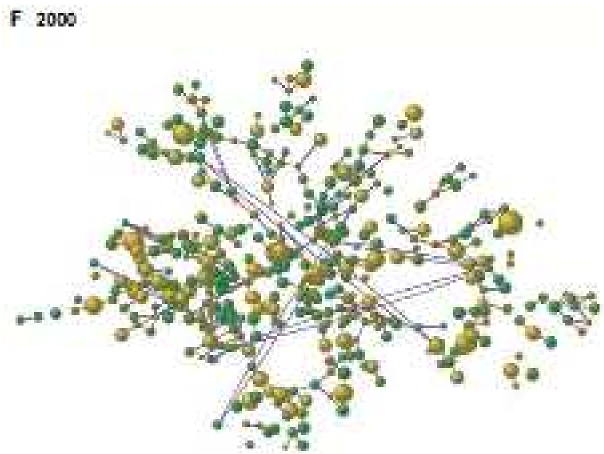
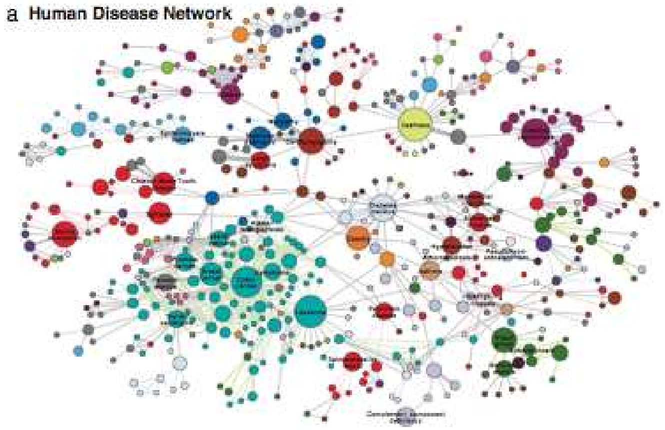




Winter School in Network Theory and Applications

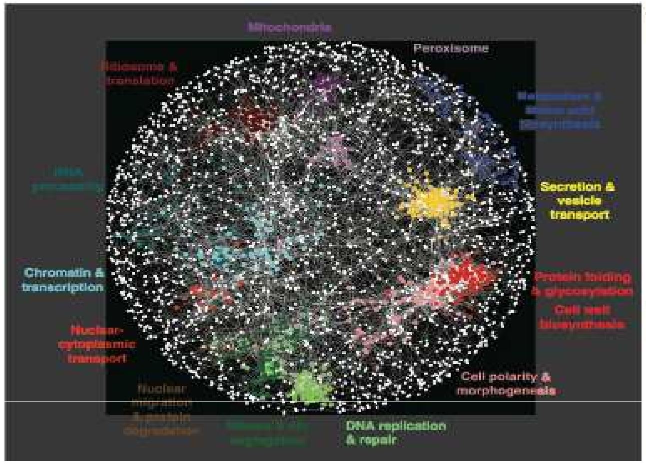
Warwick, Jan 5-9 2011

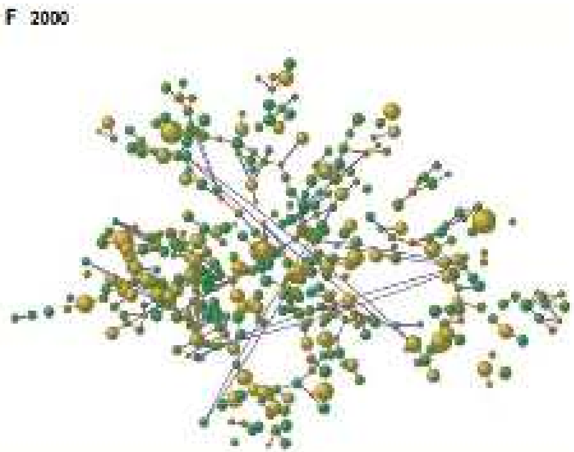
Evolution of networks

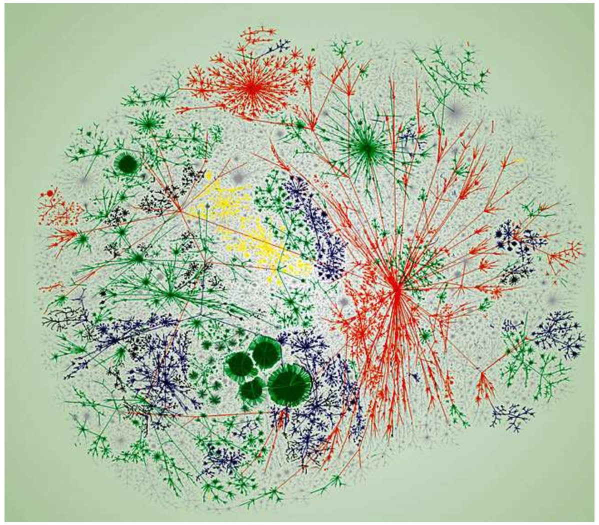
Ginestra Bianconi

Physics Department,Northeastern University, Boston,USA

Complex networks



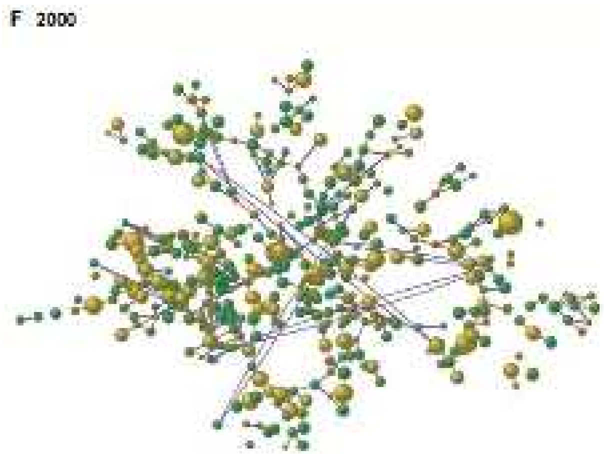




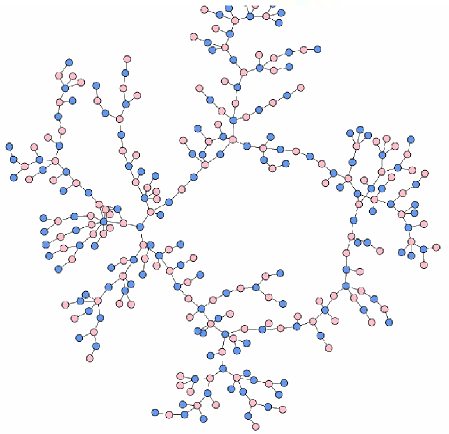
describe

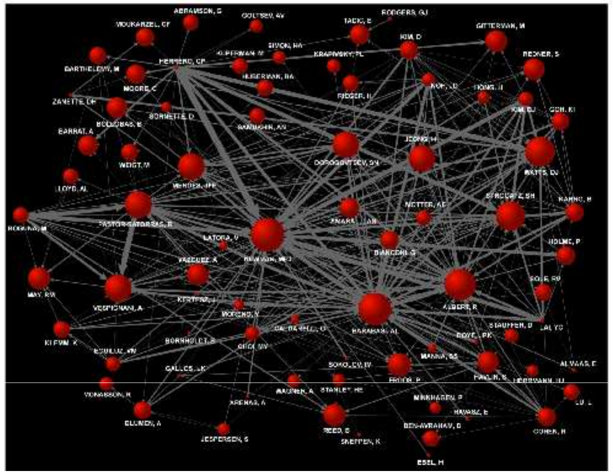
the underlying structure of interacting complex

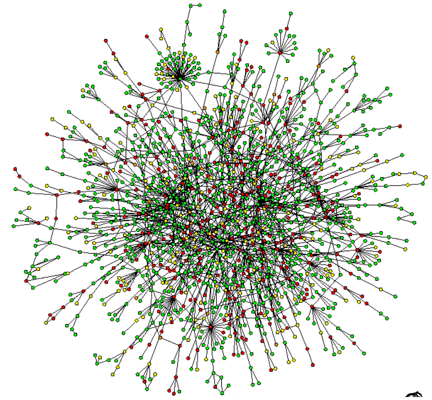
Biological, Social and Technological systems.

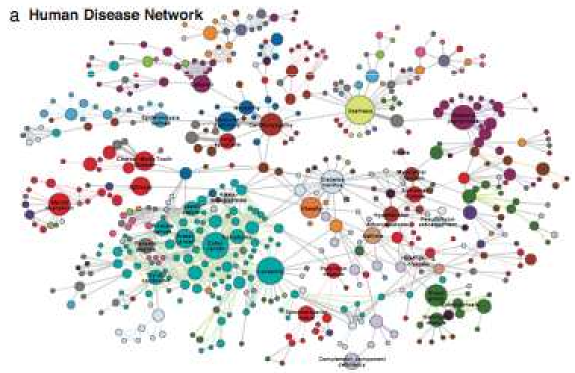


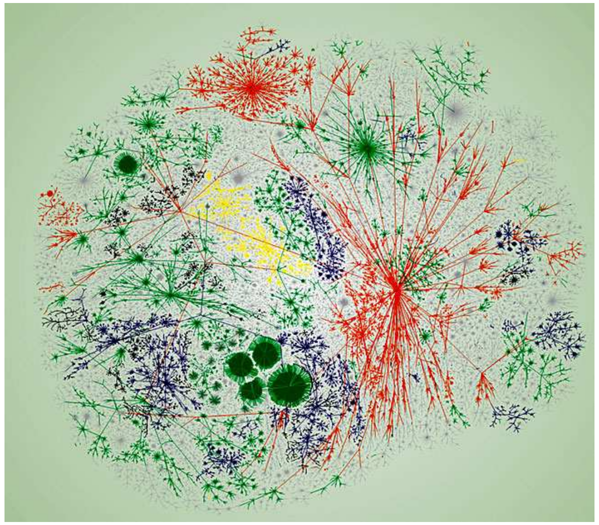
Identify your network !

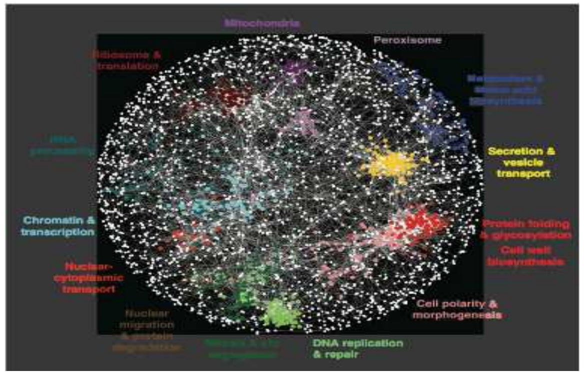












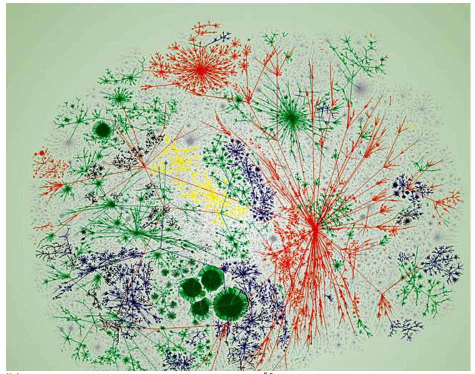
Why working on networks?

