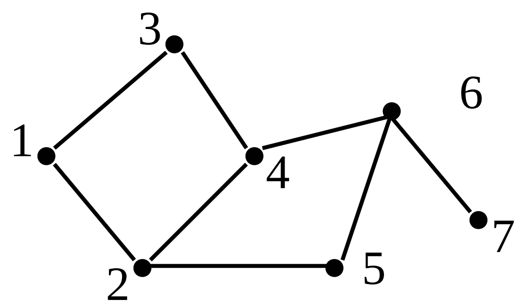
S. Milgram (1967)

John Guare, **Six Degrees of Separation**

1929, Frigyes Karinthy

“we could name any person among earth’s one and a half billion inhabitants and through at *most five* acquaintances, one of which he knew personally, he could link to the chosen one”

19 degrees of separation



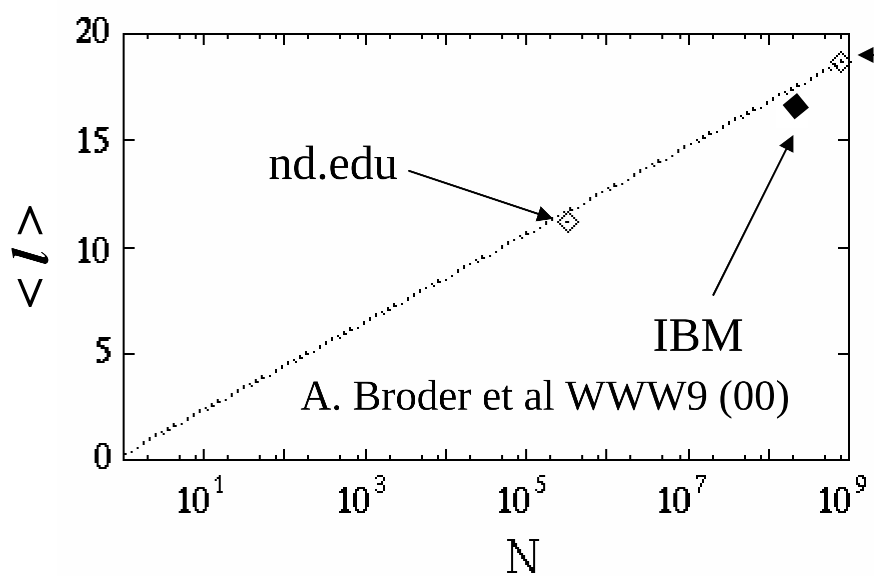
$l_{15}=2$ [1→2→5]

$l_{17}=4$ [1→3→4→6 → 7]

… $<l> = ??$

- **Finite size scaling:** create a network with N nodes with $P_{in}(k)$ and $P_{out}(k)$

$$<l> = 0.35 + 2.06 \log(N)$$



19 degrees of separation
R. Albert et al Nature (99)

based on 800 million webpages
[S. Lawrence et al Nature (99)]

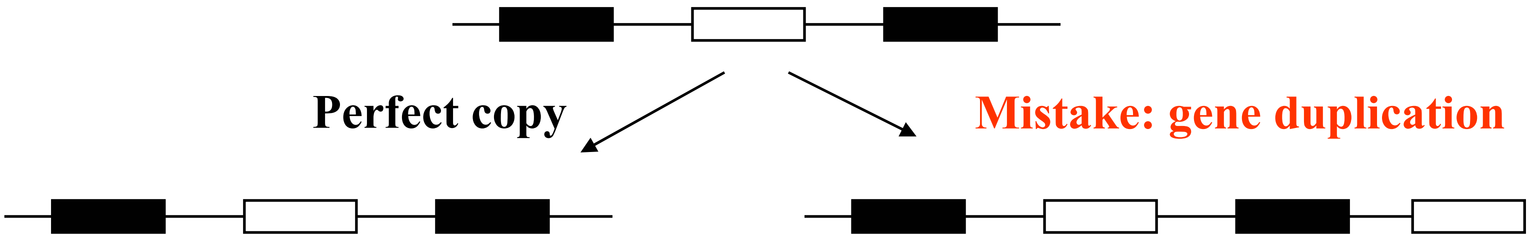
What is Complexity?

