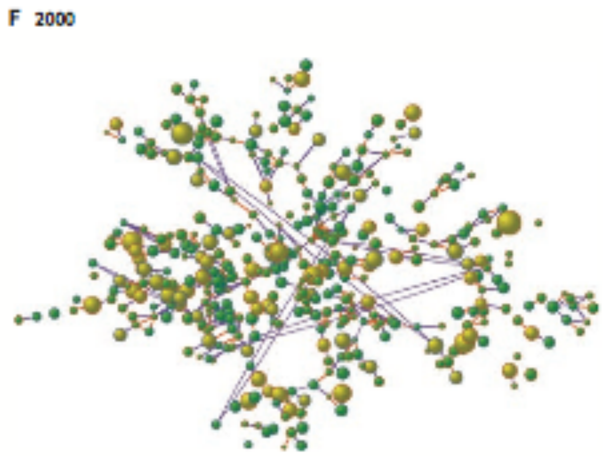
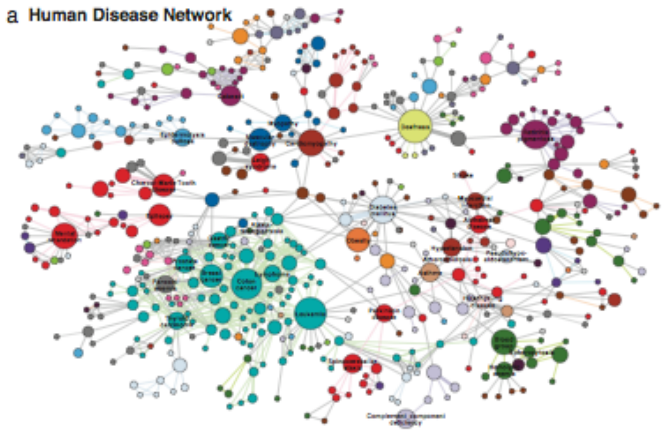




Complexity Summer School

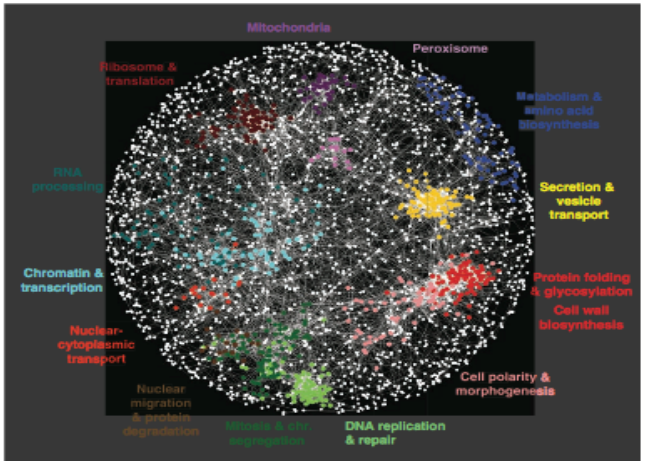
Warwick, 28 Feb-3 May 2013

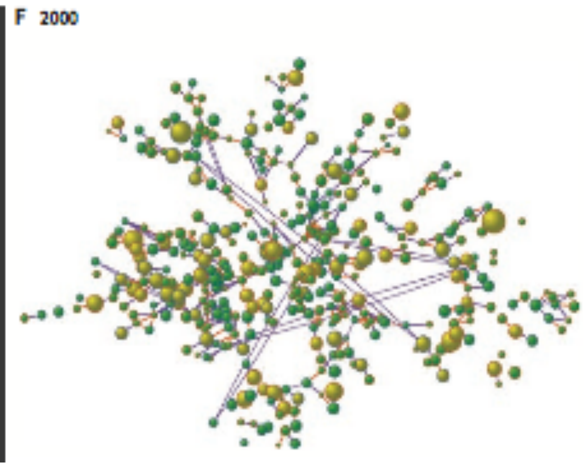
Structure of complex networks

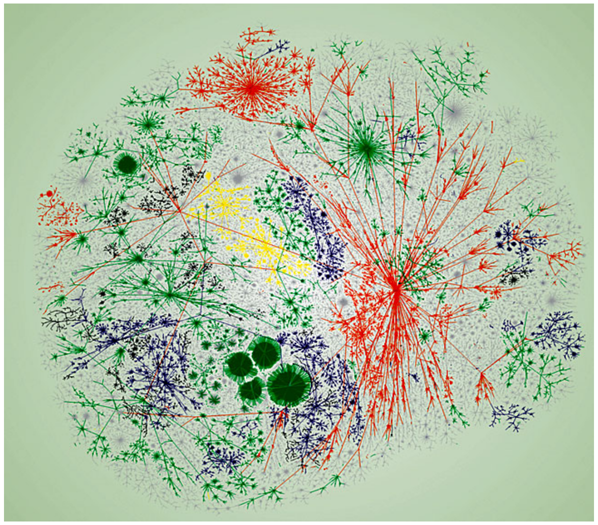
Ginestra Bianconi

School of Mathematics, Queen Mary University of London, London UK

Complex networks





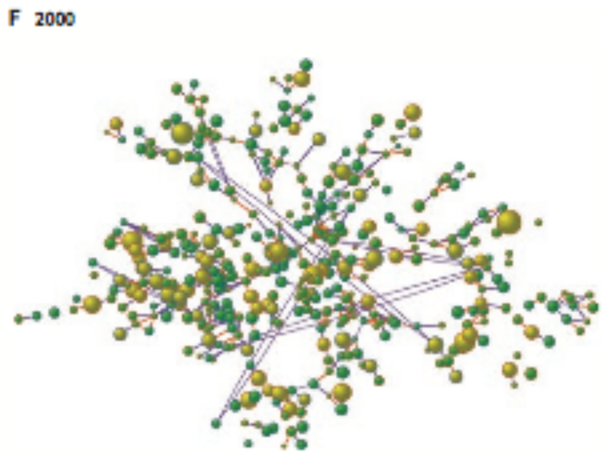


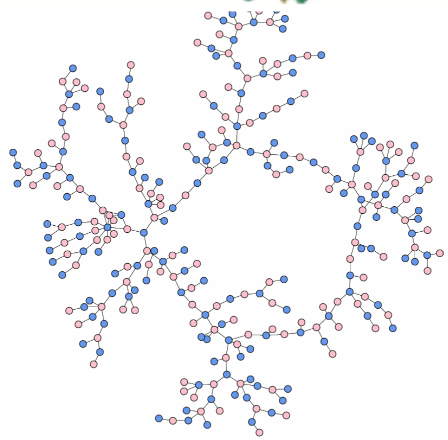
describe

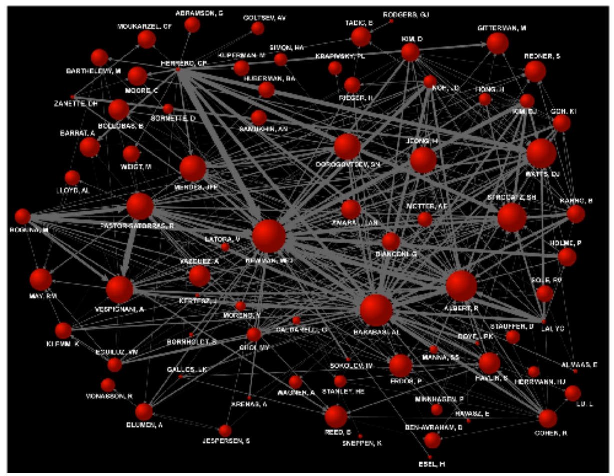
the underlying structure of interacting complex

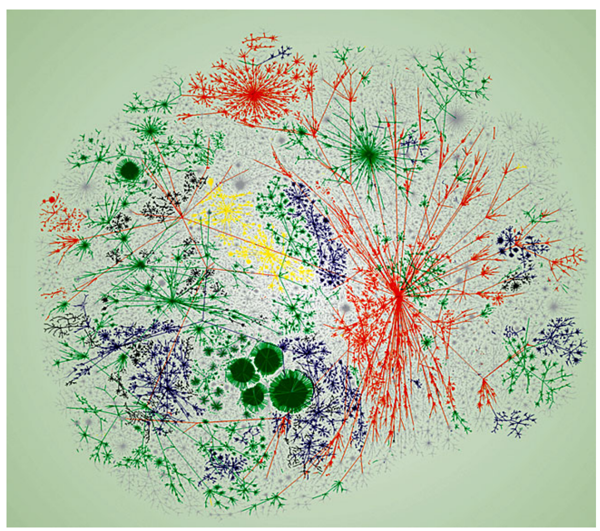
Biological, Social and Technological systems.

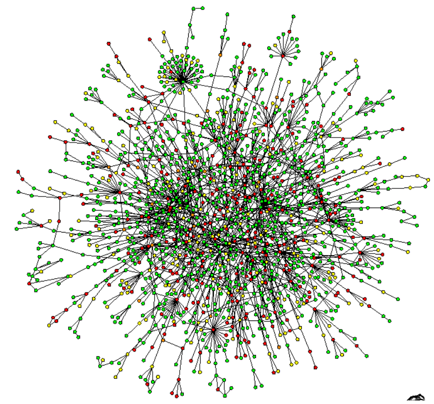
Networks

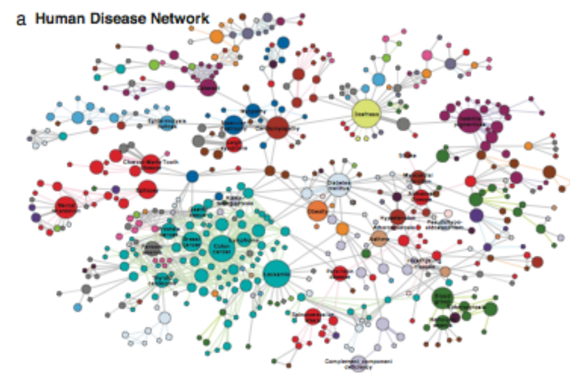

F 2000

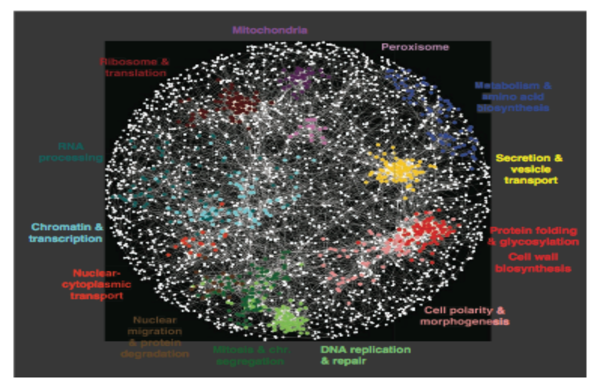









a Human Disease Network


Mitochondria
Peroxisome
Ribosome & translation
Secretion & vesicle transport
Chromatin & transcription
Protein folding & glycosylation
Cell wall biosynthesis
Nuclear-cytoplasmic transport
Cell polarity & morphogenesis
DNA replication & repair

Why working on networks?