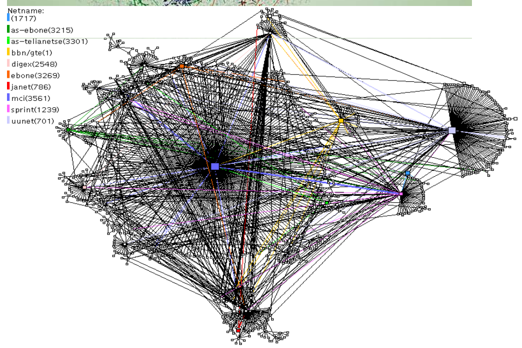
Because

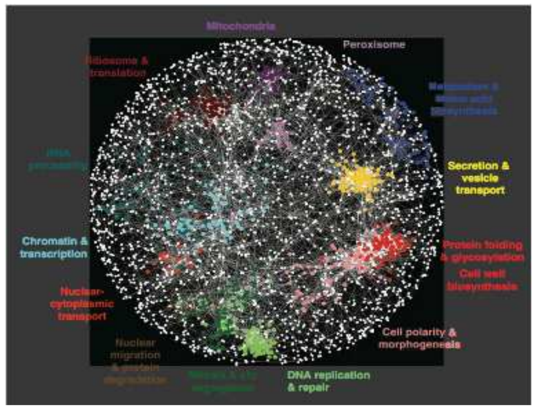
NETWORKS

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…)

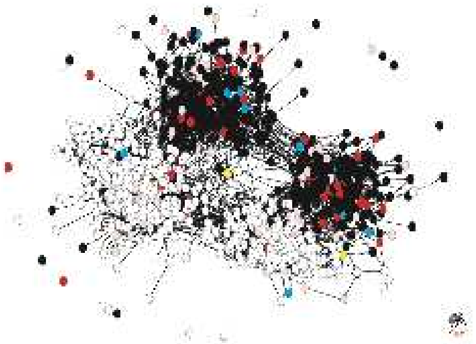
Types of networks

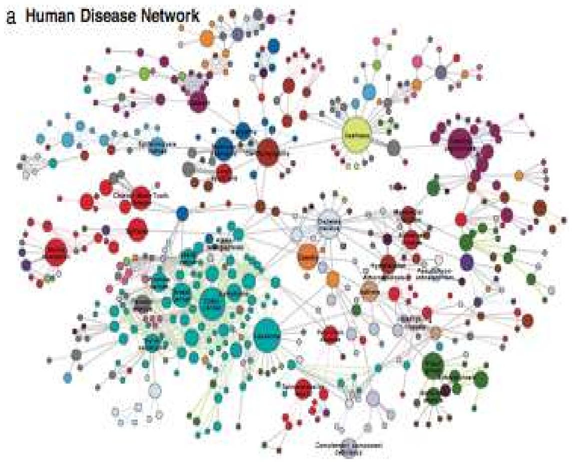
- **Weighted** The links are associated to a real number indicating their weight

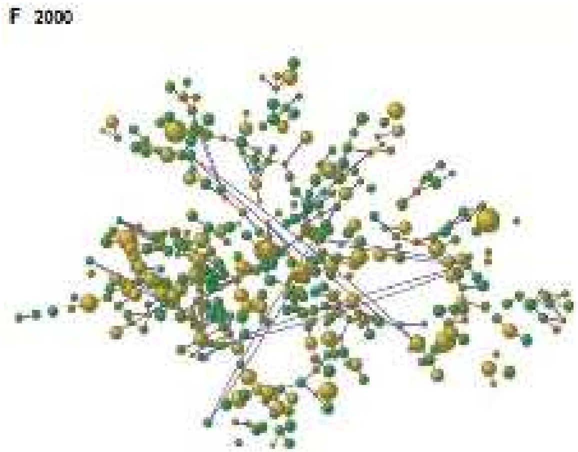
(airport networks, phone-call networks…)

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

(social networks, diseasome, ect..)







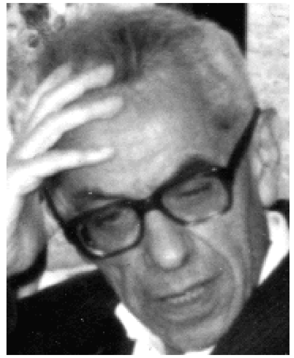
Random graphs

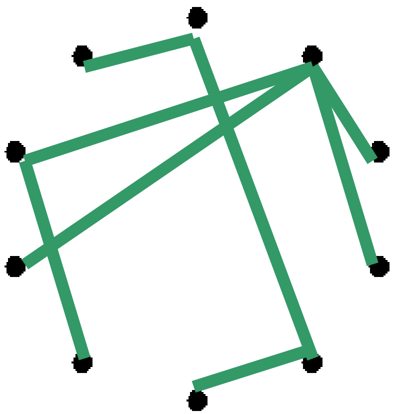
G(N,L) ensemble

Graphs with exactly N nodes and L links

G(N,p) ensemble

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



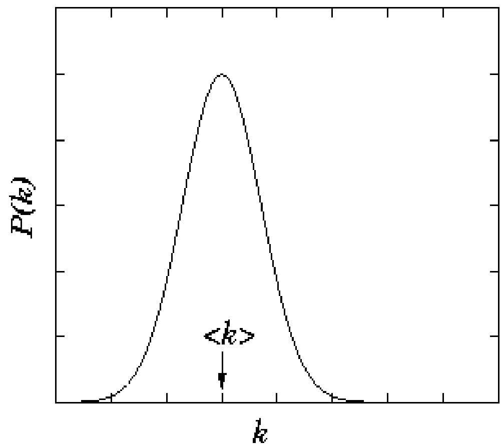


Binomial distribution

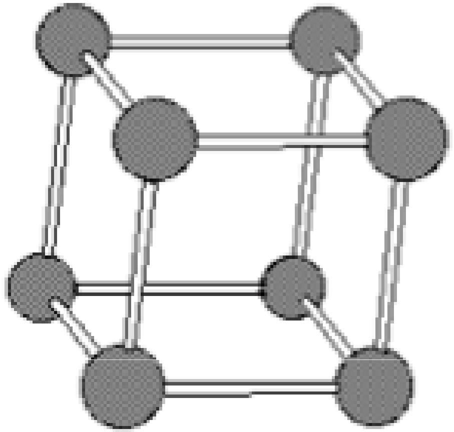
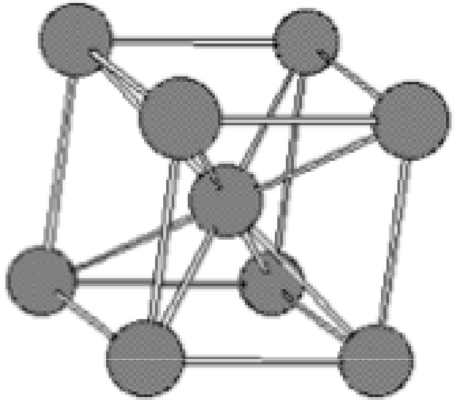
$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

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

Poisson distribution



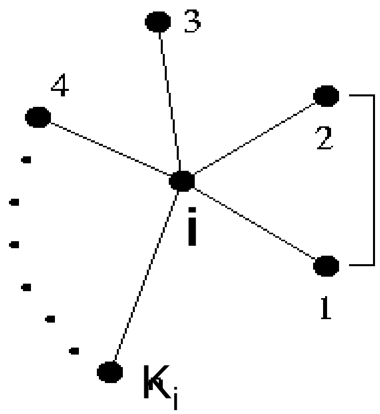
Regular lattices

d dimensions

- **Large average distance** $L \approx N^{1/d}$
- **Significant local interactions**

Universalities:Small world



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

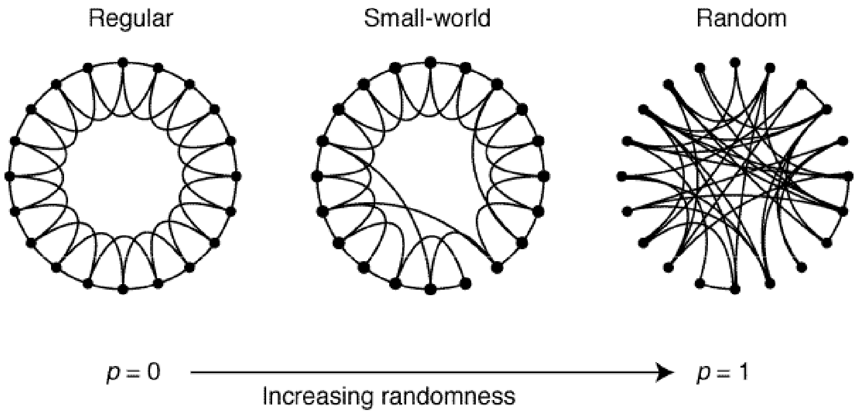
Networks are clustered (large average C_i,i.e.C)

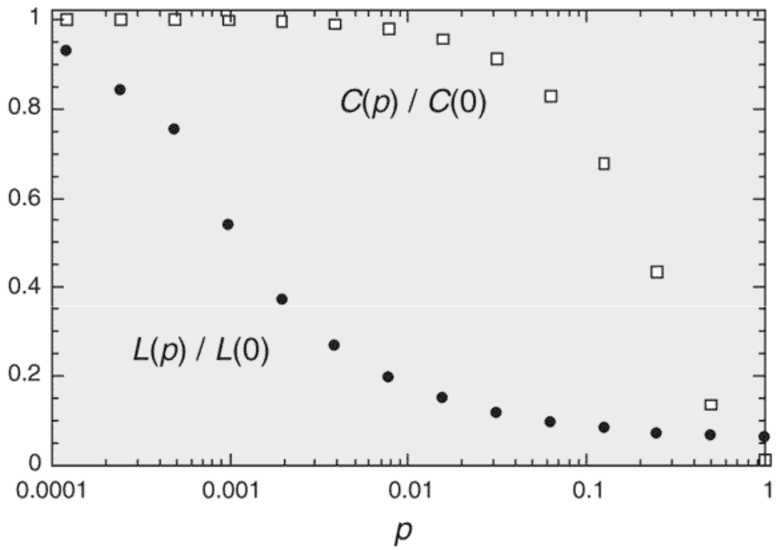
but have a small characteristic path length (small L).