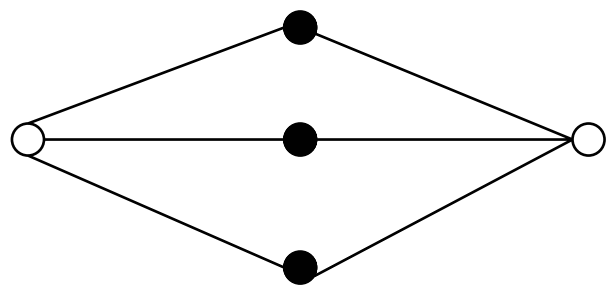
A popular paradigm:

Simple systems display complex behavior

Main Entry: [1]**com·plex**
Function: *noun*
Etymology: Late Latin *complexus* totality, from Latin, embrace, from *complecti*
Date: 1643
1 : a whole made up of complicated or interrelated parts

Origin of the scale-free topology: Gene Duplication





Proteins with more interactions are more likely to get a new link:
$\Pi(k) \sim k$
(preferential attachment).

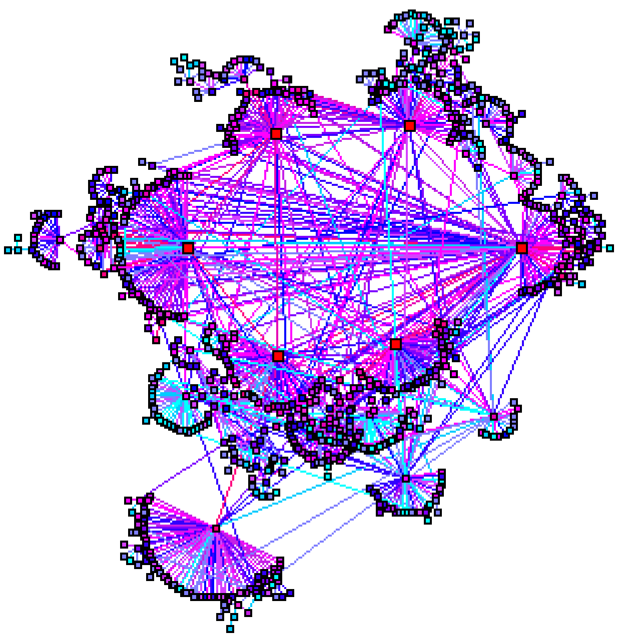
Wagner (2001); Vazquez *et al.* 2003; Sole *et al.* 2001; Rzhetsky & Gomez (2001); Qian et al. (2001); Bhan *et al.* (2002).

