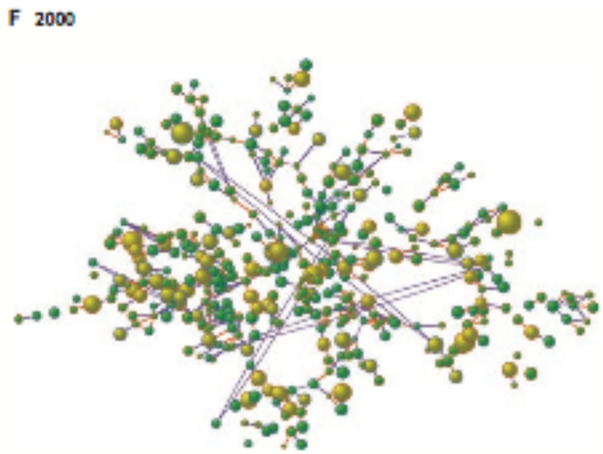
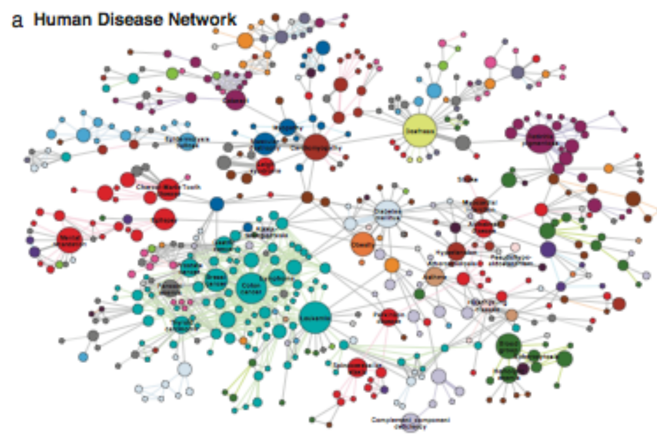