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)

Watts and Strogatz small world model





Watts & Strogatz (1998)

Variations and characterizations

Amaral & Barthélemy (1999)

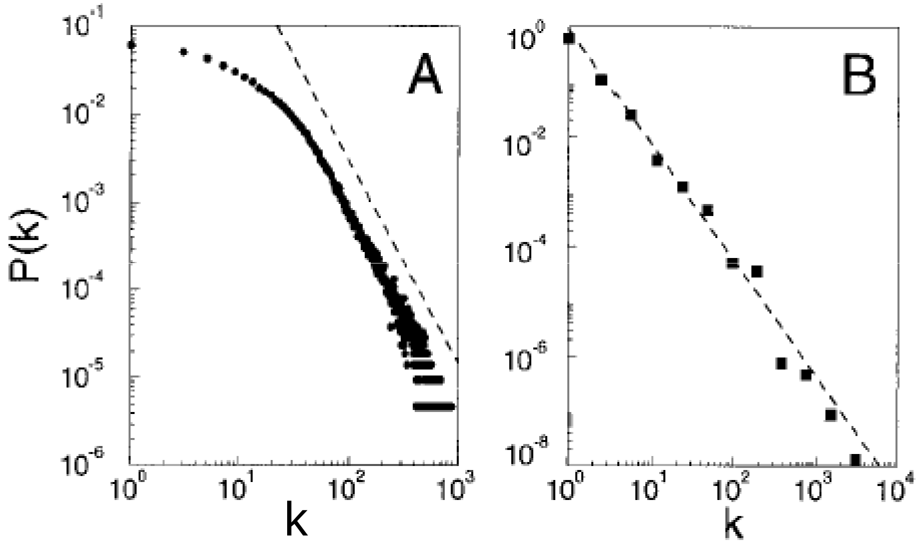
Newman & Watts, (1999)

Barrat & Weigt, (2000)

There is a wide range of values of p in which high clustering coefficient coexist with small average distance

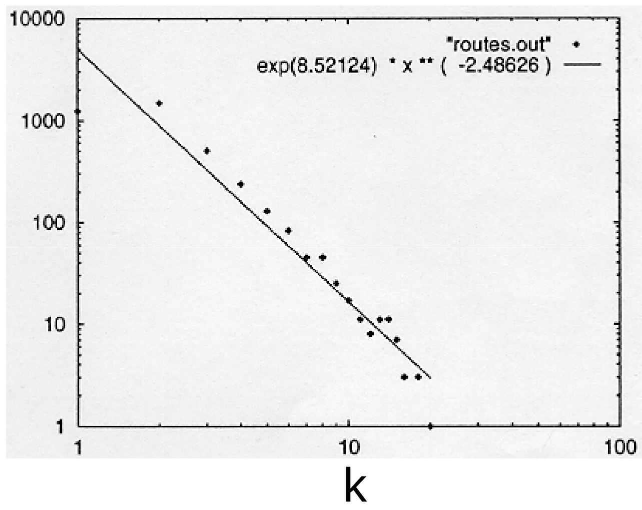
Universalities:
Scale-free degree distribution

Actor networks WWW



Barabasi-Albert 1999

Internet



Faloutsos et al. 1999

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

$$\langle k \rangle \quad finite$$

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

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

Motivation for BA model

1) The network grow

Networks continuously expand by the addition of new nodes

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

2) The attachment is **not** uniform (preferential attachment).

A node is linked with higher probability to a node that already has a large number of links.

Ex: WWW : new documents link to well known sites (CNN, YAHOO, NewYork Times, etc)
Citation : well cited papers are more likely to be cited again

BA model

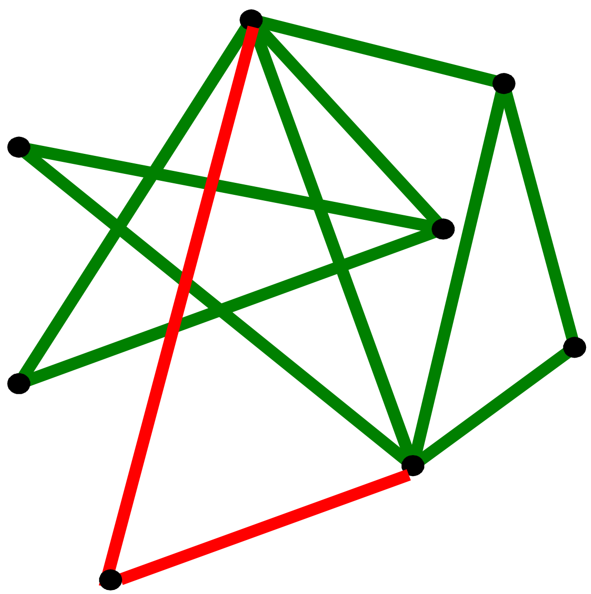
(1) GROWTH :

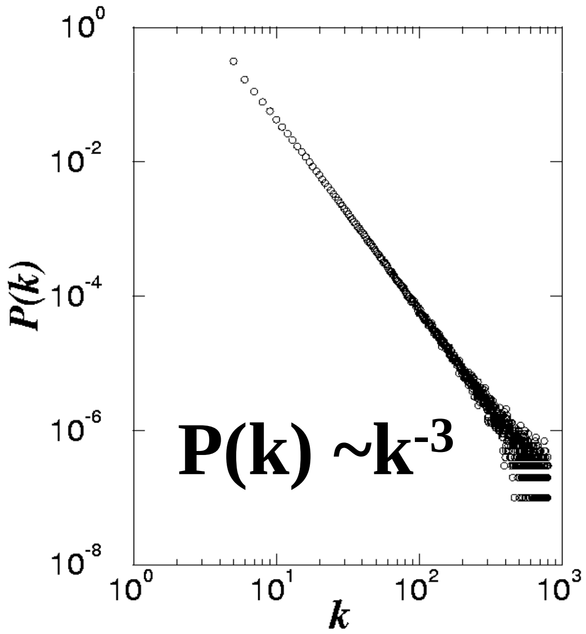
At every timestep we add a new node with *m* edges (connected to the nodes already present in the system).

(2) PREFERENTIAL ATTACHMENT :

The probability **Π** that a new node will be connected to node *i* depends on the connectivity k_i of that node

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





Barabási et al. Science (1999)

Result of the BA scale-free model

- The connectivity of each node increases in time as a power-law with exponent 1/2:

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

- The probability that a node has *k* links follow a power-law with exponent $\gamma = 3$:

$$P(k) = 2m^2 \frac{1}{k^3}$$

Initial attractiveness

The initial attractiveness can change the value of the power-law exponent γ

A preferential attachment with initial attractiveness A yields

$$\Pi_i \propto k_i + A$$

$$k_i \propto \left(\frac{t}{t_i}\right)^{\beta}$$

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

$\Rightarrow$

$$\gamma \in (2, \infty)$$

$$\beta(\gamma - 1) = 1$$

Dorogovtsev et al. 2000

Non-linear preferential attachment

The probability to link to a node is given by

$$\Pi_i \propto k_i^{\alpha}$$

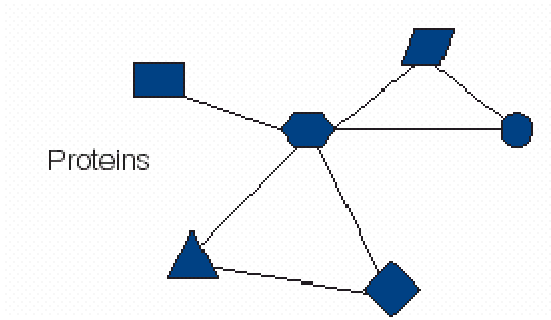
- $\alpha<1$ **Absence of power-law degree distribution**
- $\alpha=1$ **Power-law degree distribution**
- $\alpha>1$ **Gelation phenomena**

  The oldest node acquire most of the links

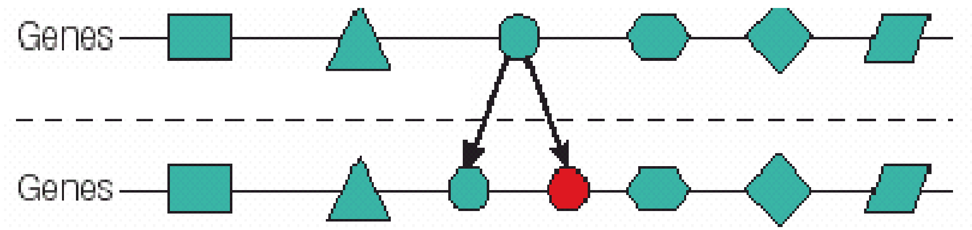
  First-mover-advantage

Krapivski et al 2000

Gene duplication model

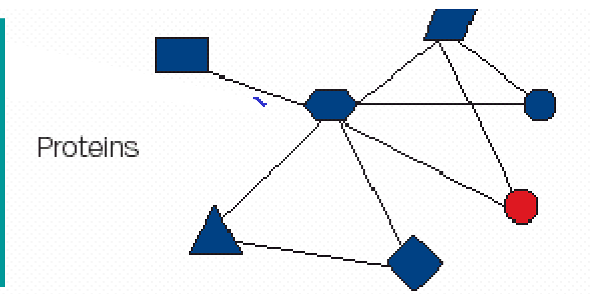


- **Duplication of a gene adds a node.**



- **New proteins will be preferentially connected to high connectivity.**

Effective preferential attachment



A. Vazquez et al. (2003).

Other variations

➢**Scale-free networks with high-clustering coefficient**

Dorogovtsev et al. 2001

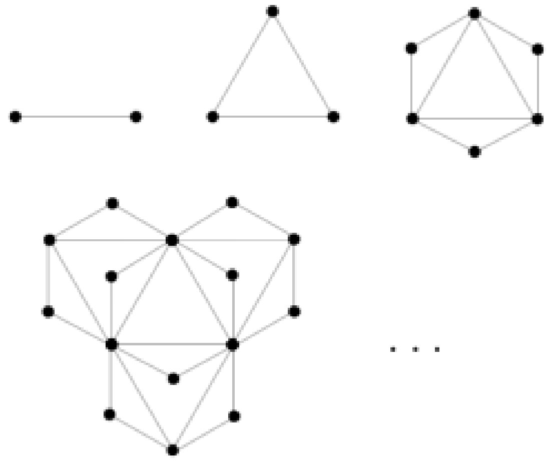
Eguiluz & Klemm 2002

➢**Aging of the nodes**

Dorogovstev & Medes 2000

➢**Pseudofractal scale-free network**

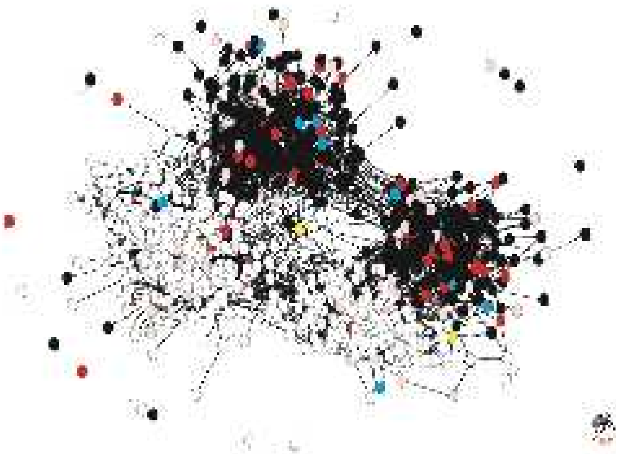
Dorogovtsev et al 2002

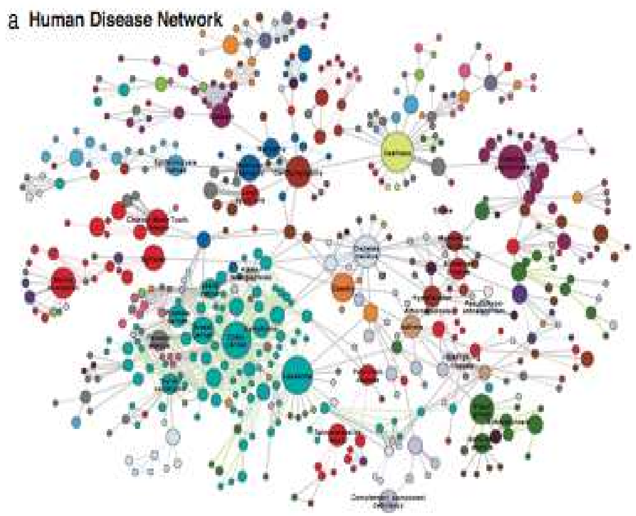


Features of the nodes

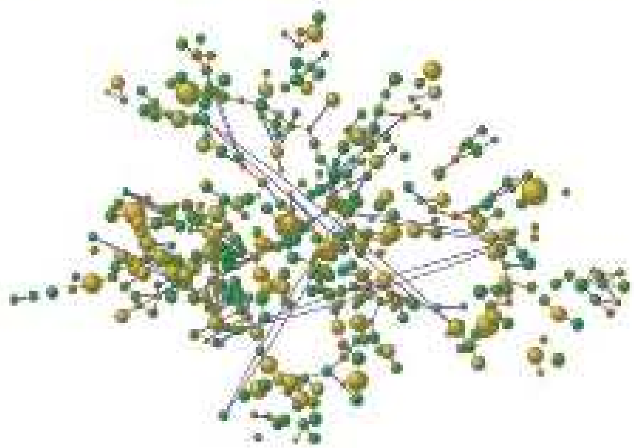
In complex networks
nodes are generally heterogeneous
and they are characterized by specific features