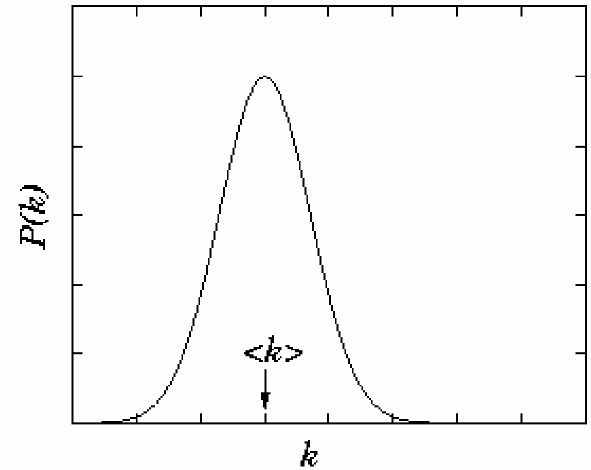
World Wide Web

Nodes: WWW documents
Links: URL links

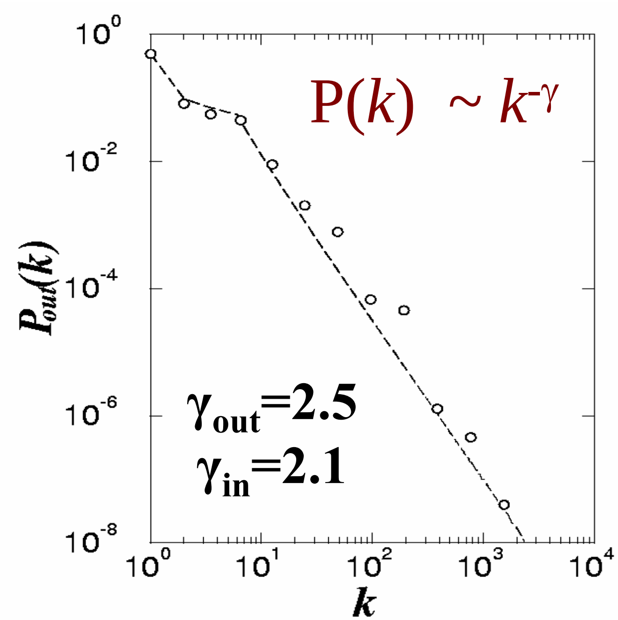
Over 3 billion documents

ROBOT: `collects all URL's found in a document and follows them recursively`





Expected

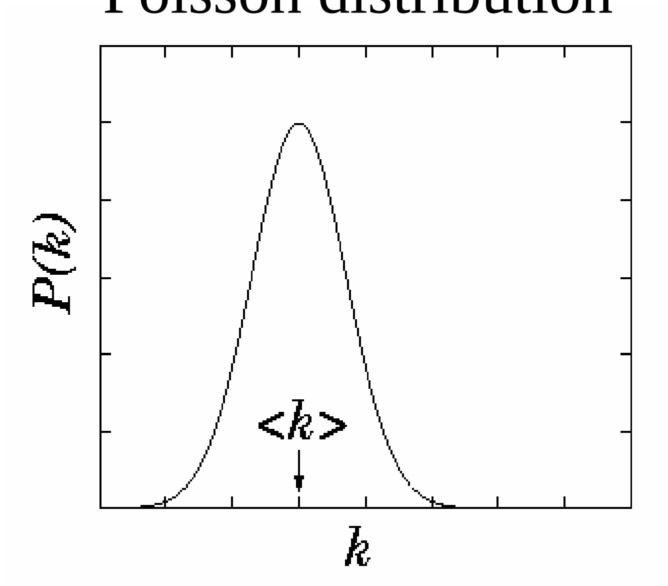


Found

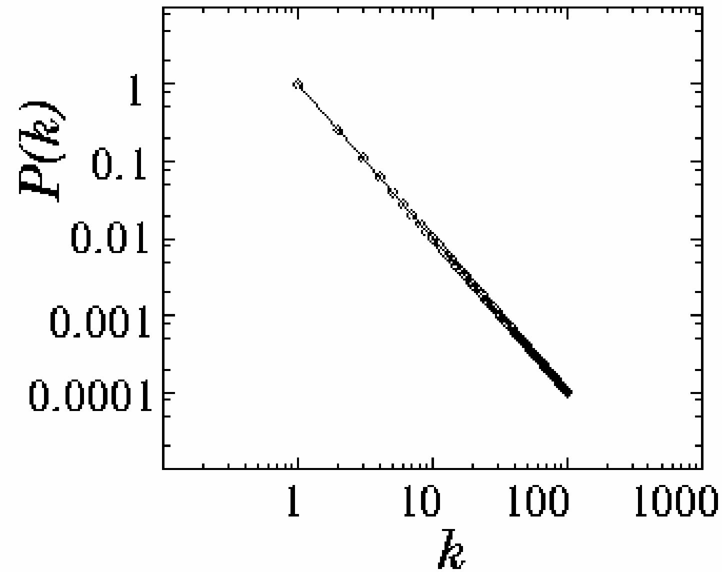
R. Albert, H. Jeong, A-L Barabasi, *Nature*, **401** 130 (1999).