Complexity Summer School

Warwick, 29 April-1 May 2013

Network models

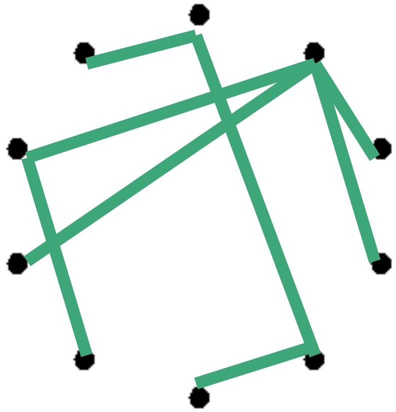
Ginestra Bianconi

School of Mathematical Sciences, Queen Mary University of London, London, UK

Random graphs

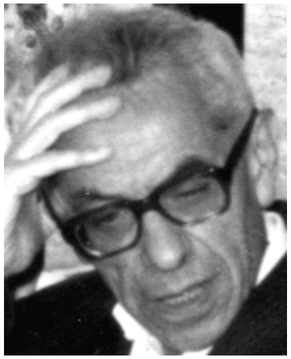
G(N,L) ensemble

Graphs with exactly
N nodes and
L links



G(N,p) ensemble

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

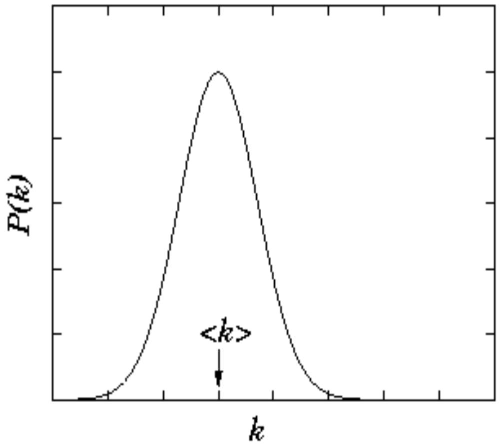


Binomial distribution

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

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

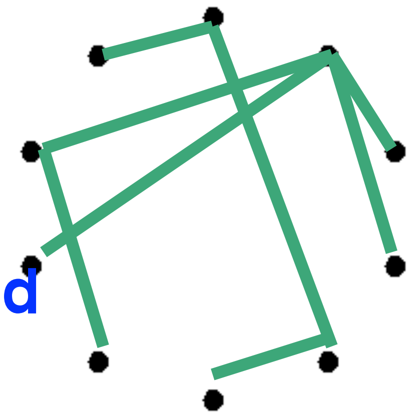
Poisson distribution

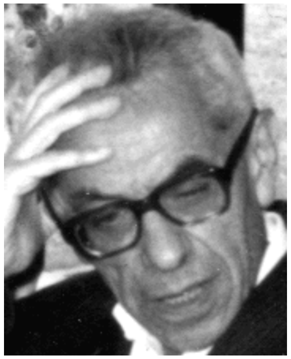


Random graphs

G(N,p) ensemble

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





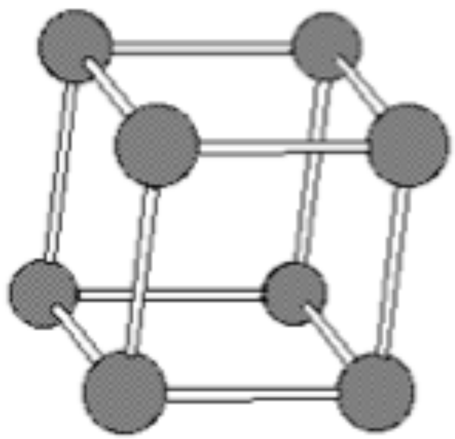
Poisson distribution

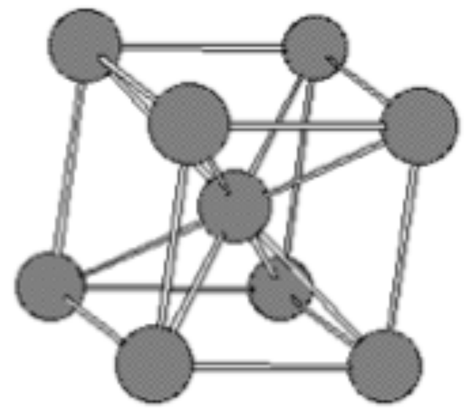
Small clustering coefficient

Small average distance

$$\Rightarrow \begin{cases} C(N) \propto \dfrac{1}{N} \\ \langle \ell \rangle \propto \dfrac{\log(N)}{\log(c)} \end{cases}$$

Regular lattices

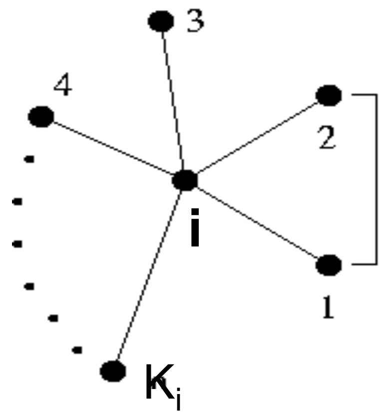




d dimensions

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

Universalities:Small world



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

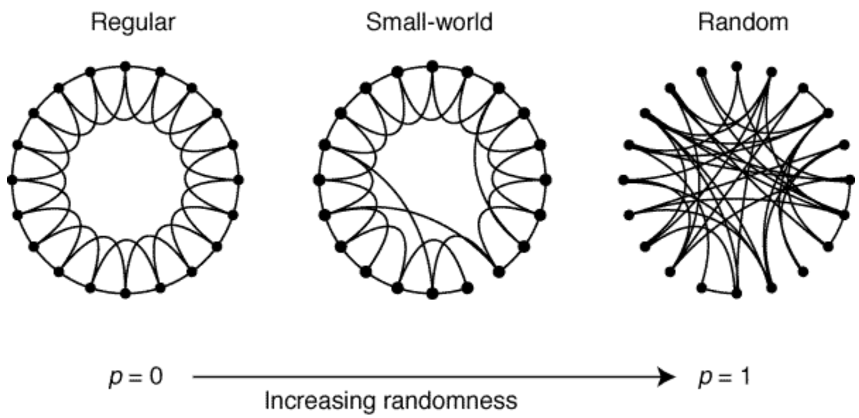
Networks are clustered (large average C_i,i.e.C)