- *Social networks*: age, gender, type of jobs, drinking and smoking habits,
- *Internet:* position of routers in geographical space, …
- *Ecological networks*: Trophic levels, metabolic rate, philogenetic distance
- *Protein interaction networks:* localization of the protein inside the cell, protein concentration





F 2000



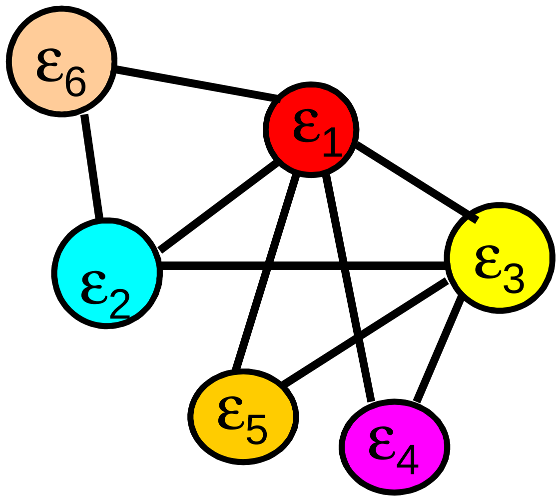
Fitness the nodes

Not all the nodes are the same!

Let assign to each node an **energy** ε *and a* **fitness** $\eta = e^{-\beta\varepsilon}$

that describes the characteristics of a node to attract new links

In the limit $\beta=0$ *all the nodes have same fitness*



The fitness model

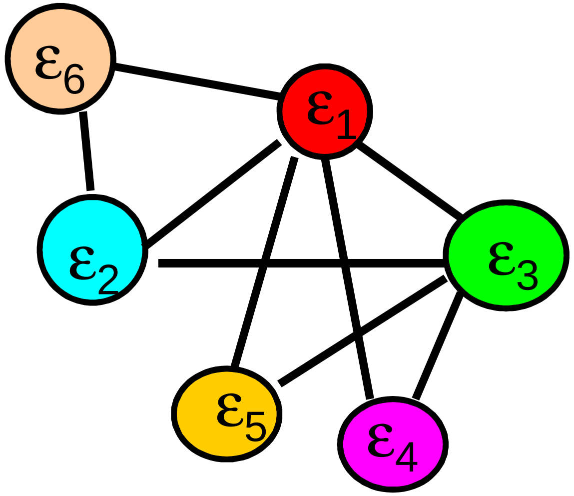
Growth:

–At each time a new node and *m* links are added to the network.

–To each node *i* we assign a energy ε_i from a $p(\varepsilon)$ distribution

Generalized preferential attachment:

–Each node connects to the rest of the network by *m* links attached preferentially to well connected, low energy nodes.



$$\Pi_i \propto e^{-\beta \varepsilon_i} k_i$$

Results of the model

- **Power-law degree distribution**

$$P(k) \approx k^{-\gamma} \quad 2 < \gamma < 3$$

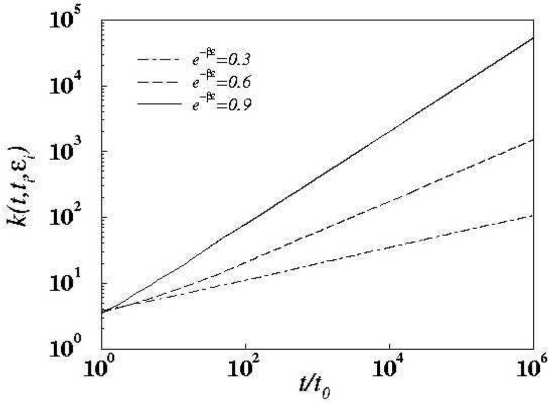
- **Fit-get-rich mechanism**

$$k_\eta(t) = m\left(\frac{t}{t_i}\right)^{\eta_i / C}$$

Fit-get rich mechanism

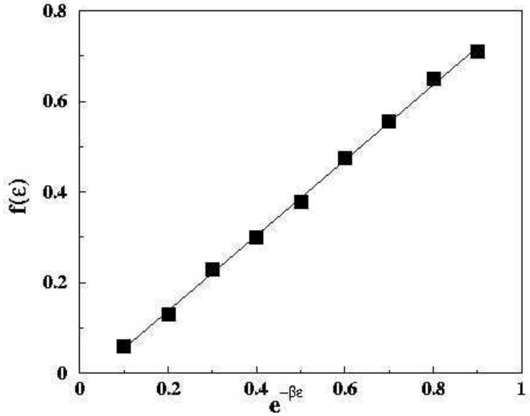
The nodes with higher fitness increases the connectivity faster

$$k_i = \left(\frac{t}{t_i}\right)^{f(\varepsilon_i)} \qquad f(\varepsilon) = e^{-\beta(\varepsilon-\mu)}$$



μ satisfies the condition

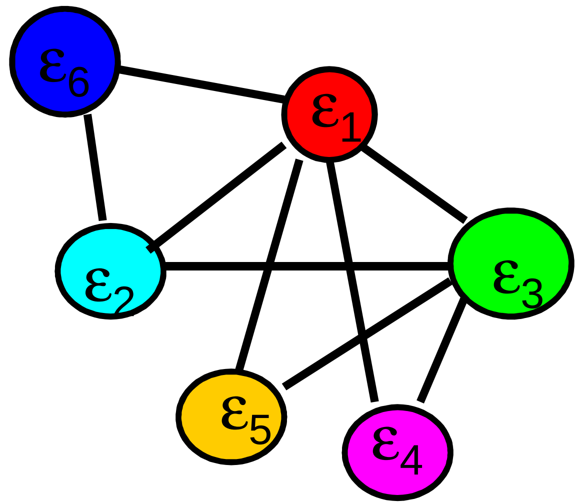
$$1 = \int d\varepsilon \; p(\varepsilon) \frac{1}{e^{\beta(\varepsilon-\mu)} - 1}.$$



Mapping to a Bose gas

We can map the fitness model to a Bose gas with

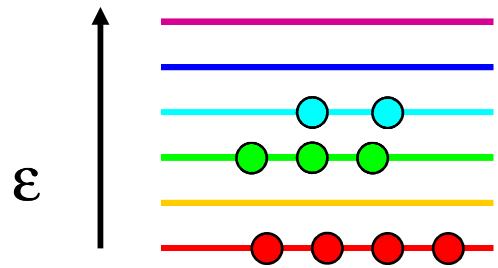
- **density of states** ***p(ε);***
- **specific volume** ***v=1;***
- **temperature** ***T=1/β*.**



Network

In this mapping,

- **each node of energy ε corresponds to an energy level of the Bose gas**
- **while each link pointing to a node of energy ε, corresponds to an occupation of that energy level.**

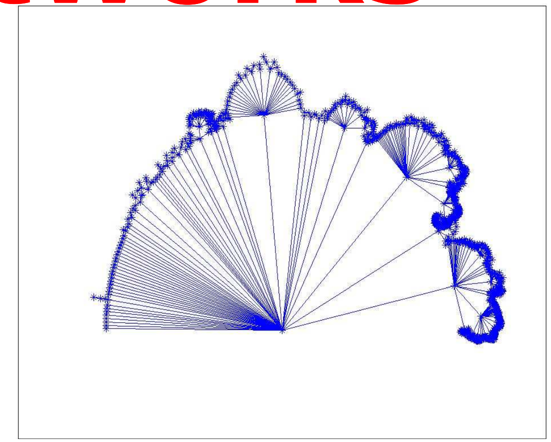


Energy diagram

G. Bianconi, A.-L. Barabási 2001

Bose-Einstein condensation in trees scale-free networks

In the network there is a critical temperature T_c such that



•for $T>T_c$ the network is in the

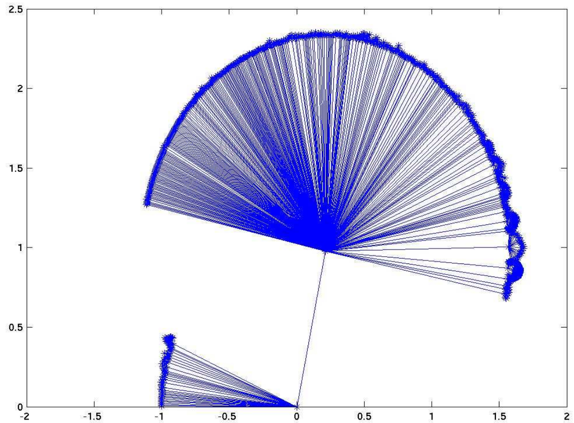
fit-get-rich phase

•for $T<T_c$ the network is in the

winner-takes-all

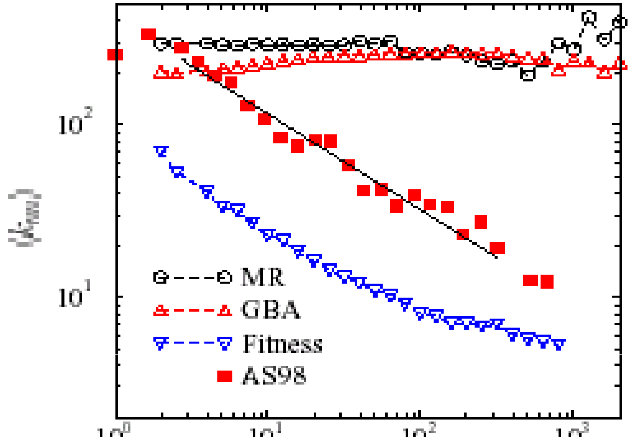
or Bose-condensate

phase

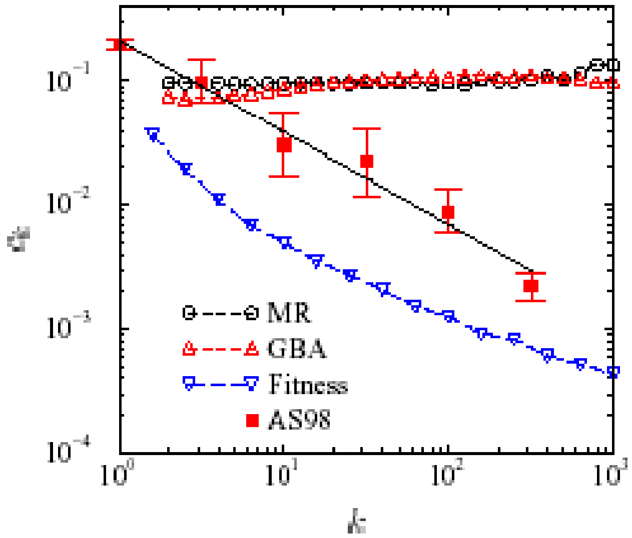


Correlations in the Internet and the fitness model

➢ k_{nn} *(k)* **mean value of the connectivity of neighbors sites of a node with connectivity k**