Because

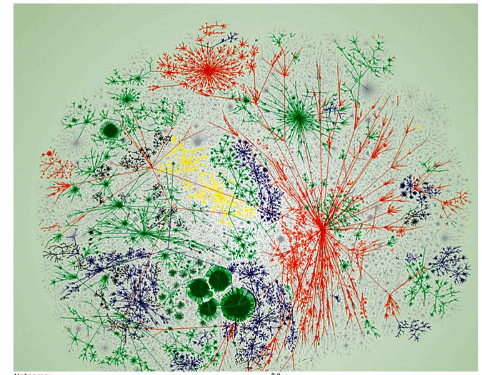
NETWORKS

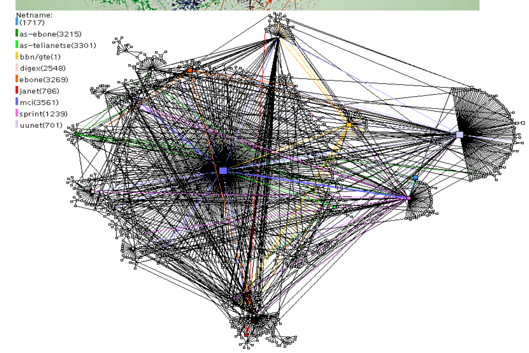
encode for the
ORGANIZATION,
FUNCTION,
ROBUSTENES
AND DYNAMICAL BEHAVIOR
of the entire complex system

Types of networks

➤Simple Each link is either existent or non existent, the links do not have directionality

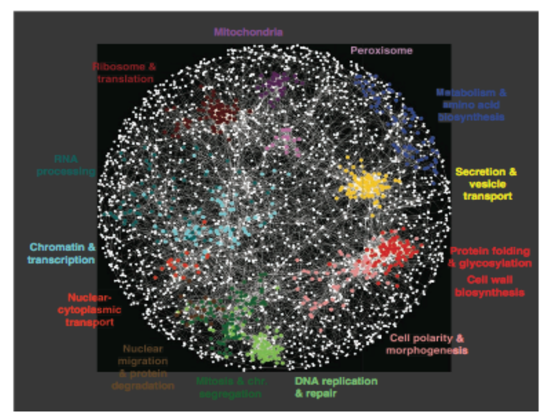
(protein interaction map, Internet,...)





➤Directed The links have directionality, i.e., arrows

(World-Wide-Web, social networks…)



➤Signed The links have a sign

(transcription factor networks, epistatic networks…)

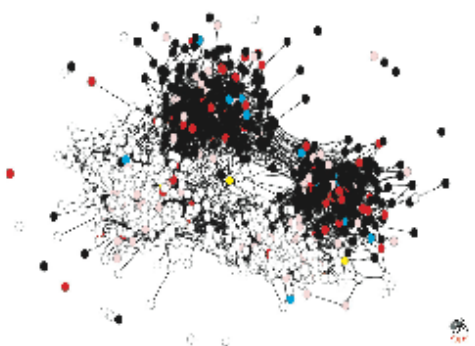
Types of networks

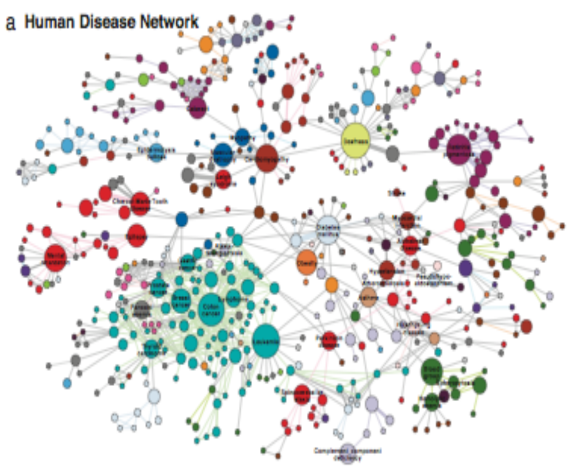
- **Weighted** The links are associated to a real number indicating their weight

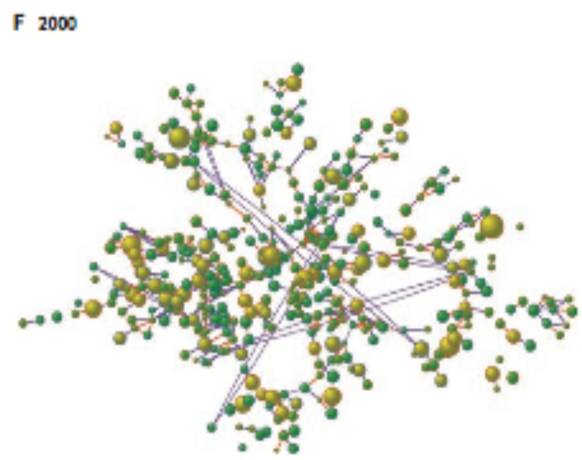
(airport networks, phone-call networks…)

- **With features of the nodes** The nodes might have weight or color

(social networks, diseasome, ect..)







Bipartite networks: ex. Metabolic network

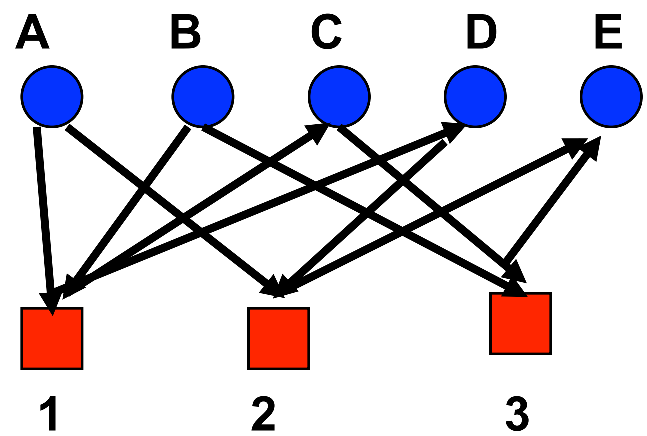
Reaction pathway

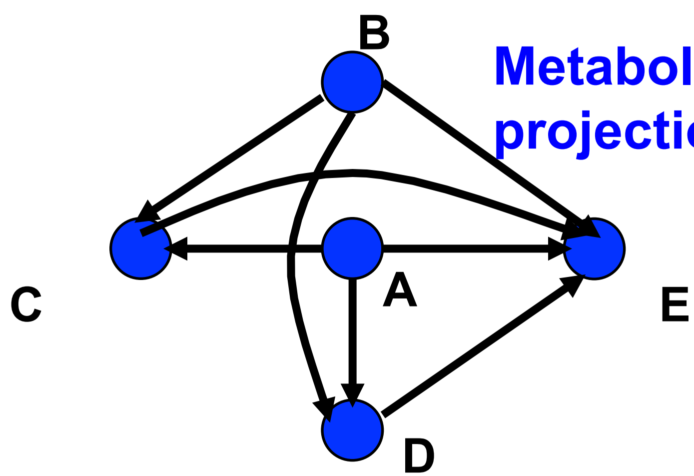
$A+B \rightarrow C+D$ 1

$A+D \rightarrow E$ 2

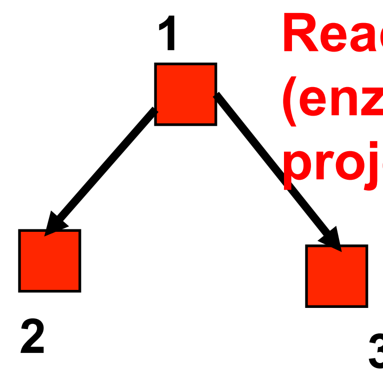
$B+C \rightarrow E$ 3

Bipartite Graph





Metabolites projection



Reactions (enzymes) projection

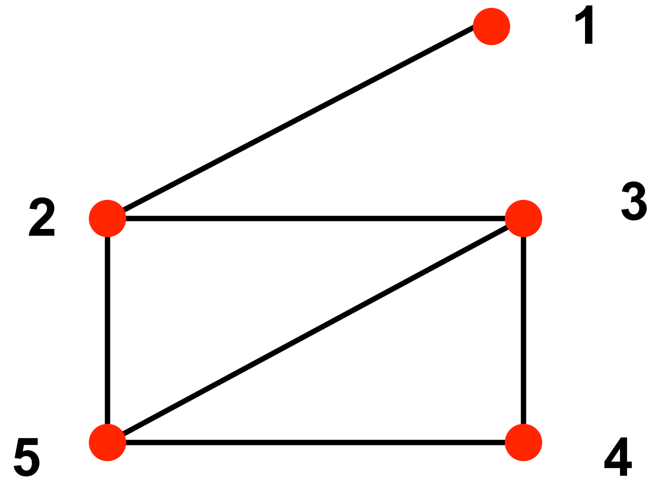
Total number of nodes N and links L

The total number of nodes N
and the total number of links L
are the most basic characteristics
of a network

Adjacency matrix

Network:
A set of labeled nodes and links between them

Adjacency matrix:
The matrix of entries a(i,j)=1 if there is a link between node i and j
a(i,j)=0 otherwise



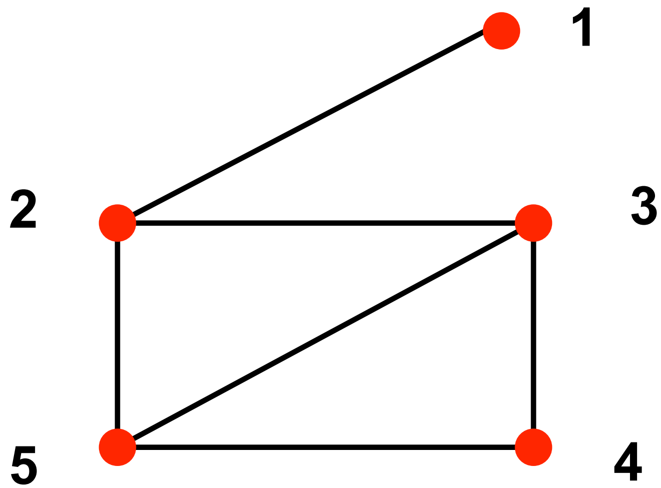
$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Degree distribution

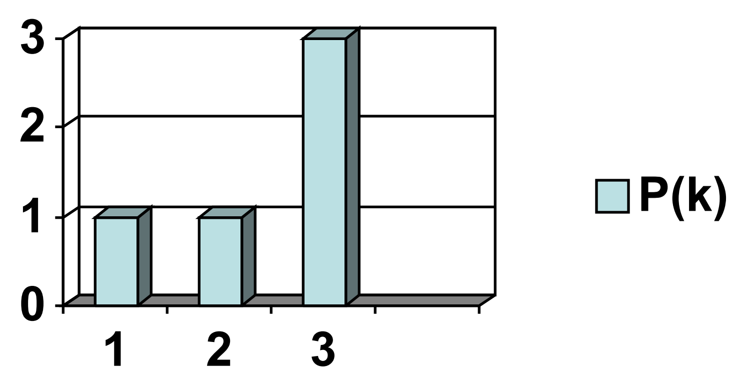
Degree if node i:
Number of links of node i

$$k_i = \sum_j a_{ij}$$

Network:



Degree distribution:
P(k): how many nodes have degree k



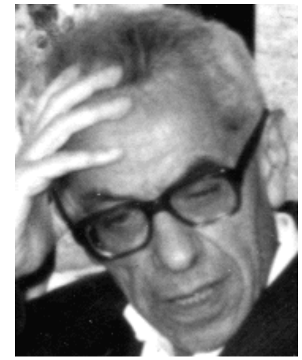
Random graphs

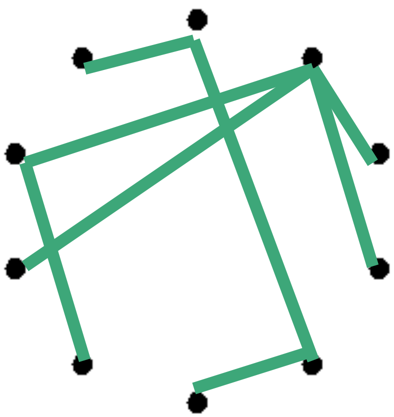
G(N,L) ensemble

Graphs with exactly N nodes and L links

G(N,p) ensemble

Graphs with N nodes
Each pair of nodes is linked with probability p



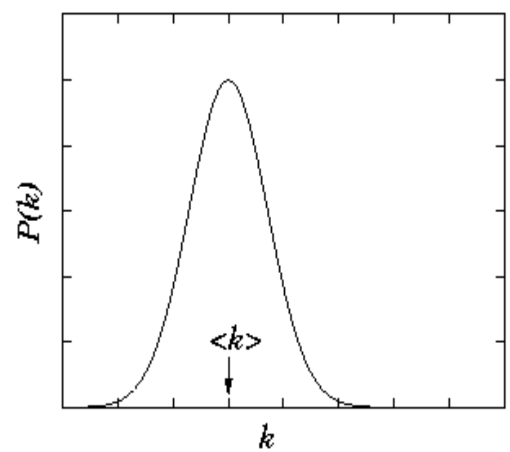


Binomial distribution

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

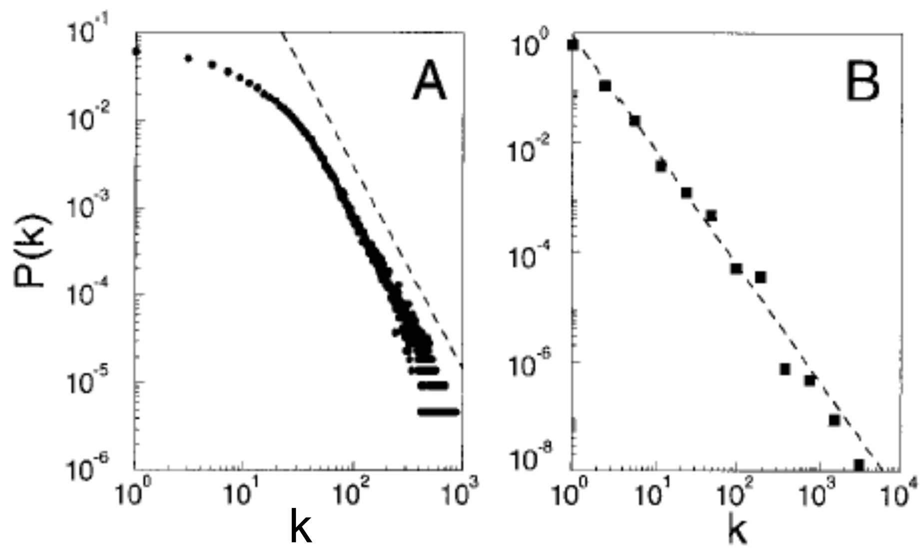
$$P(k) = \frac{1}{k!} c^k e^{-c}$$

Poisson distribution



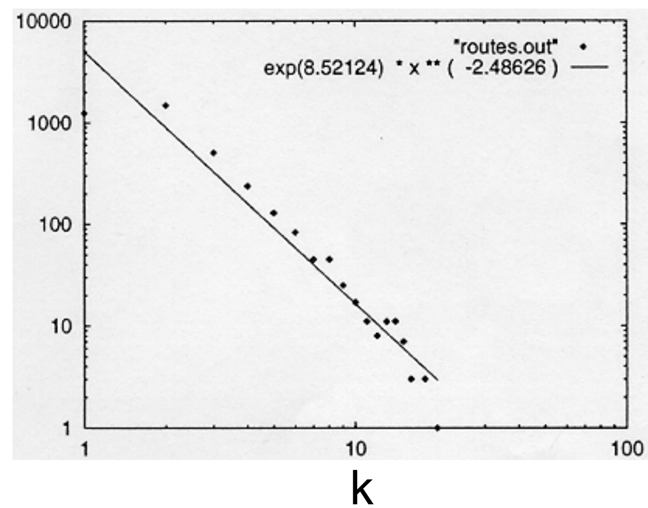
Universalities:
Scale-free degree distribution

Actor networks WWW



Barabasi-Albert 1999

Internet



Faloutsos et al. 1999

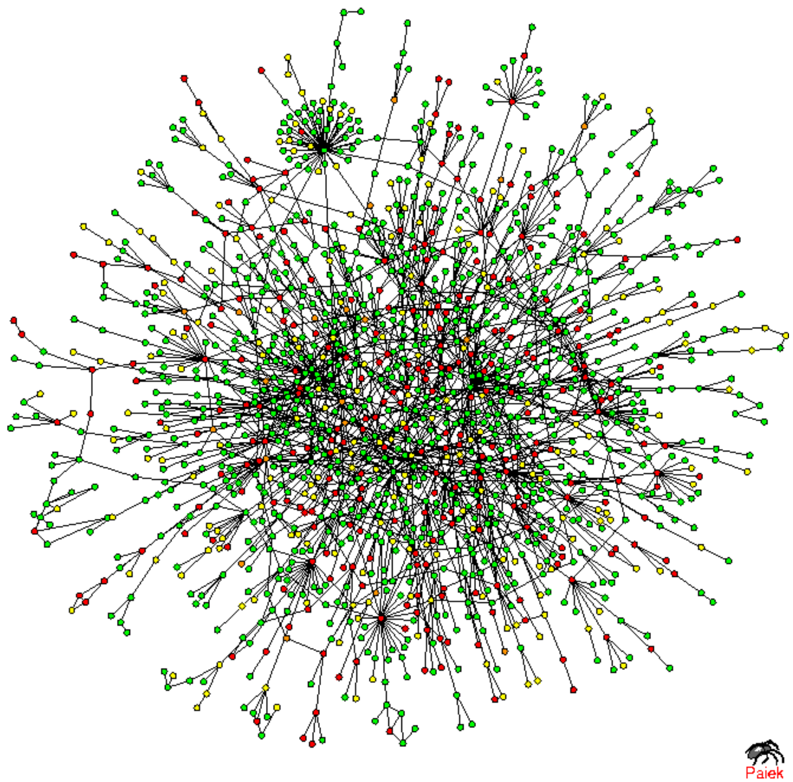
$$P(k) \propto k^{-\gamma} \quad \gamma \in (2,3)$$

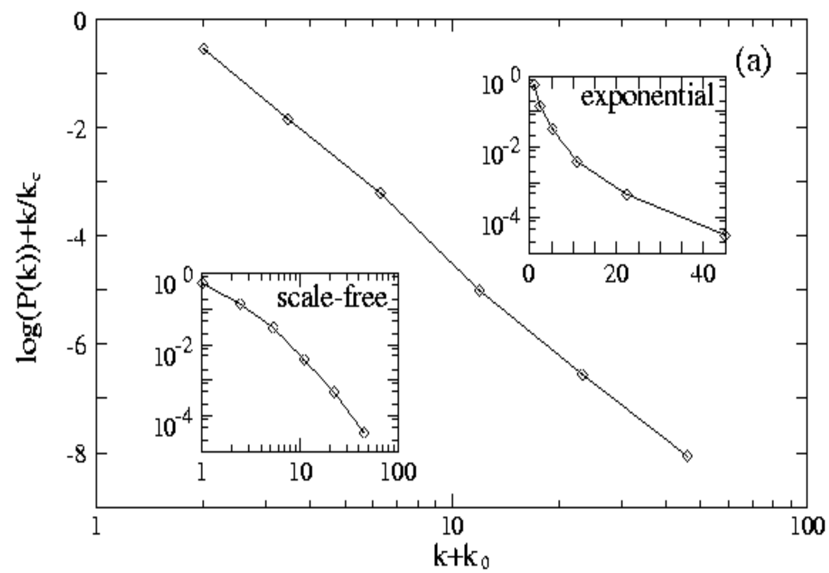
$$\langle k \rangle \quad \textit{finite}$$

$$\langle k^2 \rangle \to \infty$$

Topology of the yeast protein network

H. Jeong et al. Nature (2001)





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

$$\gamma \approx 2.5$$

Scale-free networks

- **Technological networks**:
  - Internet, World-Wide Web
- **Biological networks :**
  - Metabolic networks,
  - protein-interaction networks,
  - transcription networks
- **Transportation networks:**
  - Airport networks
- **Social networks:**
  - Collaboration networks
  - citation networks
- **Economical networks:**
  - Networks of shareholders in the financial market
  - World Trade Web

Why this universality?

- Growing networks:
  - **Preferential attachment**

Barabasi & Albert 1999,

Dorogovtsev Mendes 2000,Bianconi & Barabasi 2001, etc.

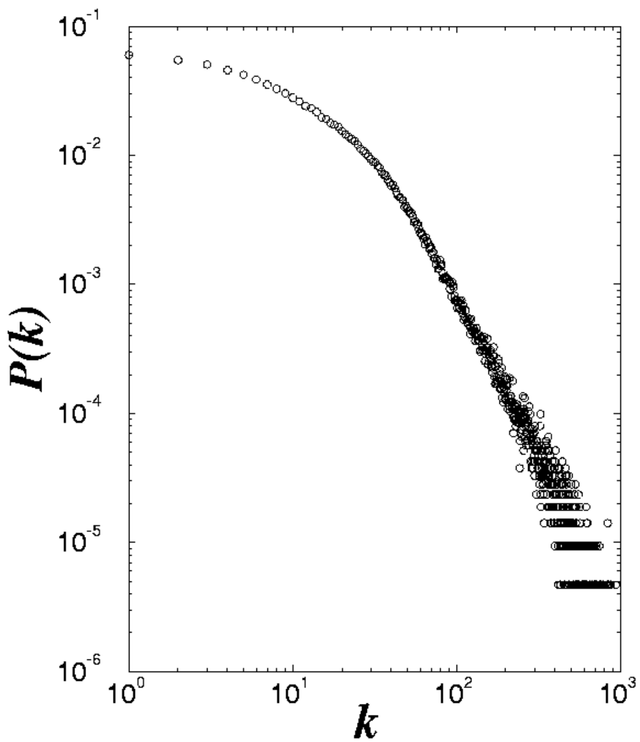
- Static networks:
  - **Hidden variables mechanism**

Chung & Lu 2002, Caldarelli et al. 2002,

Park & Newman 2003

Scale-free networks

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



with

$$\gamma > 3 \qquad \langle k \rangle \ \ finite \qquad \langle k^2 \rangle \ \ finite$$

with

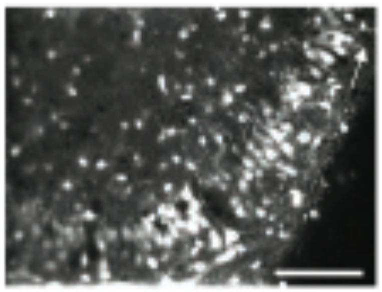
$$2 < \gamma \leq 3 \qquad \langle k \rangle \ \ finite \qquad \langle k^2 \rangle \rightarrow \infty$$

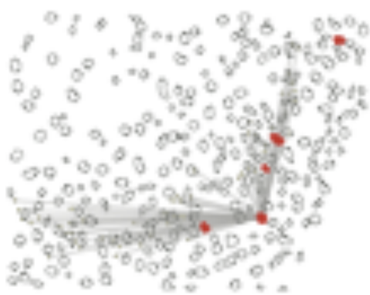
with

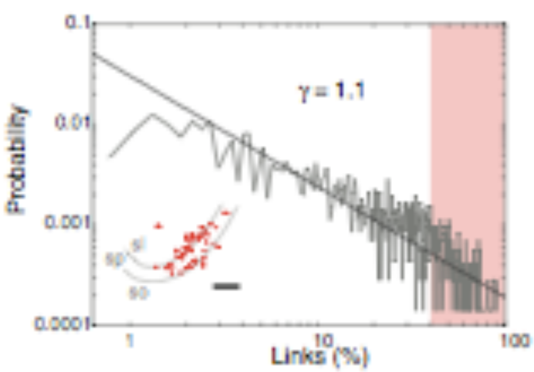
$$1 < \gamma \leq 2 \qquad \langle k \rangle \rightarrow \infty \qquad \langle k^2 \rangle \rightarrow \infty$$