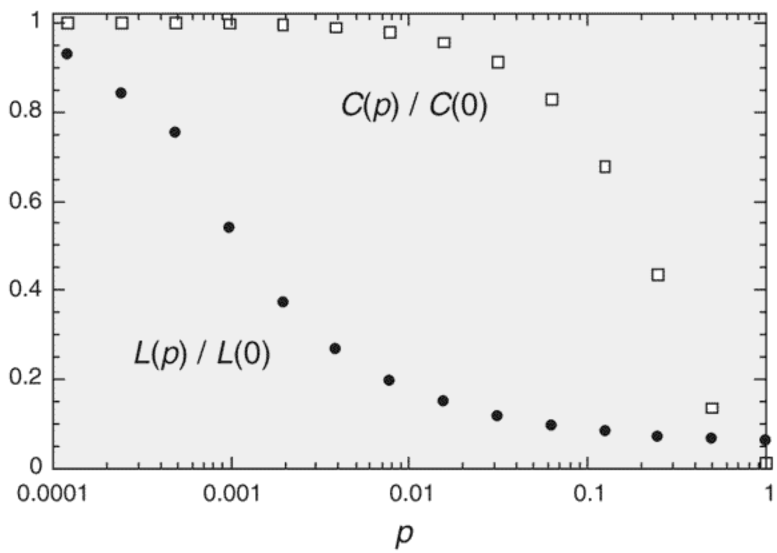
but have a small characteristic path length (small L).

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)

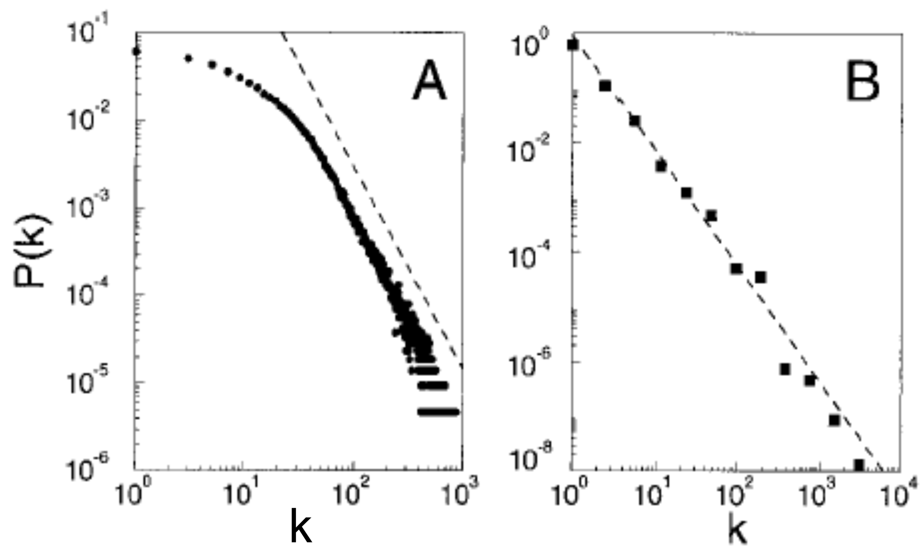
Newman & Watts, (1999)

Barrat & Weigt, (2000)

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

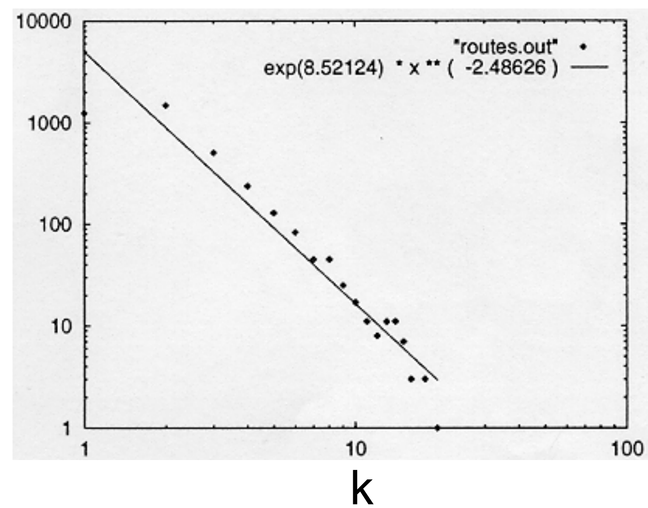
Universalities:
Scale-free degree distribution