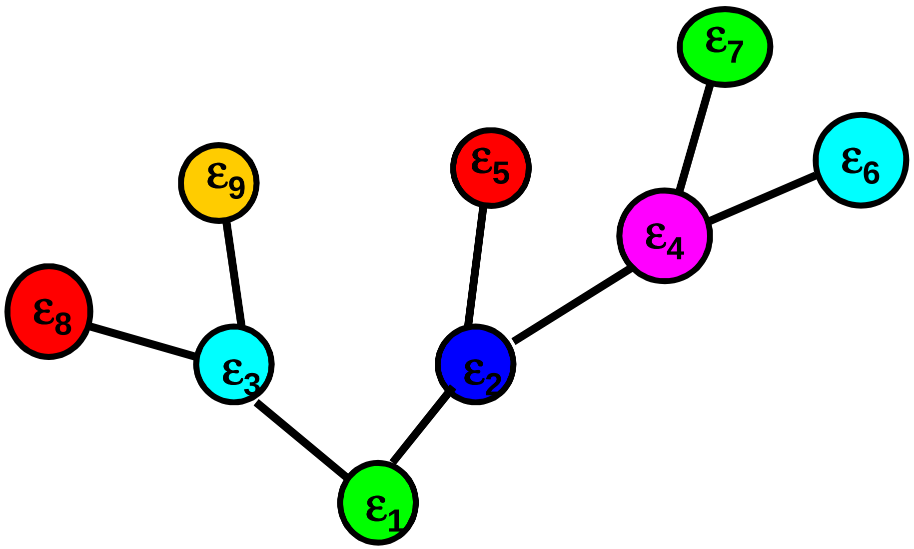


➢ ***C(k)*** **average clustering coefficient of nodes with connectivity k.**

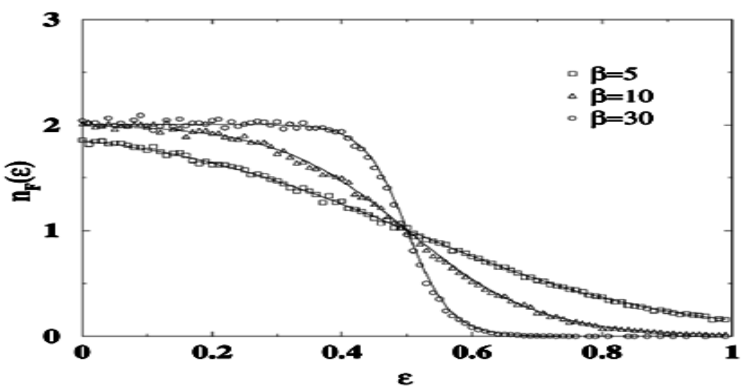


Vazquez et al. 2002

Growing Cayley-tree



Nodes at the interface



- Each node is either at the interface n_i=1 or in the bulk n_i=0
- At each time step a node at the interface is attached to *m* new nodes with energies ε from a *p(ε)* distribution.
- High energy nodes at the interface are more likely to grow.
- The probability that a node *i grows is given by*

$$\Pi_i \propto e^{\beta \varepsilon_i} n_i$$

G. Bianconi 2002

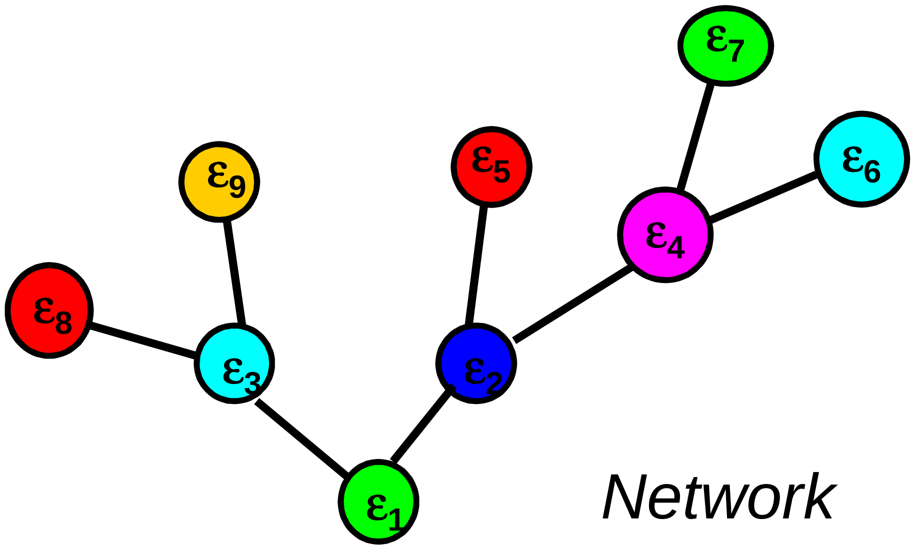
Mapping to a Fermi gas

The growing Cayley tree network can be mapped into a Fermi gas

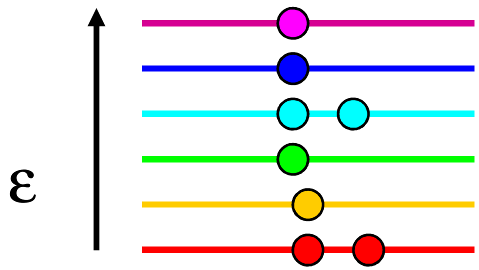
- **with density of states** ***p(ε);***
- **temperature** ***T=1/β*;**
- **specific volume** ***v=1-1/m*.**

➢In the mapping the **nodes** corresponds to the **energy levels**

➢ the **nodes at the interface** to the **occupied energy levels**



Network



Energy diagram

Weight distribution and weight-degree correlations

The weights of a network might correlate with the degree of the nodes which are linked by them:

- Strength S_k
  - Is a measure of the inhomogeneity of the distribution of the weights among the nodes of the network.

- *Disparity* Y^2_k of the weights of the links attached to nodes of degree k,
  - Is a measure of the inhomogeneity of the weights ending at one node.

Strength of the node and weight-degree correlations

The strength of the nodes is the average weight of links connecting that node

$$s(k,i) = \sum_j w_{i,j} \approx \begin{cases} k \\ k^{1+\eta} \end{cases}$$

All weight distribute themselves evenly between the nodes

Weight-degree correlations:
Highly weighted links are connected to highly connected nodes

The inhomogeneity of the weights ending at one node

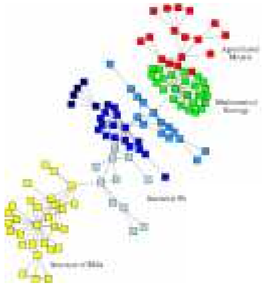
$$Y^2(k,i) = \sum_j \left(\frac{w_{i,j}}{s_i} \right)^2 \approx \begin{cases} 1/k & \text{All weights contribute equall} \\ 1 & \text{There is a dominating link} \end{cases}$$

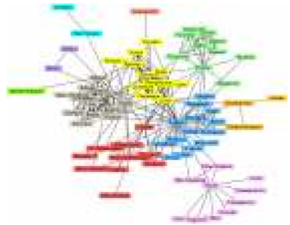
In many networks one has

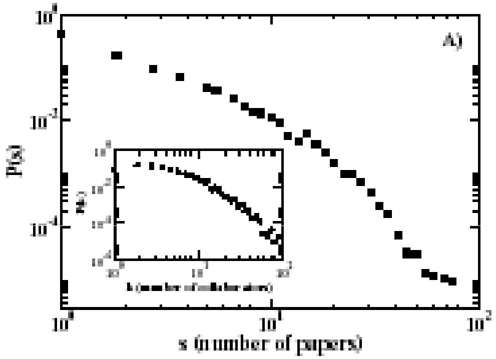
$$Y^2(k) = 1/k^{\alpha} \quad \text{with } \alpha \in (0,1)$$

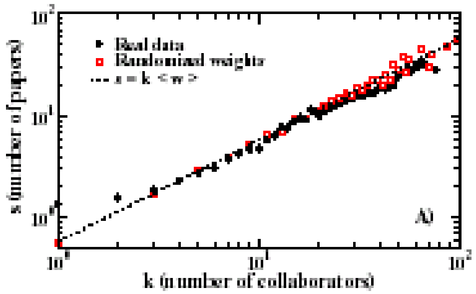
indicating a hierarchy of weights

Coauthorship networks



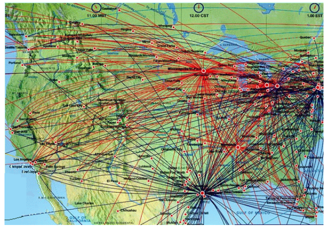






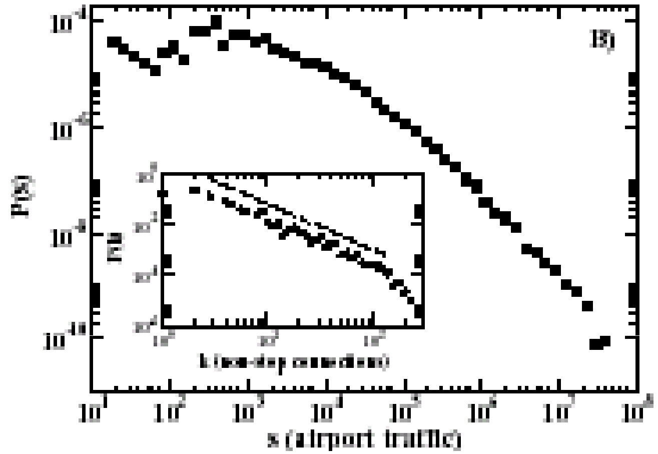
- The coauthorship network is a scale-free network
- With power-law strength distribution and
- Dependence of strength versus k growing linearly.