Hyppocampus functional neural network

Bonifazi et al . Science 2009







Dense scale-free networks

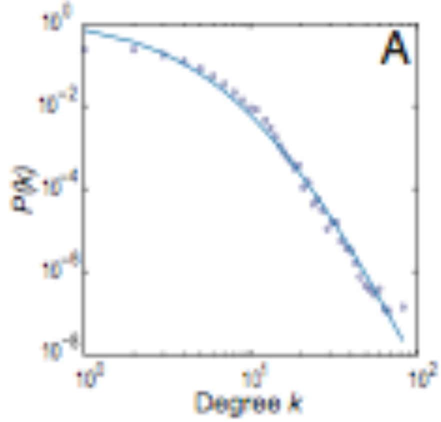
$$P(k) \propto k^{-\gamma}$$

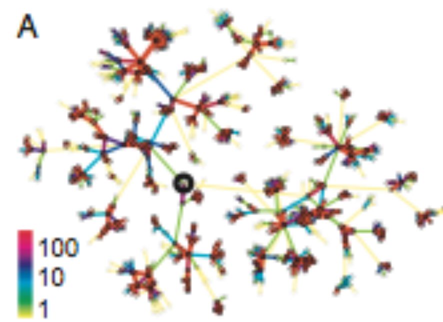
$$\gamma \in (1,2]$$

$$\langle k \rangle \rightarrow \infty$$

$$\langle k^2 \rangle \rightarrow \infty$$

Social network of phone calls





J. P Onnela
PNAS 2007

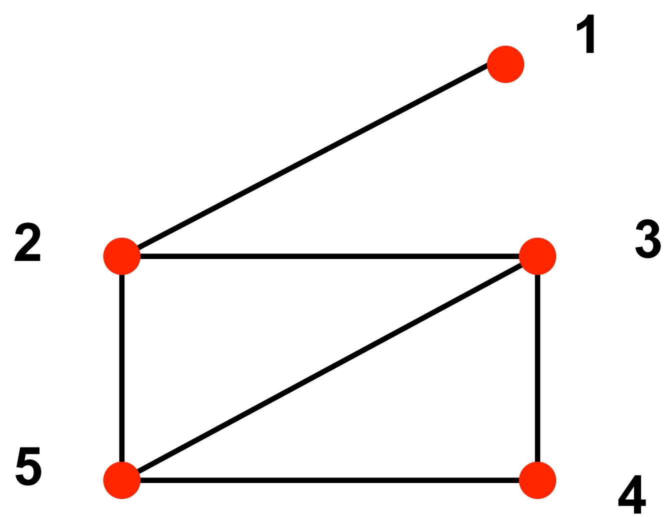
$$P(k) = \frac{a}{(k+k_0)^{\gamma}} e^{-k/k_c}$$

$$\gamma = 8.4$$

Finite scale network

Shortest distance

The **shortest distance** between two nodes *is the minimal number of links than a path must hop to go from the source to the destination*



The shortest distance

- **between node 4 and node 1 is 3**
- **between node 3 and node 1 is 2**

Diameter and average distance

The **diameter D** *of a network is the maximal length of the shortest distance between any pairs of nodes in the network*

The **average distance** $\langle \ell \rangle$ *is the average length of the shortest distance between any pair of nodes in the network*

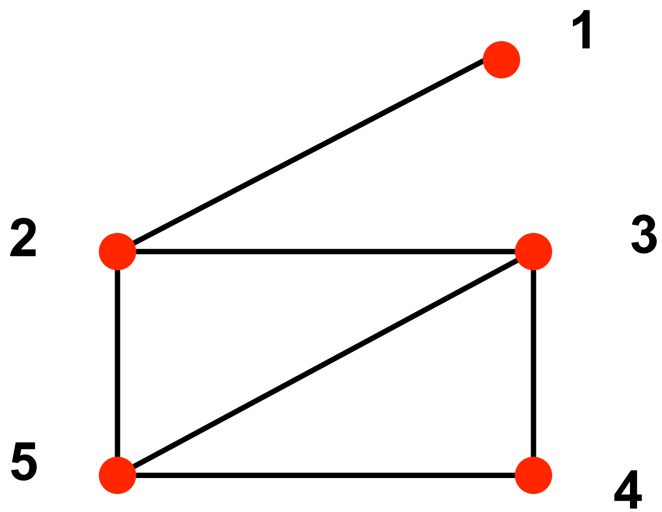
$$\boxed{D \geq \langle \ell \rangle}$$

Clustering coefficient

Definition of local clustering coefficient

$$C_i = \frac{|\#\, of\ \ triangles\, through\ i\,|}{k_i(k_i - 1)/2}$$

Network

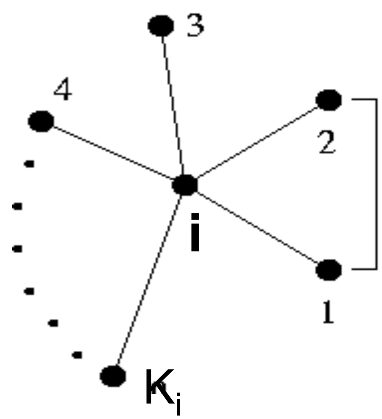


Clustering coefficient of nodes 2,3

$$C_2 = \frac{1}{3}$$

$$C_3 = \frac{2}{3}$$

Universality: Small world



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

Networks are clustered (large average C_i, i.e. C)

but have a small characteristic path length (small L).

Network	C	C_{rand}	L	N
WWW	0.1078	0.00023	3.1	153127
Internet	0.18-0.3	0.001	3.7-3.76	3015-6209
Actor	0.79	0.00027	3.65	225226
Coauthorship	0.43	0.00018	5.9	52909
Metabolic	0.32	0.026	2.9	282
Foodweb	0.22	0.06	2.43	134
C. elegance	0.28	0.05	2.65	282

Watts and Strogatz (1999)

Small world in social networks

1967 Milligran experiment

People from Nebraska and Kansas were asked to contact a person in Boston though their network of acquaintances

The strength of weak ties

(Granovetter 1973)

Weak ties help navigate the social networks

Small-world model

(Watts & Strogatz 1998)

The model for coexistence of high clustering and small distances in complex networks

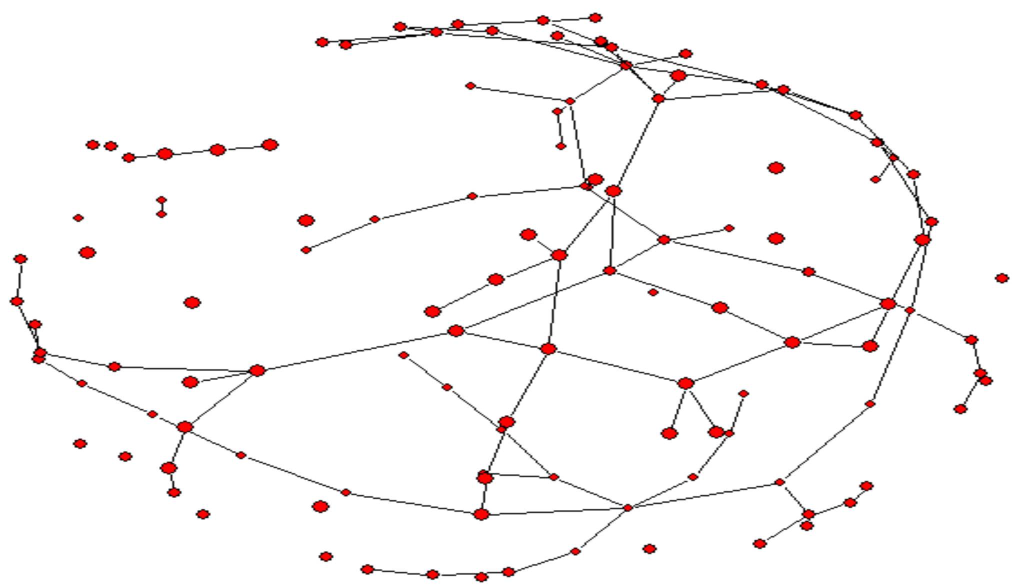
Diameter and average path length

	Diameter	Average distance
Poisson networks	$D \approx \frac{\log(N)}{\log(<k>)}$	$\langle \ell \rangle \approx \frac{\log(N)}{\log(<k>)}$
Scale-free networks	$D \approx \log(N)$	$\langle \ell \rangle \approx \begin{cases} \frac{\log N}{\log <k>} & if\ \gamma > 3 \\ \frac{\log(N)}{\log(\log(N))} & if\ \gamma = 3 \\ \log(\log(N)) & if\ \gamma < 3 \end{cases}$
	Chung & Lu 2004	*Cohen &Havlin 2003*
	Small world property	Ultra small world property

Giant component

➢ *A* **connected component** *of a network is a subgraph such that for each pair of nodes in the subgraph there is at least one path linking them*

➢ *The* **giant component** *is the connected component of the network which contains a number of nodes of the same order of magnitude of the total number of links*



Molloy-Reed principle

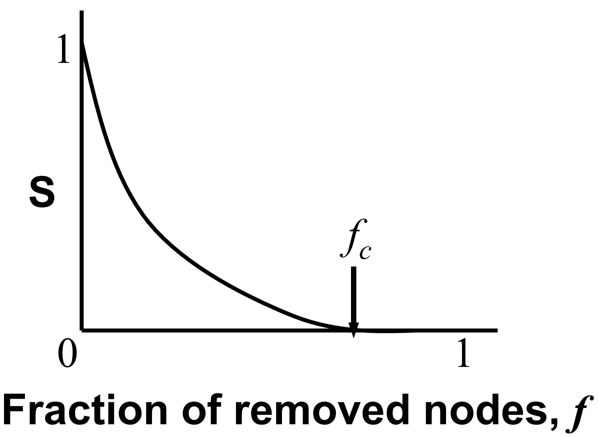
A network has a giant component if $\frac{<k^2>}{<k>} \geq 2$