What does it mean?

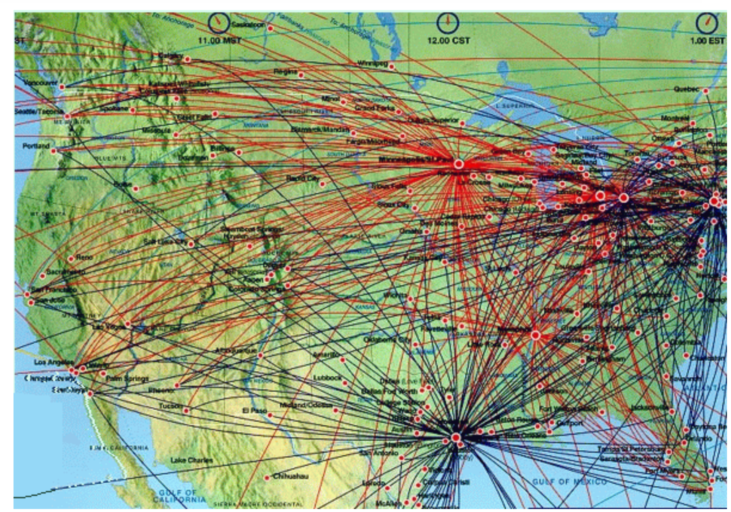
Poisson distribution



Power-law distribution



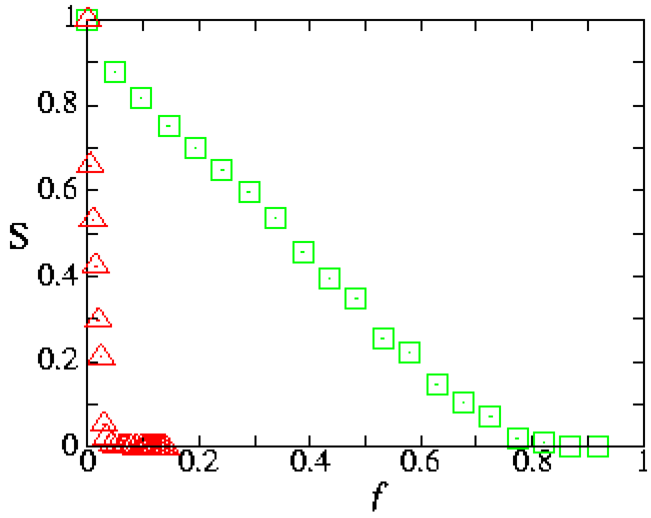
Exponential Network

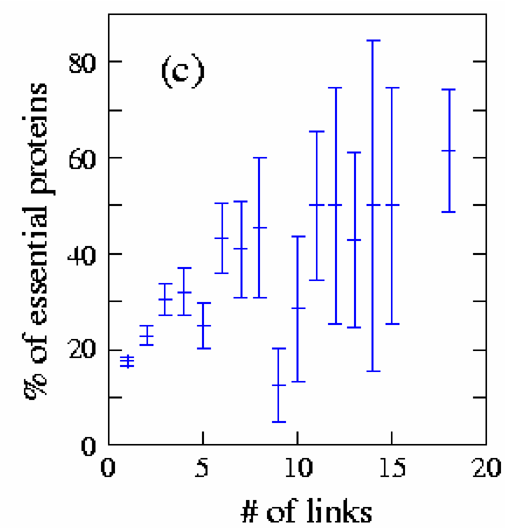