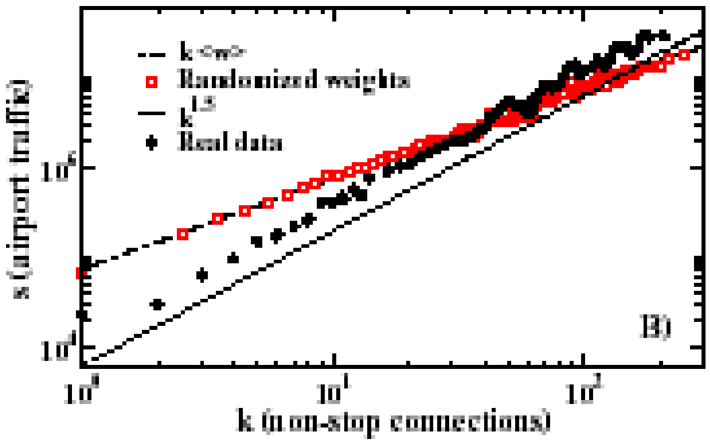
A.Barrat et al. PNAS 2004



Airport network

- **The airport network is a scale-free network**
- **With power-law strength distribution and**
- **Dependence of strength versus k growing faster than linear.**



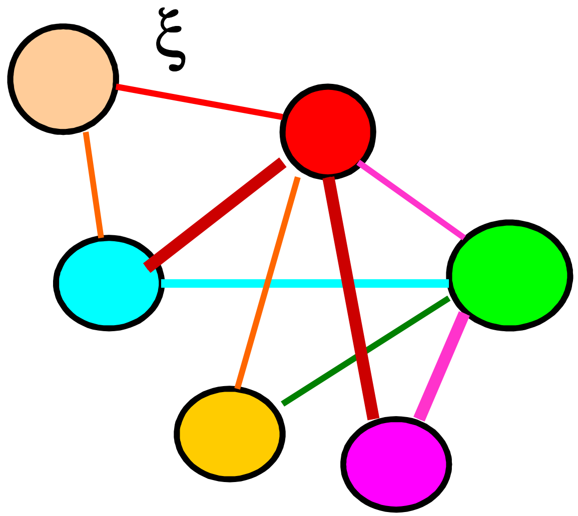


A.Barrat et al. PNAS 2004

Fitness model with reinforcement of the links

- A fitness $\xi_{i,j}$ is assigned to each link
- At each time m' links chosen are reinforced.
- They are chosen preferentially in within high fitness and high weights links with probability

$$\Pi_{i,j} \propto \xi_{ij} w_{ij}$$



Different cases

1. Uncorrelated fitness of the links and of the nodes

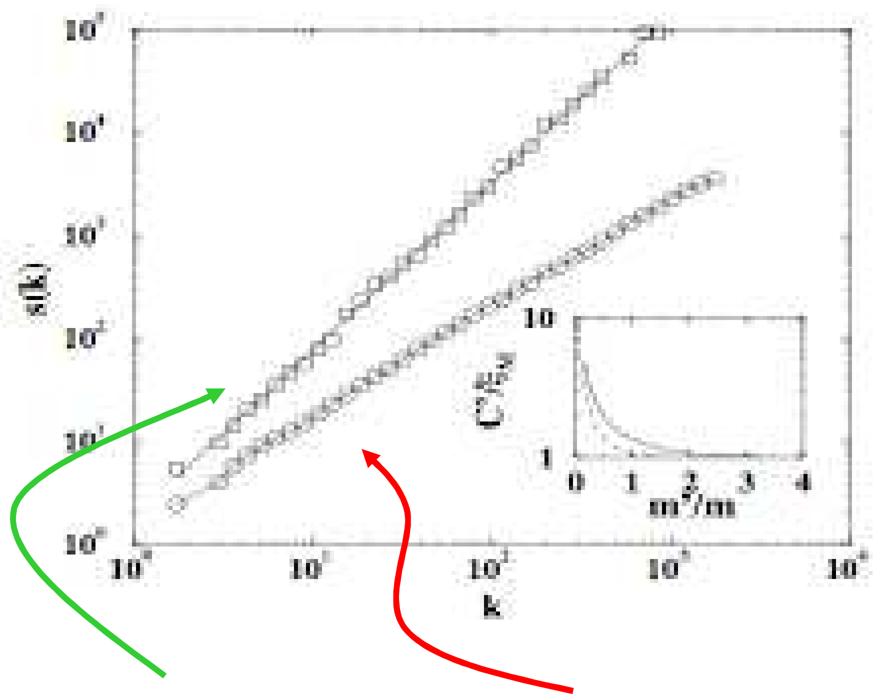
$$\xi_{i,j} \quad uncorrelated\ with\ \eta_i$$

2. Correlated fitness of the links and of the nodes

-*Simple cases*

$\xi_{i,j} = \eta_i + \eta_j$
$\xi_{i,j} = \eta_i \eta_j$

Additive case



(m,m')=(1,10)   (m,m')=(2,1)

The fitness of the links is the sum of the fitness of the nodes

$$\xi_{i,j} = \eta_i + \eta_j$$

As a function of m'/m

The strength s versus connectivity k can grow

linearly

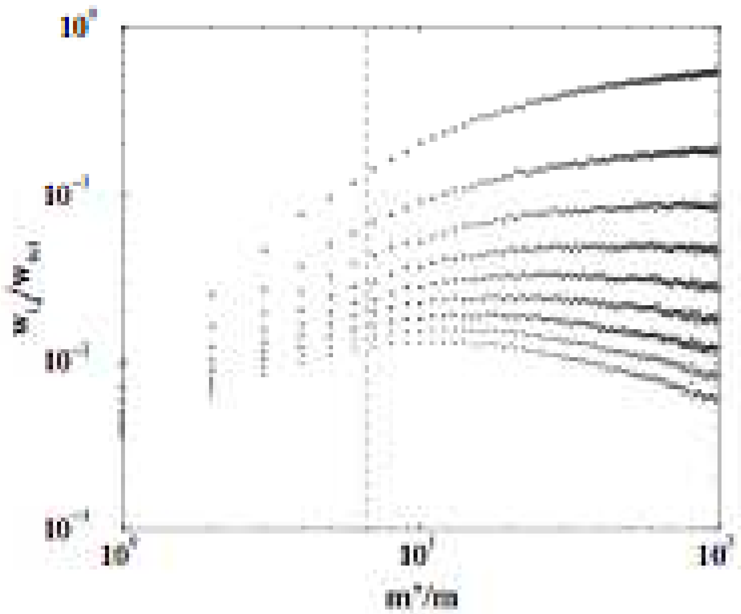
faster than linearly

Condensation of the links

When instead is not anymore possible to solve the equation of C' as a function of m'/m

One can predict the condensation of the weights of a finite fraction of the links

As a function of m'/m one observe that this is in fact encountered in the model for m'/m>r_0=3.28



Why this universality?

- Growing networks:
  - **Preferential attachment**

Barabasi & Albert 1999,
Dorogovtsev Mendes 2000, Bianconi & Barabasi 2001, etc.

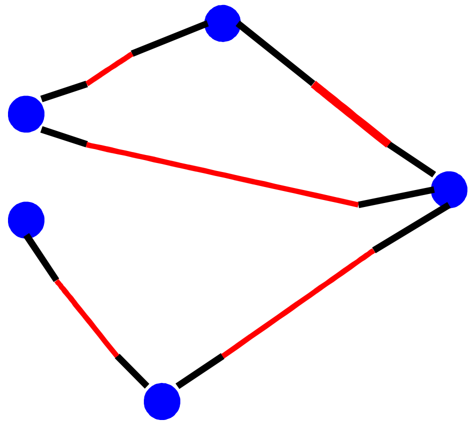
- Static networks:
  - **Hidden variables mechanism**

Chung & Lu 2002, Caldarelli et al. 2002,
Park & Newman 2003

Molloy Reed configuration model

$$P(G) = \frac{1}{\Sigma_1} \prod_i \delta(k_i - \sum_j a_{ij})$$

Networks with given degree distribution



- Assign to each node a degree from the given degree distribution
- Check that the sum of stubs is even
- Link the stubs randomly
- If tadpoles or double links are generated repeat the construction

Molloy & Reed 1995

Caldarelli et al. hidden variable model

- Every nodes is associated with an hidden variable x_i
- The each pair of nodes are linked with probability

$$p_{ij} = f(x_i, x_j)$$

$$k(x) = N \int dy\, \rho(y)\, f(x, y)$$

Caldarelli et al. 2002
Soderberg 2002
Boguna & Pastor-Satorras 2003

Park & Newman
Hidden variables model

$$H = \sum_i \theta_i k_i = \sum_{i,j} (\theta_i + \theta_j) a_{i,j}$$

The system is defined through an Hamiltonian

$$p_{ij} = \frac{e^{\theta_i + \theta_j}}{1 + e^{\theta_i + \theta_j}}$$

p_{ij} is the probability of a link

J. Park and M. E. J. Newman (2004).

- **The "hidden variables" θ_i are quenched and distributed through the nodes with probability $\rho(\theta)$.**

- **There is a one-to-one correspondence between θ and the average connectivity of a node**

$$\langle k_i \rangle = (N-1) \int d\theta' \rho(\theta') \frac{1}{e^{\theta_i + \theta'} + 1}$$

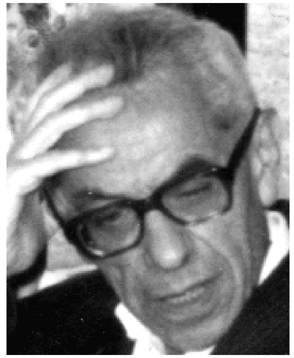
Random graphs

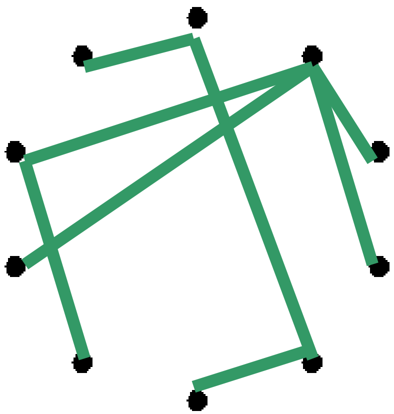
G(N,L) ensemble

Graphs with exactly N nodes and L links

G(N,p) ensemble

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



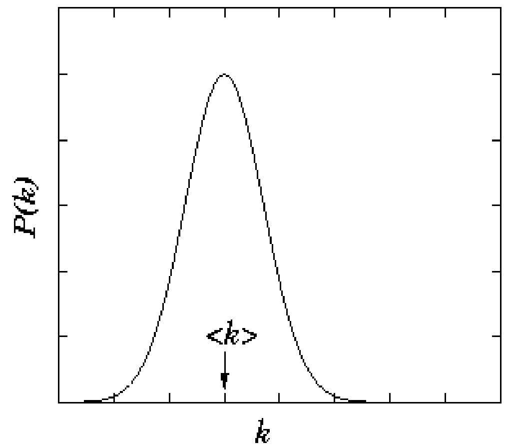


Binomial distribution

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

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

Poisson distribution



Statistical mechanics and random graphs

Statistical mechanics		Random graphs	
Microcanonical Ensemble	**Configurations with fixed energy E**	**G(N,L) Ensemble**	**Graphs with fixed # of links L**
Canonical Ensemble	**Configurations with fixed average energy <E>**	**G(N,p) Ensemble**	**Graphs with fixed average # of links <L>**

Gibbs entropy and entropy of the G(N,L) random graph

Statistical mechanics **Microcanonical ensemble**	**Random graphs** **G(N,L) ensemble**
$$S = k\,log(\Omega(E))$$	$$\Sigma = \frac{1}{N} log(Z)$$
Gibbs Entropy	**Entropy per node of the G(N,L) ensemble**
$$\Omega(E)$$	$$Z = \binom{N(N-1)/2}{L}$$
Total number of microscopic configurations with energy E	Total number of graphs in the G(N,L) ensembles

Shannon entropy and entropy of the G(N,p) random graph

Statistical mechanics Canonical ensemble	Random graphs G(N,p) ensemble
$S=-\sum_{E} p(E)\ln p(E)$	$S=-\frac{1}{N}\sum_{\{a_{ij}\}} p(a_{ij})\ln p(a_{ij})$
Shannon Entropy	**Entropy per node of the G(N,p) ensemble**
$p(E)=\frac{1}{Z}e^{-bE}$	$\Sigma=-\frac{c}{2}\ln c+\frac{N}{2}\ln N-\frac{(N-c)}{2}\ln(N-c)$
Typical number of microscopic configurations with temperature β	Total number of typical graphs the G(N,p) ensembles

Complexity of a real network

Hypothesis:

Real networks are

single instances

of an ensemble of possible networks

which would equally well perform the function of the existing network

The "complexity" of a real network

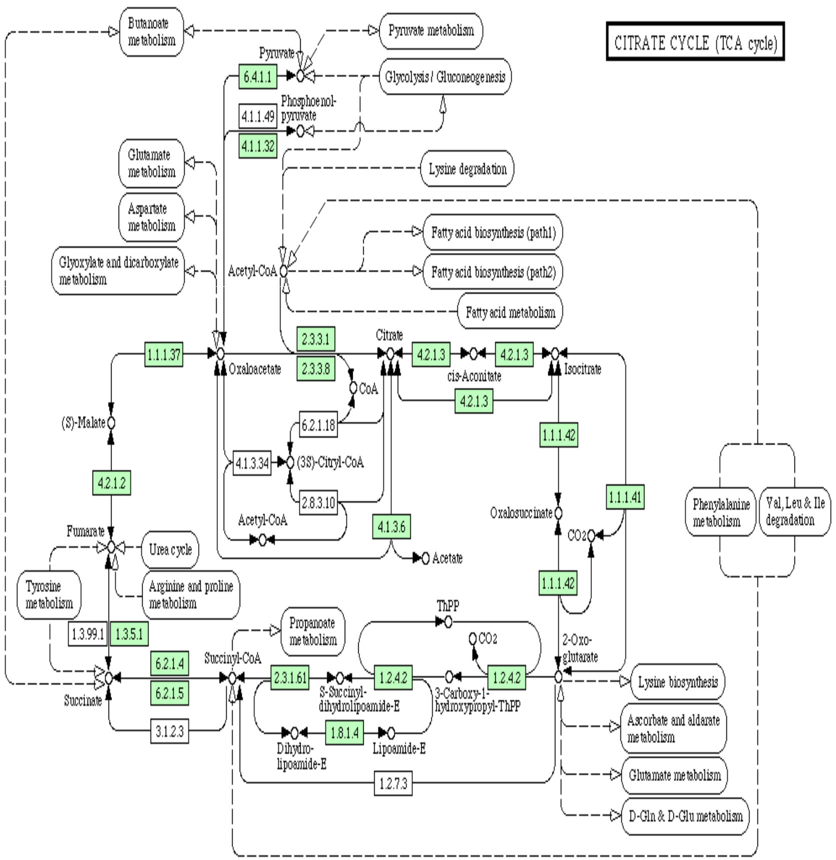
is a decreasing function

of the entropy of this ensemble

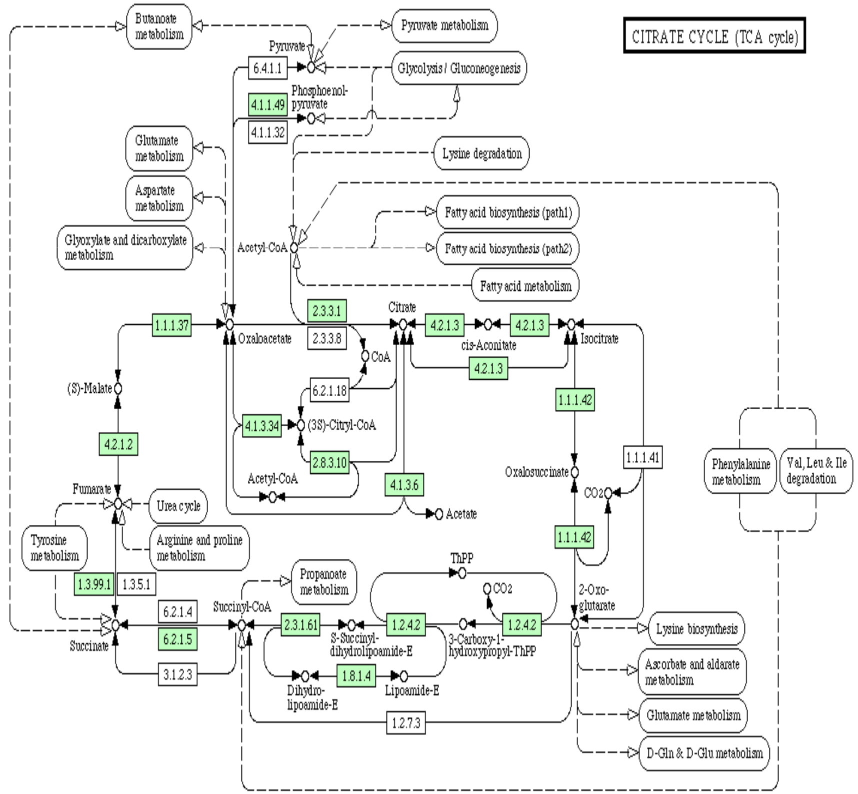
G. Bianconi EPL (2008)

Citrate cycle:highly preserved

Escherichia coli



Homo Sapiens



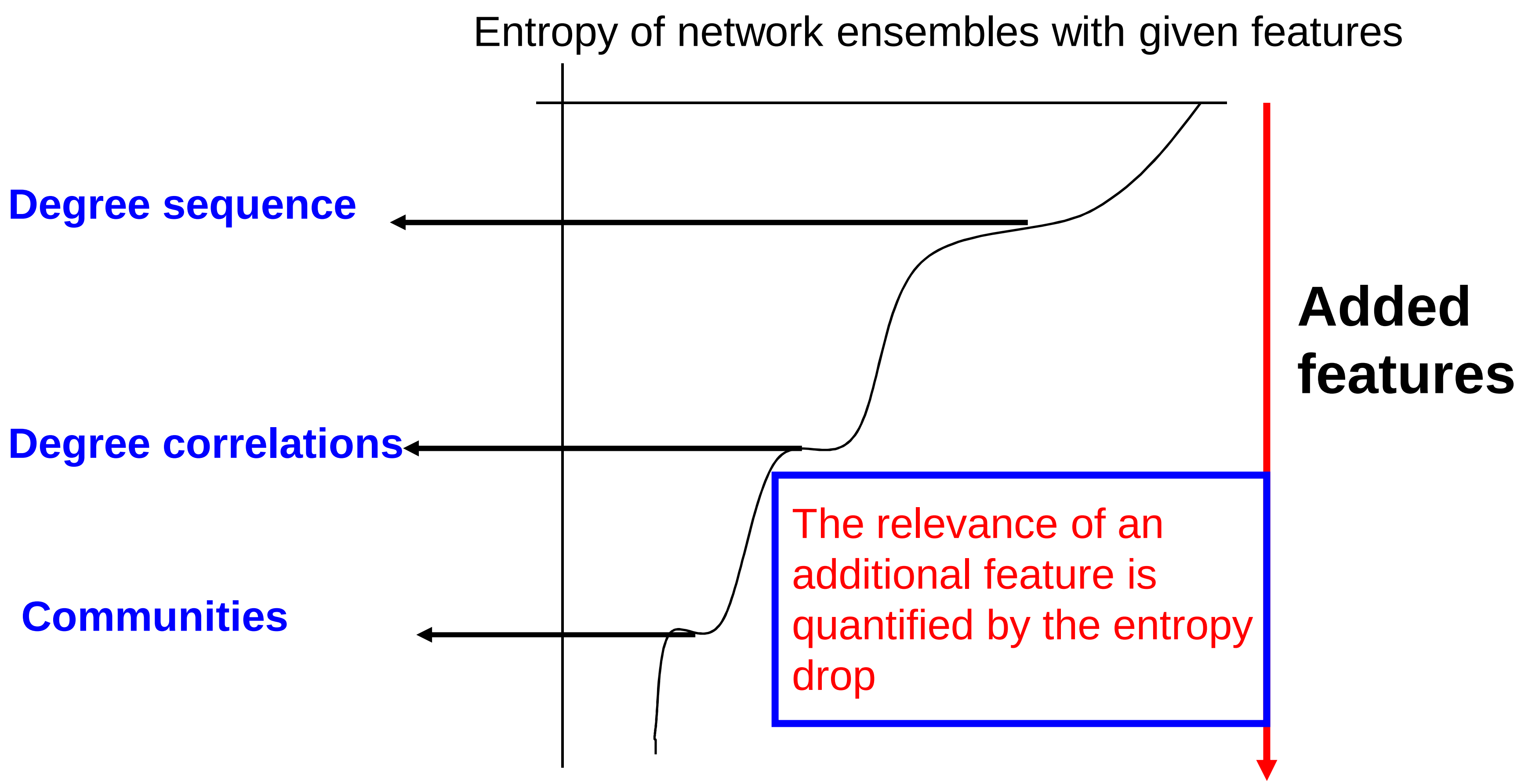
The complexity of networks is indicated by their organization at different levels

- **Average degree of a network**
- **Degree sequence**
- **Degree correlations**
- **Loop structure**
- **Clique structure**
- **Community structure**
- **Motifs**

Relevance of a network characteristics

How many networks

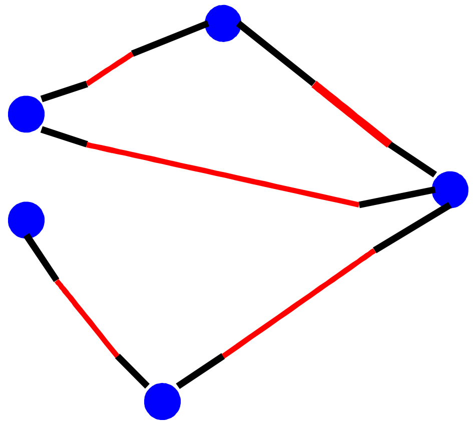
have the same:



Networks with given degree sequence

Microcanonical ensemble

$$P(G)=\frac{1}{\Sigma_1}\prod_i \delta(k_i-\sum_j a_{ij})$$

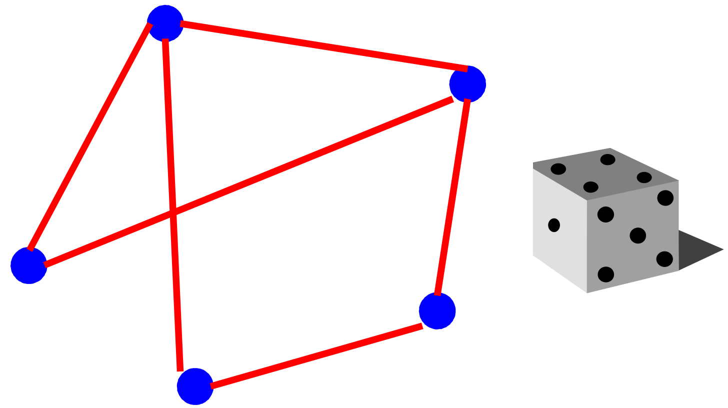


Ensemble of network with exactly M links

Molloy-Reed

Canonical ensemble

$$P(G)=\prod_{i<j} p_{ij}^{\ a_{ij}}(1-p_{ij})^{1-a_{ij}}$$



Ensemble of networks with average number of links M

Hidden variables

Shannon Entropy of canonical ensembles

$$S = -\frac{1}{N}\left[\sum_{ij} p_{ij}\ln p_{ij} + (1-p_{ij})\ln(1-p_{ij})\right]$$

We can obtain canonical ensembles by maximizing this entropy conditional to given constraints