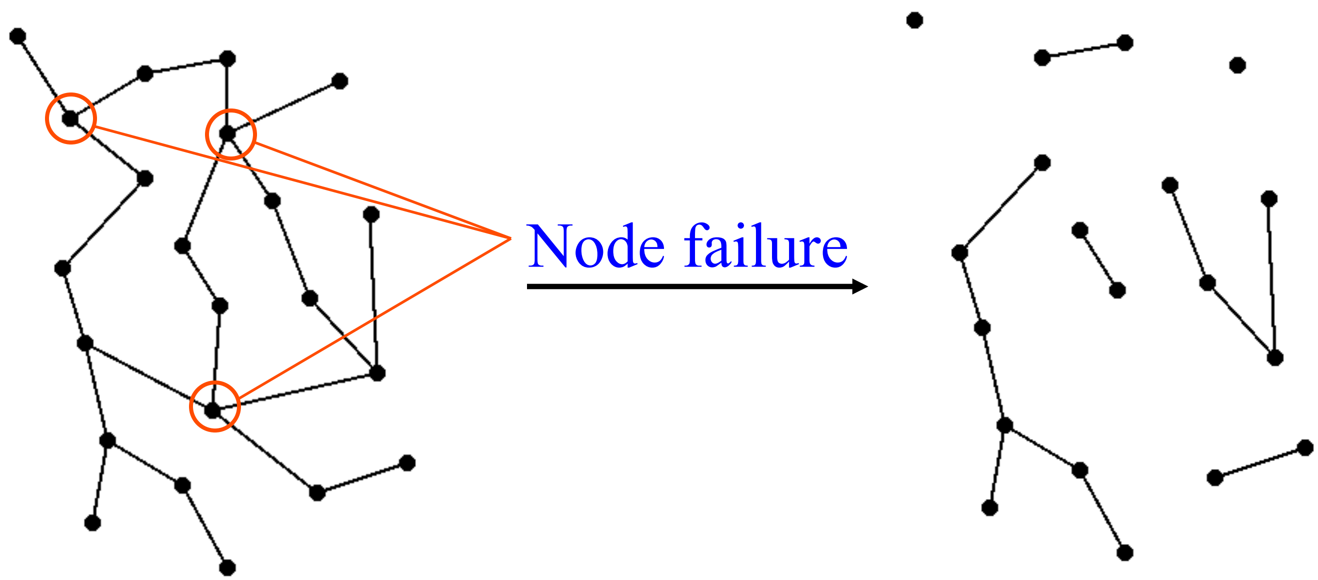
Molloy-Reed 1995

	Second moment $<k^2>$	Molloy-Reed condition
Poisson networks	$<k^2> = <k>^2 + <k>$	$<k> \geq 1$
Scale-free Networks	$<k^2> \approx \begin{cases} \log(N) & \text{if } \gamma = 3 \\ N^{(3-\gamma)/2} & \text{if } \gamma < 3 \end{cases}$	*There is always a giant component as long as* $\gamma \leq 3$

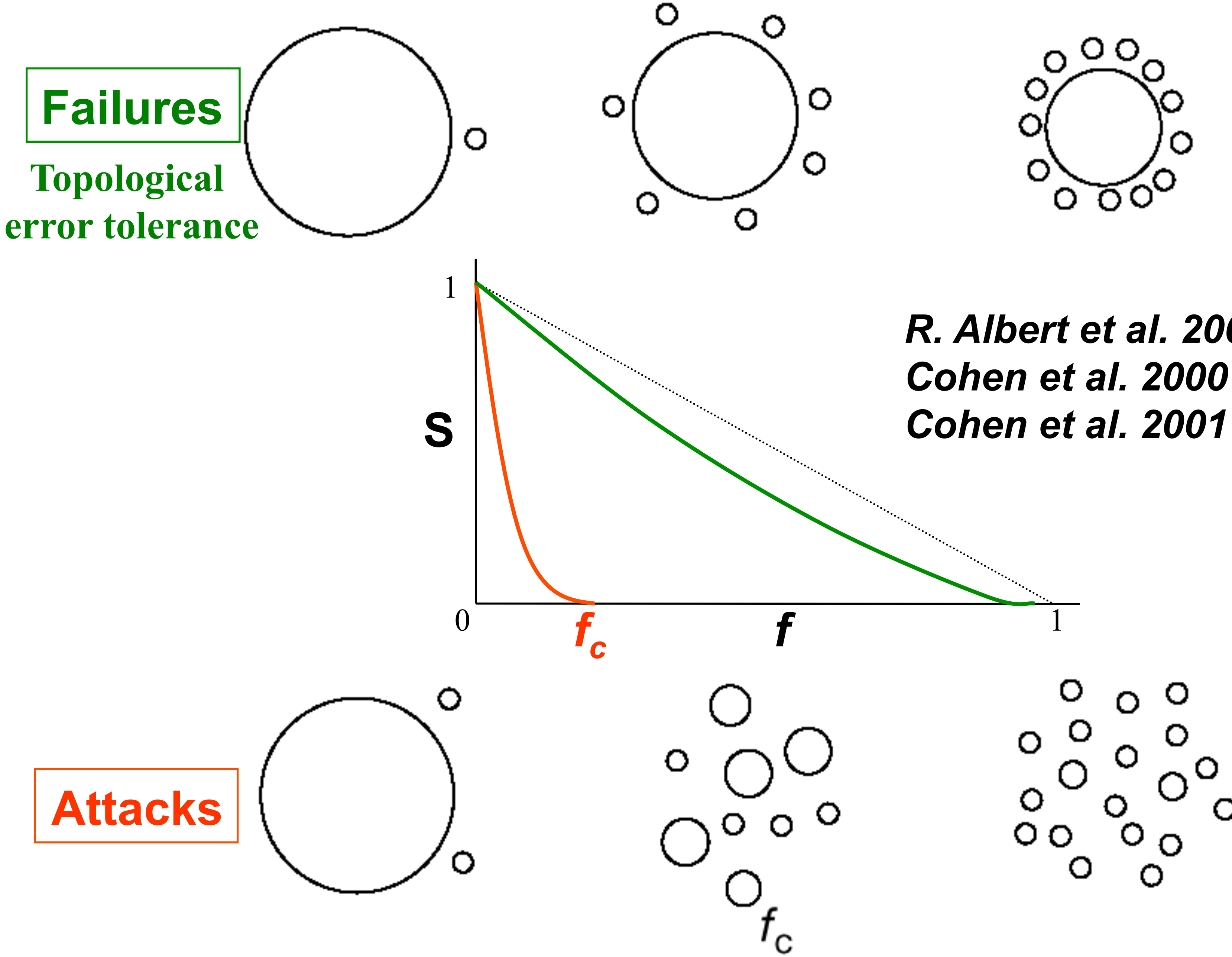
Robustness

Complex systems maintain their basic functions
even under errors and failures



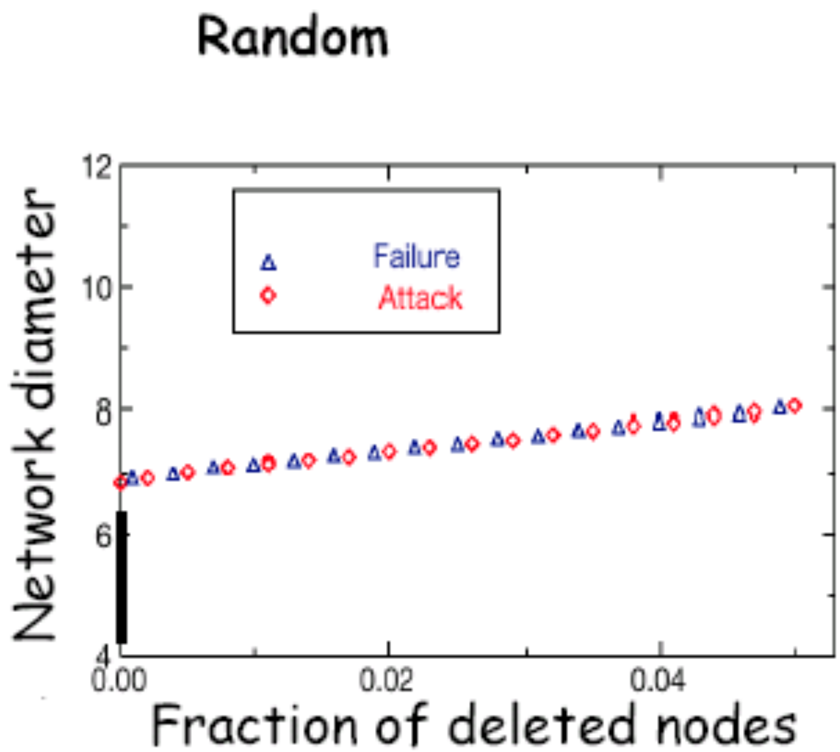


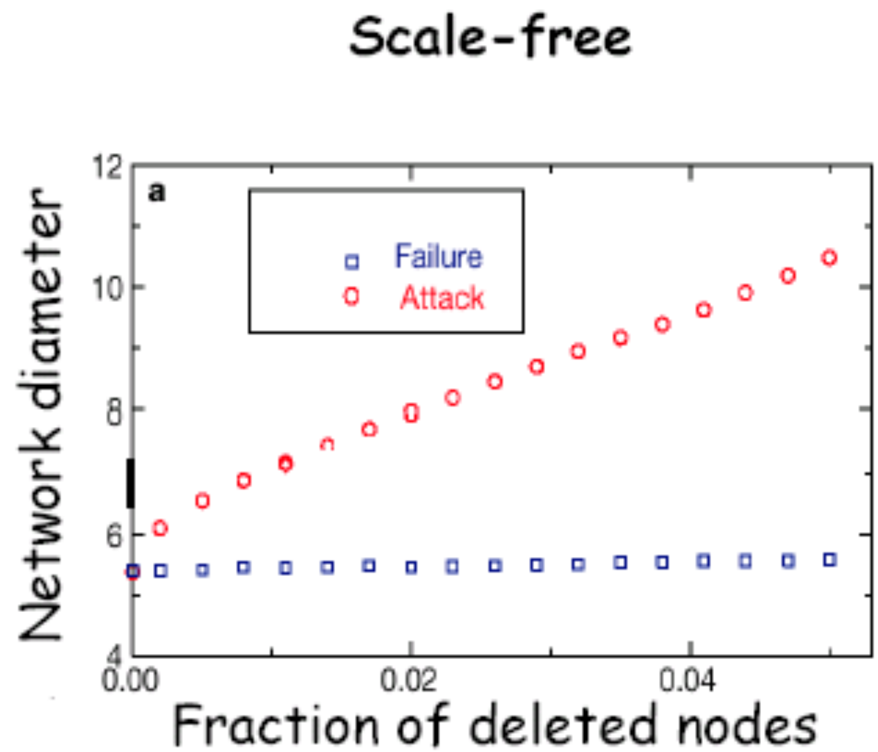
Robustness of scale-free networks



R. Albert et al. 2000
Cohen et al. 2000
Cohen et al. 2001

Robustness





Failure=Attack

Robust against Failure
Weak against Attack

R. Albert et al. 2000

Betweenness Centrality

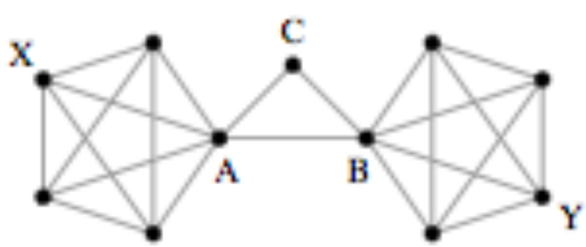
The betweenness centrality *of a node v*

is given by

$$B(v) = \sum_{s,t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths between s and t