Scale-free Network

Yeast protein network

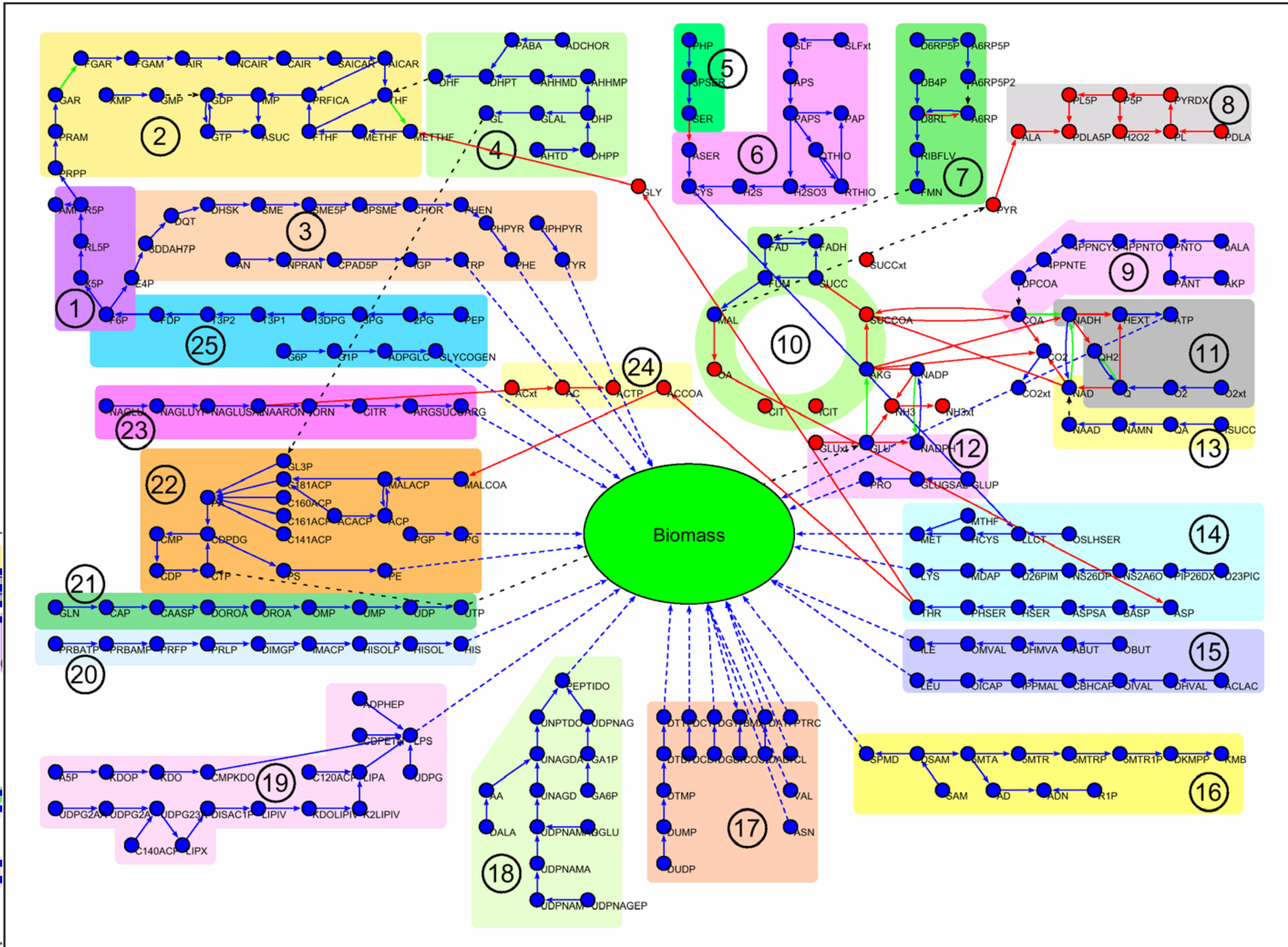
- lethality and topological position -





Highly connected proteins are more **essential (lethal)**...