Link probabilities

Constraints	Link probability
➢ Total number of links L=cN	$p_{ij}=\frac{c}{N}$
➢ Degree sequence $\{k_i\}$	$p_{ij}=\frac{\theta_i\theta_j}{1+\theta_i\theta_j}$
➢ Degree sequence $\{k_i\}$ and number of links in within and in between communities $\{q_i\}$	$p_{ij}=\frac{\theta_i\theta_j W(q_i,q_j)}{1+\theta_i\theta_j W(q_i,q_j)}$
➢ In spatial networks, degree sequence $\{k_i\}$ and number of links at distance d	$p_{ij}=\frac{\theta_i\theta_j W(d_{ij})}{1+\theta_i\theta_j W(d_{ij})}$

Anand & Bianconi PRE 2009

Partition function randomized microcanonical network ensembles

$$Z_{\kappa} = \sum_{\{a_{ij}\}} \prod_{k} \delta(\ldots \text{constraint}_k \ldots)\, e^{\sum_{i,j} h_{ij} a_{ij}}$$

h_{ij} auxiliary fields

Statistical mechanics
on the adjacency matrix of the network

$$a_{ij} = 1 \quad \text{if } i \text{ is linked to } j$$
$$a_{ij} = 0 \quad \text{otherwise}$$

The entropy of the randomized ensembles

Gibbs Entropy per node of a randomized network ensemble

$$\Sigma_{\kappa} = \frac{1}{N} \log(Z_{\kappa})\Big|_{h=0}$$

Probability of a link.

$$p_{ij} = \frac{\partial \log(Z_{\kappa})}{\partial h_{ij}}\Bigg|_{h=0}$$

The link probabilities in microcanonical and canonical ensembles are the same

-Example

Microcanonical ensemble **Canonical ensemble**

Regular networks Poisson networks

$$p_{ij}=\frac{c}{N}$$

$$p_{ij}=\frac{c}{N}$$

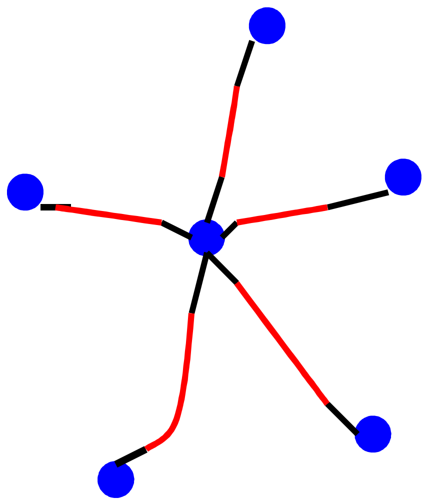
but

$$\Sigma < S$$

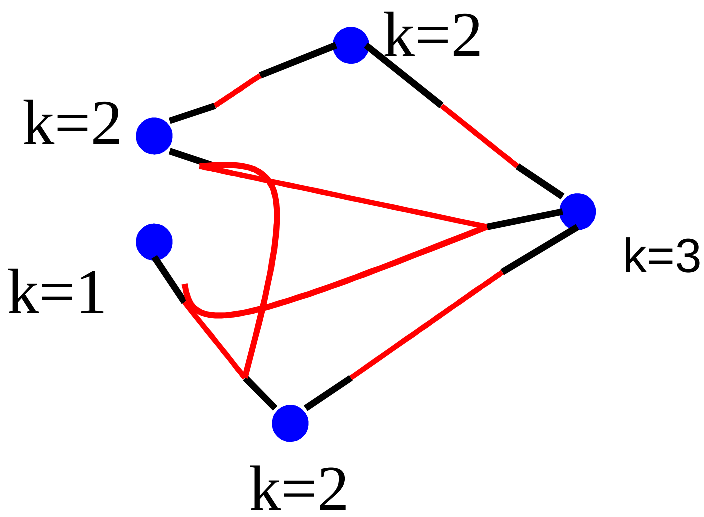
Anand & Bianconi 2009

Two examples of given degree sequence

Zero entropy

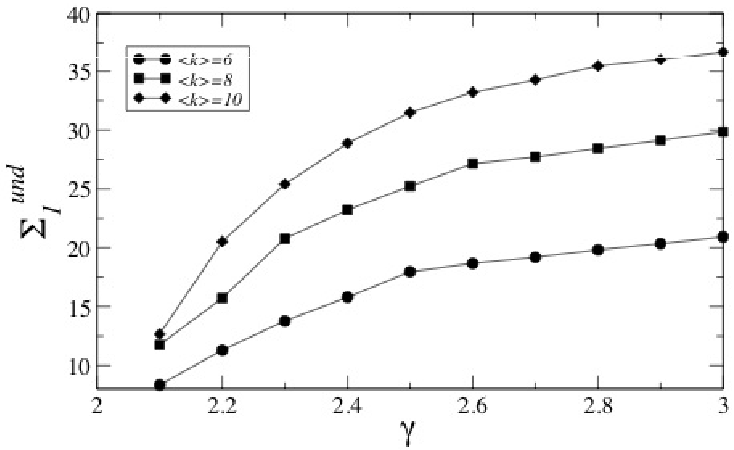


Non-zero entropy



The entropy of random scale-free networks

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



The entropy decreases as γ decrease toward 2
quantifying a higher order in networks with fatter tails

Bianconi 2008

Chance and Necessity:

Randomness is not all

Selection

or

non-equilibrium processes

have to play a role

in the evolution of

highly organized networks

Quantum statistics in equilibrium network models

➢**Simple networks**

Fermi-like distribution

$$p_{ij} = \frac{\theta_i \theta_j W(d_{ij})}{1+\theta_i \theta_j W(d_{ij})} = \frac{1}{1+e^{\beta \varepsilon_{ij}}}$$

$$\beta \varepsilon_{ij} = -\ln \theta_i - \ln \theta_j - \ln W(d_{ij})$$

➢**Weighted network**

Bose-like distributions

$$\langle w_{ij} \rangle = \frac{1}{e^{\beta(x_i + x_j)} - 1}$$

G. Bianconi PRE 2008
D. Garlaschelli, Loffredo PRL 2009

Other related works

- Ensembles of networks with clustering, acyclic

Newman PRL 2009, Karrer Newman 2009

- Entropy origin of disassortativity in complex networks

Johnson et al. PRL 2010

- Assessing the relevance of node features for network structure

Bianconi et al. PNAS 2009

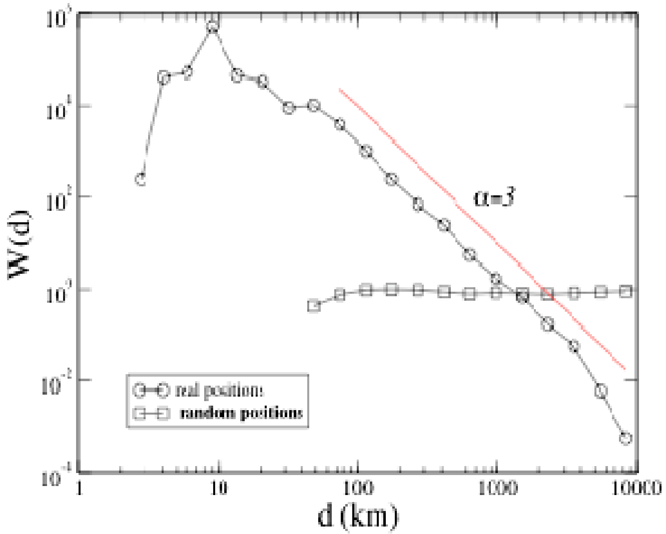
- Finding instability in the community structure of complex networks

Gfeller et al. PRE 2005

The spatial structure of the airport network

Link probability

$$p_{ij} = \frac{\theta_i \theta_j W(d_{ij})}{1 + \theta_i \theta_j W(d_{ij})}$$

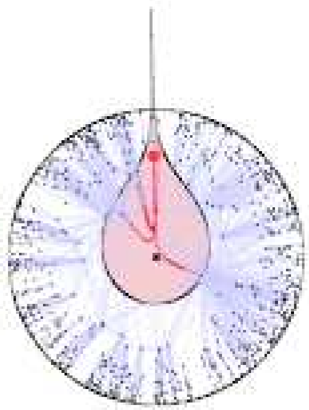


$$W(d) \approx d^{-\alpha}$$

$$\alpha \approx 3$$

G. Bianconi et al. PNAS 2009

Models in hidden hyperbolic spaces



The linking probability is taken to be dependent on the hyperbolic distance x between the nodes

$$p(x) = \frac{1}{1 + e^{\beta(x-R)}}$$

$$x = r + r' + \frac{2}{\zeta} \ln \sin \frac{\Delta\theta}{2}$$

Krioukov et al. PRE 2009

Dynamical networks

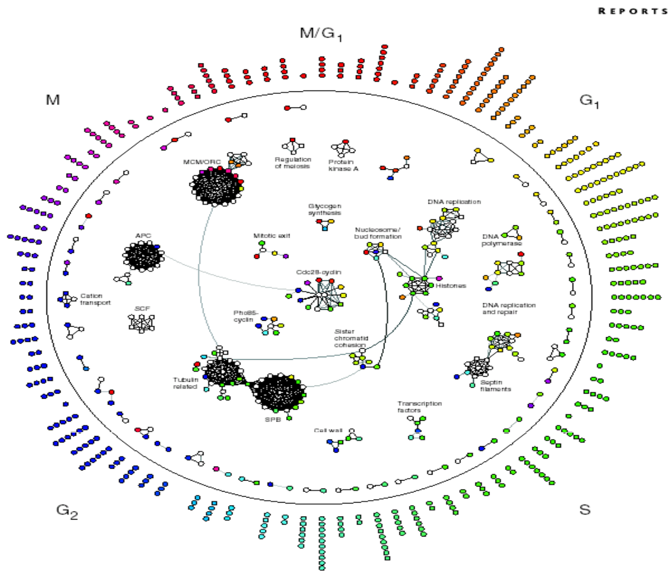
At any given time the network looks disconnected

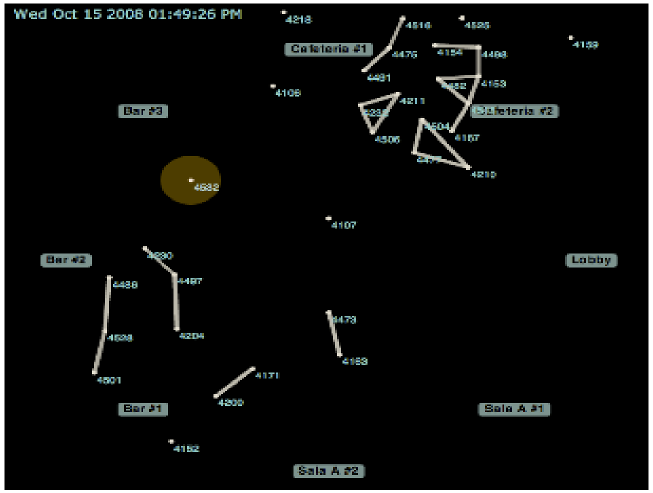
➢ **Protein complexes during the cell cycle of yeast**

De Lichtenberg et al.2005

➢ **Social networks (phone calls, small gathering of people)**

Barrat et al.2008





Conclusions

The modeling of complex networks is a continuous search to answer well studied questions as

- **Why we observe the universality network structure?**
- **How can we model a network at a given level of coarse-graining?**

And new challenging questions…

- **What is the relation between network models and quantum statistics?**
- **Space: What is the geometry of given complex networks?**
- **Time: How can we model the dynamical behavior of complex social and biological networks?**