$\sigma_{st}(v)$ are the number of such paths passing through v

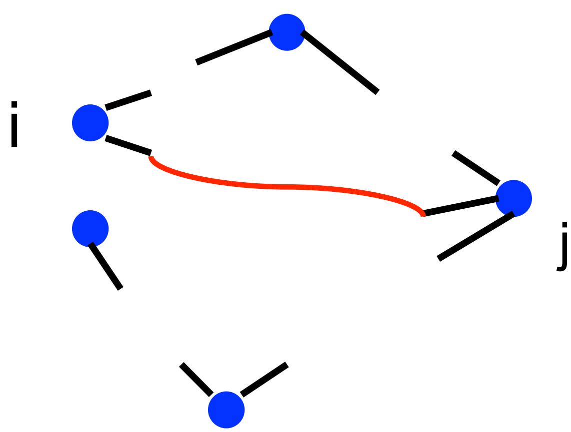


The nodes A and B are also called bridges and have high betweeness

Link probability in uncorrelated networks

In uncorrelated networks *the probability that a node i is linked to a node j is given by*

$$p_{ij} = \frac{k_i k_j}{<k>N}$$

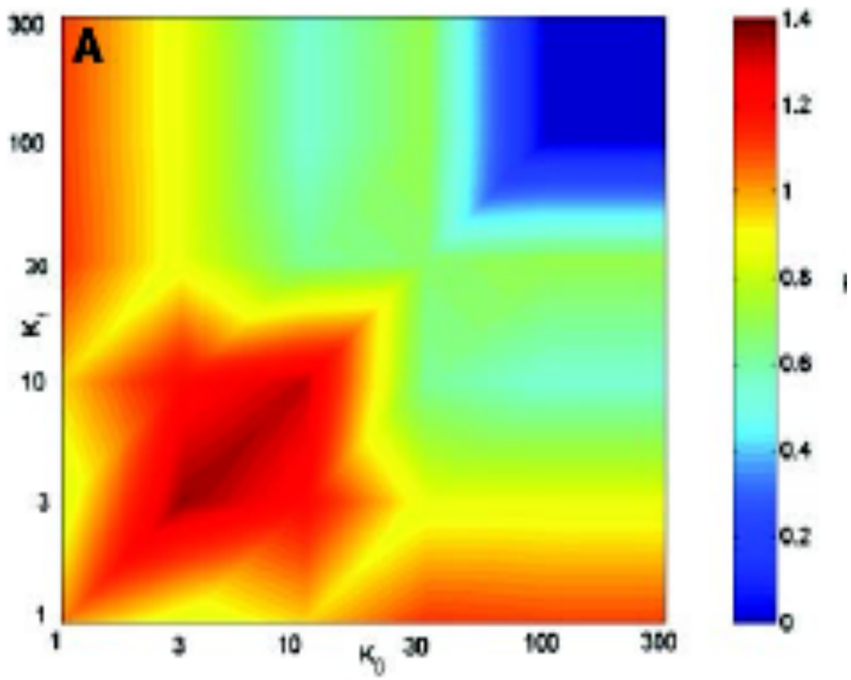


$$p_{ij} = 2\frac{3}{\langle k \rangle N}$$

Degree correlations

$$P_R(k,k') = \frac{kk'}{\langle k \rangle N} P(k)P(k')$$

Link probability between nodes of degree classes k and k'



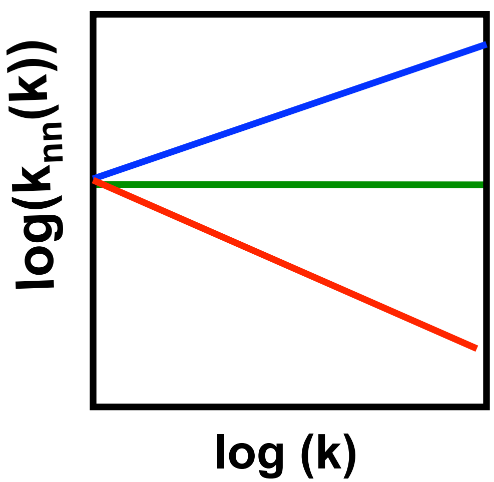
S. Maslov and K. Sneppen Science 2002

The map of

$$\frac{P(k,k')}{P_R(k,k')}$$

reveals correlations in the protein interaction map

Specific characteristic of networks: Degree correlations



$$k_{nn}(k) \approx k^{\alpha}$$

Assortative networks $\alpha>0$

Uncorrelated networks $\alpha=0$

Disassortative networks $\alpha<0$

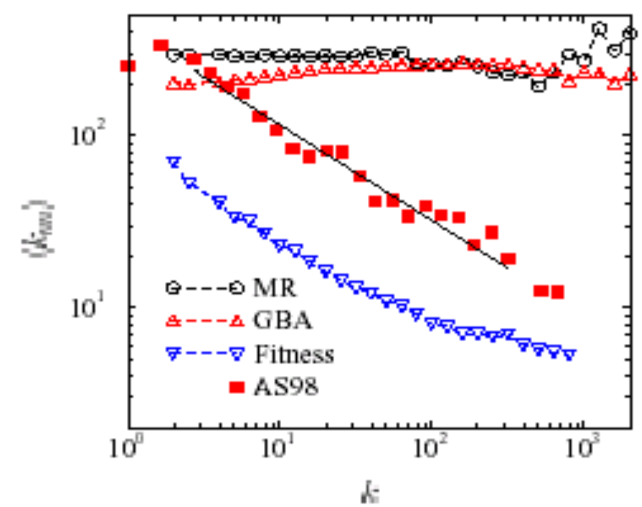
Average degree of a neighbor of a node of degree k

$$k_{nn}(k) = \left\langle \frac{1}{k_i} \sum_{j \in N(i)} k_j \right\rangle_{k_i = k}$$

Degree-degree correlations in the Internet at the AS level

$$k_{nn}(k) \approx k^{\alpha} \quad \alpha < 0$$

the network is disassortative

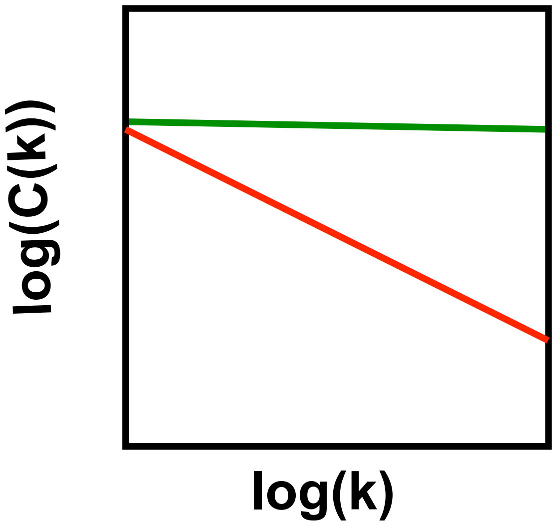


Vazquez et al. PRL 2001

Average degree of a neighbor of a node of degree k

$$k_{nn}(k) = \left\langle \frac{1}{k_i} \sum_{j \in N(i)} k_j \right\rangle_{k_i = k}$$

Correlations and clustering coefficient



$$C(k) \approx k^{-\delta}$$

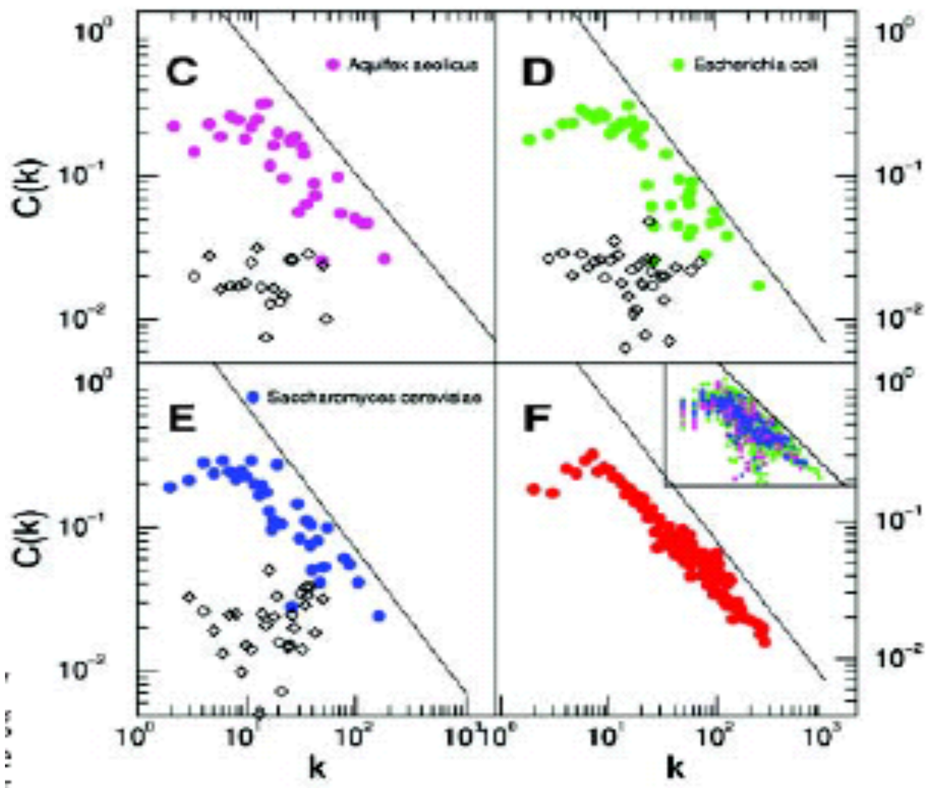
Uncorrelated networks $\delta=0$

Modular networks $\delta>0$

Average clustering coefficient C(k) of nodes of degree k

$$C(k) = \left\langle \frac{\sum_{i,j,r} a_{i,j} a_{jr} a_{r,i}}{k(k-1)} \right\rangle_{i|\langle k(i)\rangle = k}$$

Clustering coefficient of metabolic networks



Ravasz,et al. Science (2002).

$$C(k) \approx k^{-\delta}$$

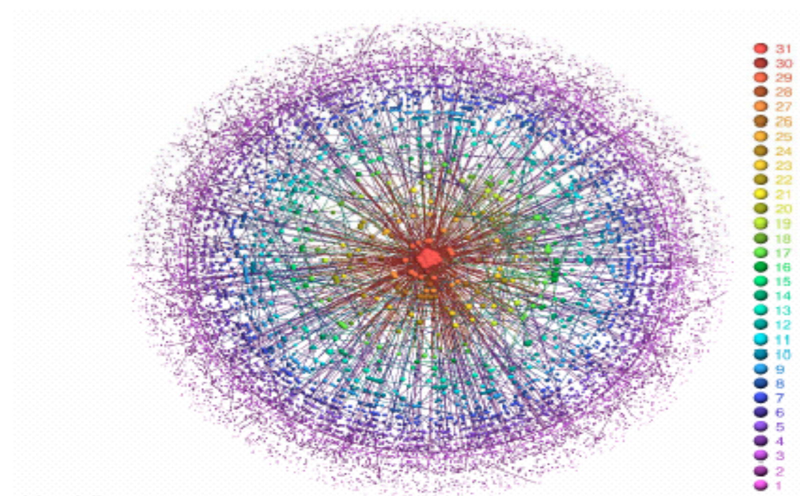
- There are important correlations in the network!
- The network is not 'random'
- Highly connected nodes link more "distant" nodes of the network

K-core of a network

A **K-core** *of a network*

is the subgraph of a network obtained by
removing the nodes with connectivity k_i<K
iteratively until the network
has only nodes with $k_i \geq K$

K-core of the Internet
DIMES Internet data
Visualization:
Alvarez-Hamelin et al. 2005