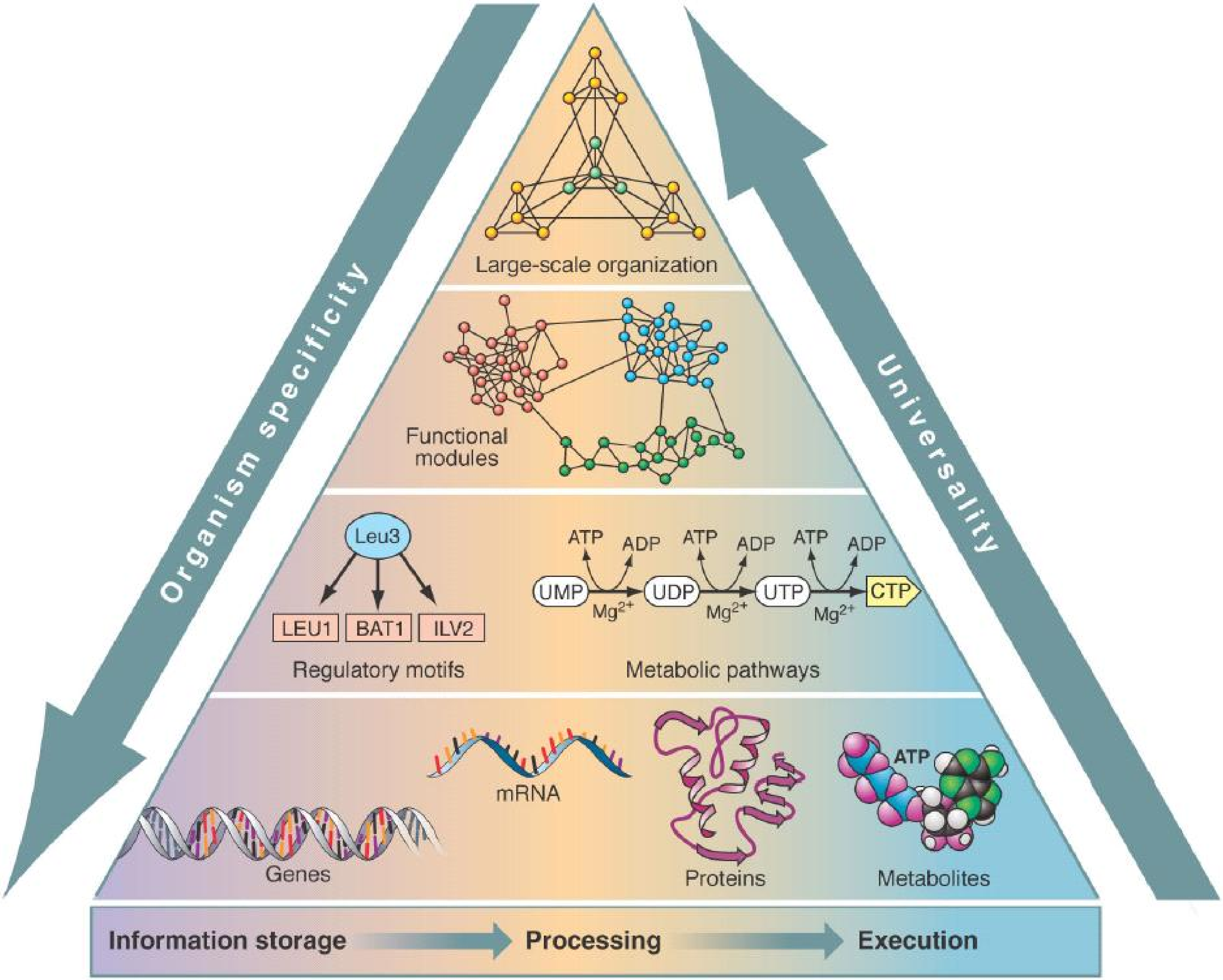
H. Jeong, S.P. Mason, A.-L. Barabasi, Z.N. Oltvai, *Nature* 411, 41-42 (2001)



Glutamate rich substrate

Succinate rich substrate

Life's Complexity Pyramid



Z.N. Oltvai and A.-L. B. *Science*, 2002.