Using LaNet-Vi
http://xavier.informatics.indiana.edu/lanet-vi

K-cores on trees with given degree distributions

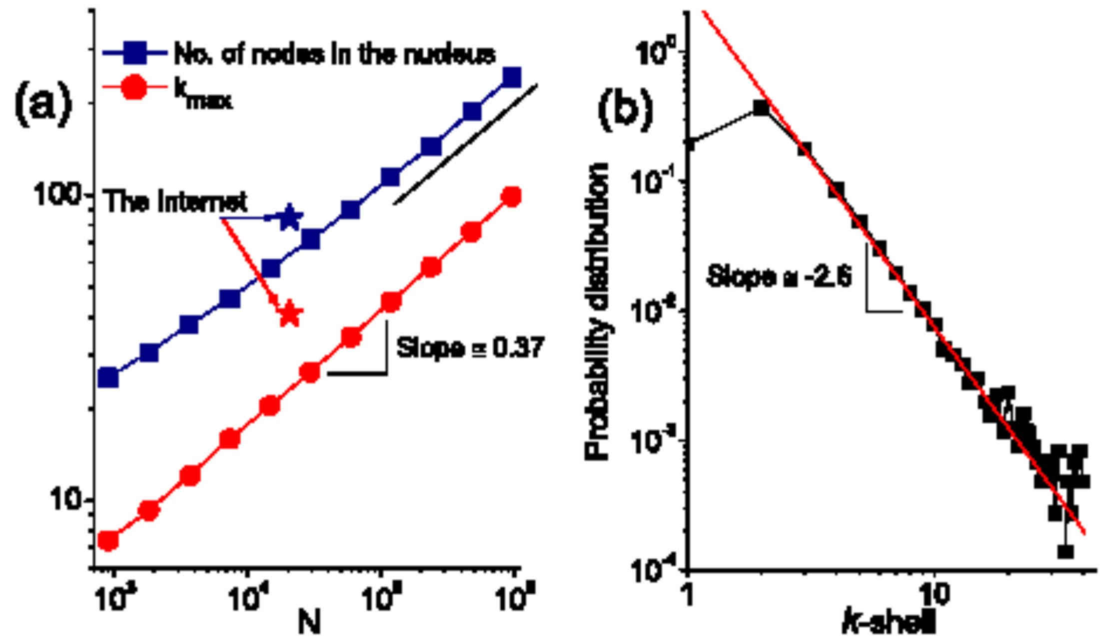
Scale-free networks

➢ ***Maximal K of the K-cores***

$$K_{\max} \approx pm^{\gamma-2}\, K_{cut}^{\;3-\gamma}$$

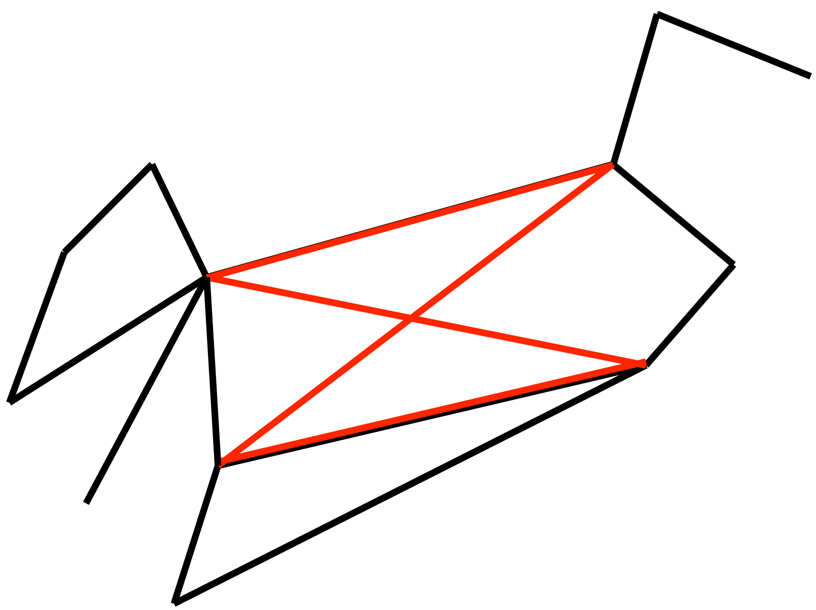
➢ ***Size of the minimal K-core***

$$M(K_{\max}) \approx p\left(\frac{m}{K_{cut}}\right)^{\gamma-1}$$



S. Dorogovtsev, et al. PRL (2006) Carmi et al. PNAS (2007)

How to build a null model form a given network: swap of connections



- Choose two random links linking four distinct nodes
- If possible (not already existing links) swap the ends of the links

Maslov & Sneppen 2002

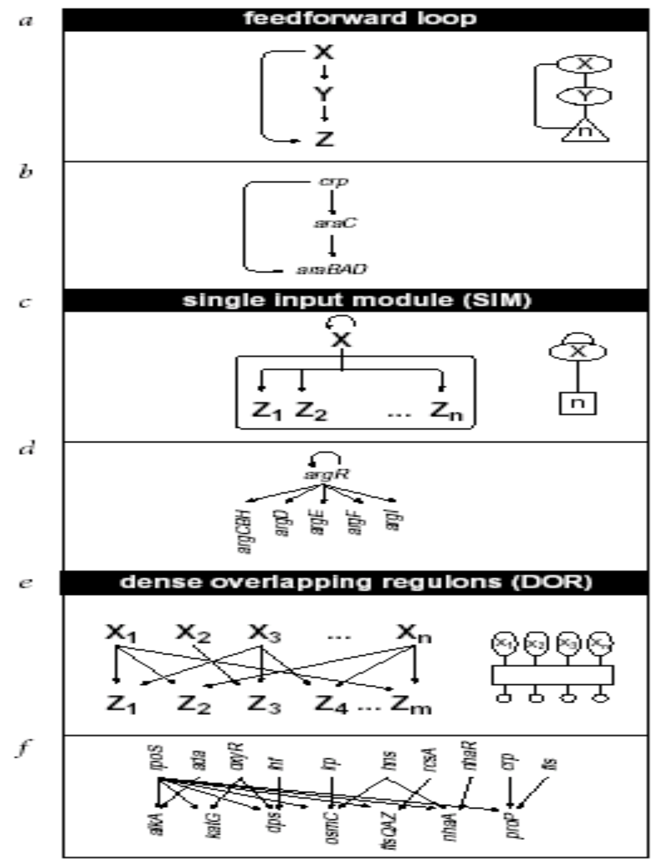
Motifs

The **motifs** *are subgraphs which appear with higher frequency in real networks than in randomized networks*

In biological networks the motifs are told to be selected by evolution

and are relevant to understand the function of the network.

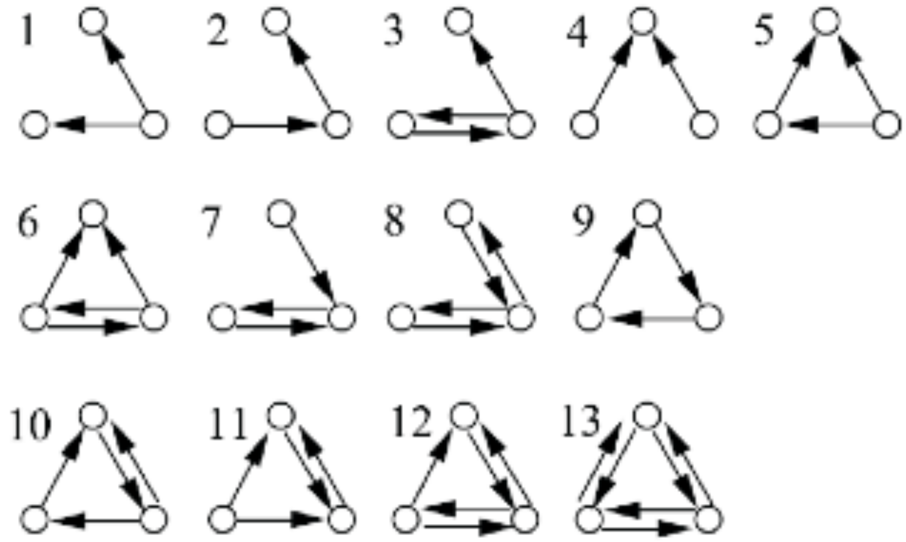
Milo et al. 2002



Motifs in the transcriptome network of e.coli.

S.S. Shen-Orr, et al., 2002

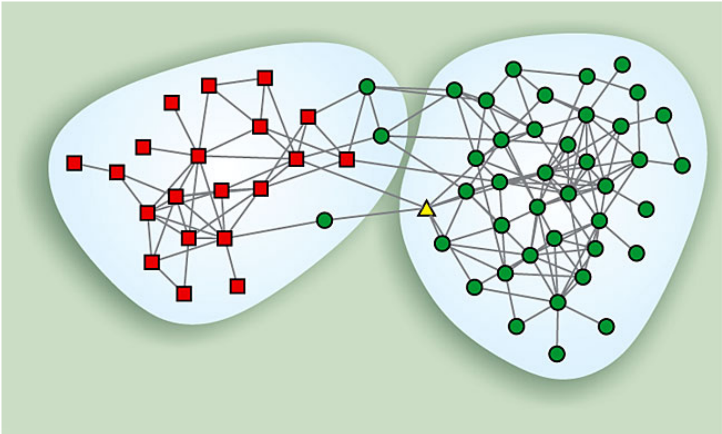
Motifs of size 3 in directed networks



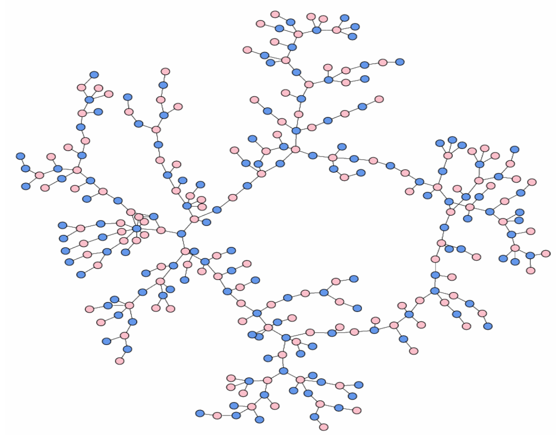
The number of distinguishable possible motifs increases exponentially with the motif size limiting the extension of this method to large subgraphs

Specific characteristics of a network: **communities**

Dolphins social network



High-school dating networks



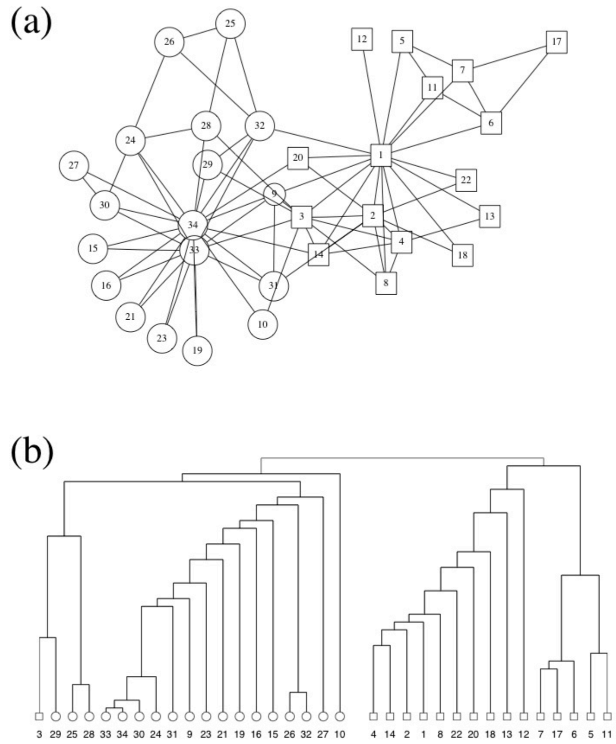
A community of a network define a set of nodes
with similar connectivity pattern.

S. Fortunato Phys. Rep. 2010

Girvan and Newman algorithm

The algorithm

1. The betweenness of all existing edges in the network is calculated first.
2. The edge with the highest betweenness is removed.
3. The betweenness of all edges affected by the removal is recalculated.
4. Steps 2 and 3 are repeated until no edges remain.



Girvan and Newman PNAS 2002

Modularity

- Assign a community s_i=1,2,…K to each node,
- The modularity Q is given by

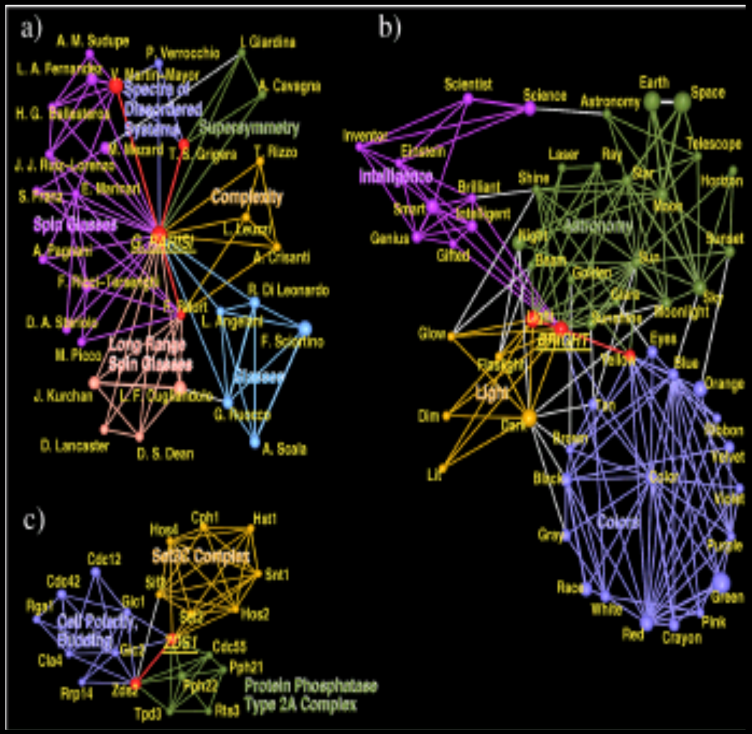
$$\boxed{Q = \frac{1}{\langle k \rangle N} \sum_{ij} \left[a_{ij} - \frac{k_i k_j}{\langle k \rangle N} \right] \delta(s_i, s_j)}$$

- Tight communities can be found by maximizing the modularity function

Newman PNAS 2006

Cliques for community detection

Overlapping set of cliques can be helpful to identify community structures in different networks

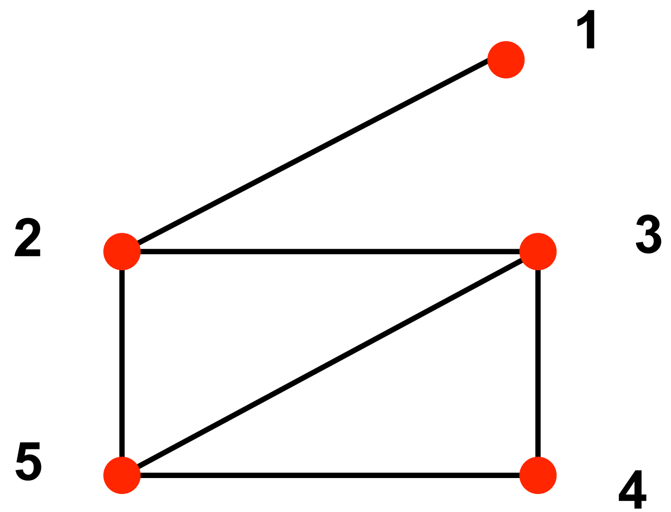


For this method to work systematically networks must have many cliques (ex. Scale-free networks)

Palla et al. Nature (2005)

Loops or cycles of size L

A loop or a cycle of size L
is a path of the network that start at one point and ends after hopping L links on the same point without crossing the intermediate nodes more than one time

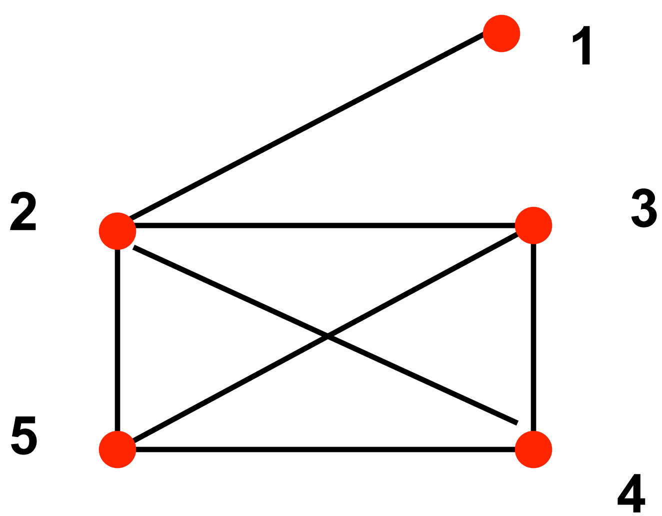


This network has

- ***2 loops of size 3***
- ***and 1 loop of size 4***

Cliques of size c

Clique of size c
is a fully connected subgraph of the network of c nodes and c(c-1)/2 links



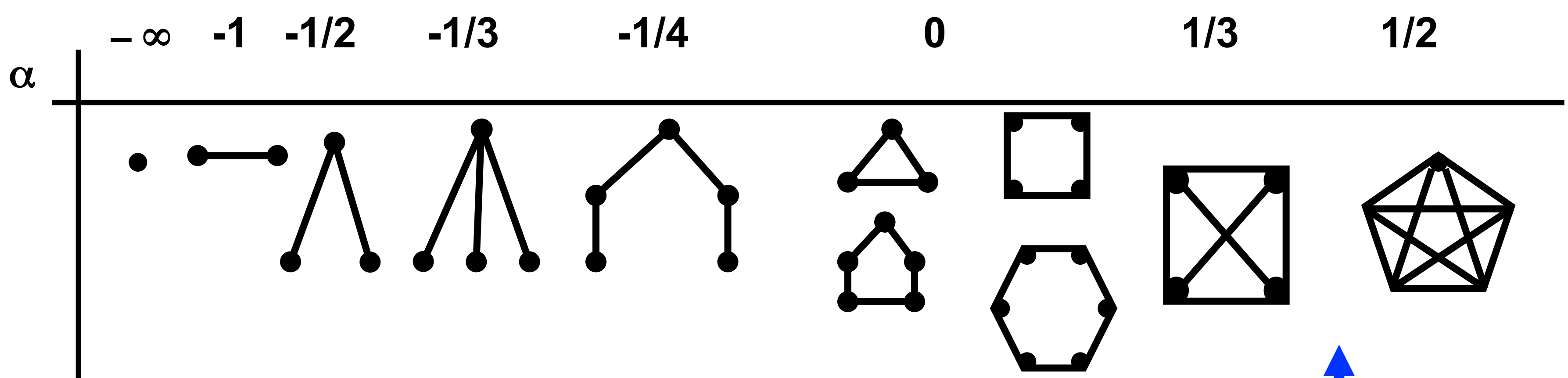
This network contains
- ***4 cliques of size 3 (triangles)***
- ***and 1 clique of size 4***

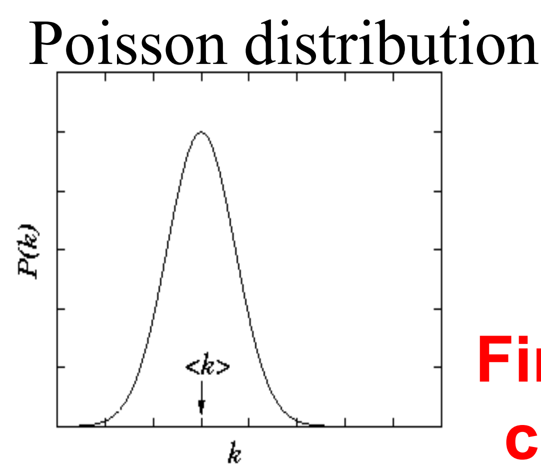
Small subgraphs appear abruptly when we increase the average number of links in the random graphs

$$c = \frac{L}{N} \approx N^{\alpha}$$

Subgraph thresholds

$$c \approx N^{\alpha}$$





Finite average connectivity

Diverging average connectivity

Small loops in the Internet at the AS level

The number of loops

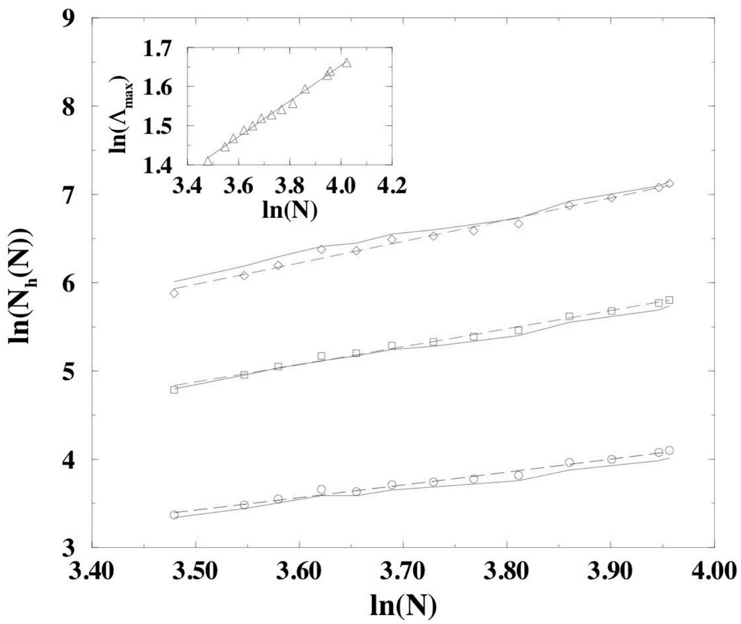
of size L= 3,4,5

grows with the network size N as

$$\aleph_L \propto N^{\xi(L)}$$

In Poisson networks instead they are a fixed number

Independent on N

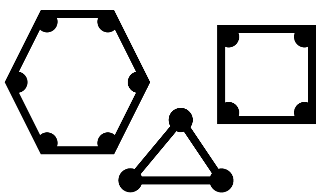
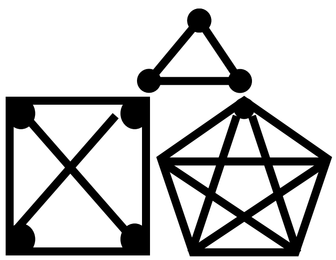


G. Bianconi et al.PRE 2005

Small subgraphs in SF networks

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

$\langle k \rangle$ *finite* $\qquad\qquad$ $\langle k \rangle \to \infty$

γ	∞ – 3	3 – 2	2 – 1
 Small loops Of length L	$\langle \aleph_L \rangle$ *finite*	$\langle \aleph_L \rangle \approx \frac{1}{2L} N^{(3-\gamma)L/2}$	$\langle \aleph_L \rangle \approx \frac{1}{2L} N^{(3-\gamma)L/\gamma}$
 Cliques of Size C	$C_{max}=3$	$\langle \aleph_c \rangle \approx \left(\frac{N^{(3-\gamma)/2}}{c\,(c-\gamma)} \right)^c$	$\langle \aleph_c \rangle \approx \left(\frac{N^{(1/\gamma)}}{c\,(c-\gamma)} \right)^c$
		$c_{max} \xrightarrow{N \to \infty} \infty$	$c_{max} \xrightarrow{N \to \infty} \infty$

G. Bianconi and M. Marsili (2004), (2005)

Why working on networks?

Because

NETWORKS

encode for the
ORGANIZATION,
FUNCTION,
ROBUSTENES
AND DYNAMICAL BEHAVIOR
of the entire complex system