Actor networks WWW



Barabasi-Albert 1999

Internet



Faloutsos et al. 1999

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

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

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

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

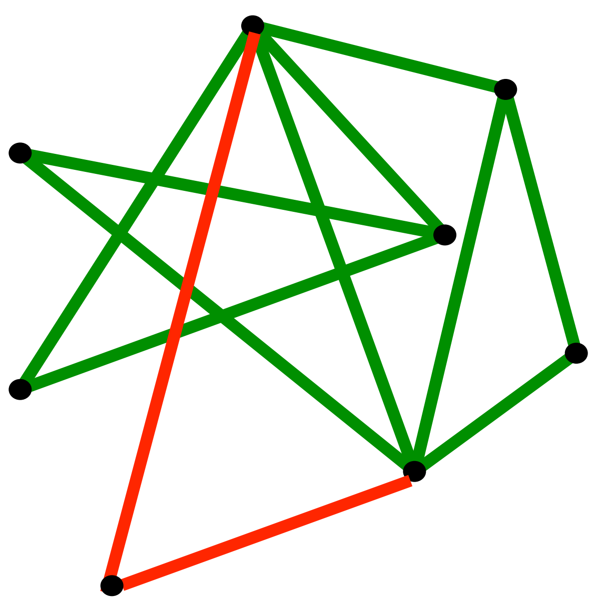
(1) GROWTH :

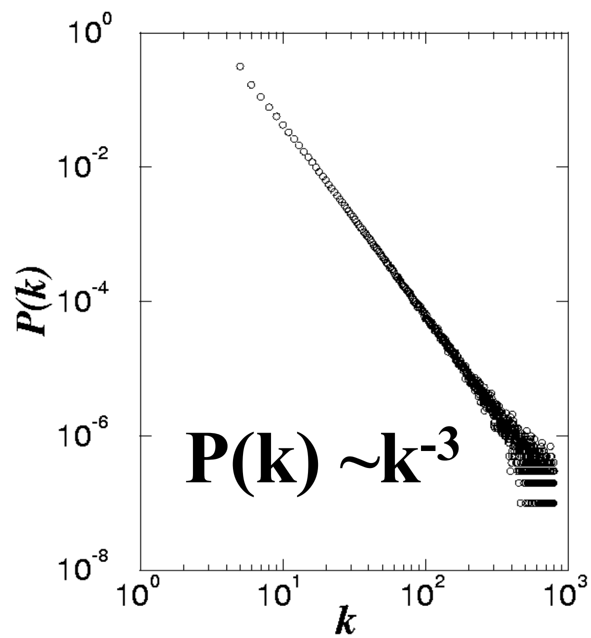
At every timestep we add a new node with *m* edges (connected to the nodes already present in the system).

(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)

The mean-field approach in a nutshell

We write the equation for the average degree of node i at time t $k_i(t)$ and we derive the degree distribution in the mean-field approximation

$$\frac{dk_i}{dt} = m\frac{k_i}{\sum_j k_j} \approx m\frac{k_i}{2mt}, \quad k_i(t_i) = m$$

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

$$P(k_i(t) > k) = P(t_i < tm^2/k^2) = \frac{m^2}{k^2}$$

$$P(k) = -\frac{dP(k_i(t) > k)}{dk} = 2\frac{m^2}{k^3}$$

Result of the mean-field solution of the BA model

- In the mean-field approximation 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 $\gamma=3$:

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

Barabasi Albert (1999)

Master equation approach

- The exact degree distribution is obtained through the master equation approach.
- The master equation is an equation for the average number of nodes of degree k at time t, $N_k(t)$ and reads for *m=1* and *t>>1*

$$N_k(t+1) = N_k(t) + \frac{k-1}{2t} N_{k-1}(t) - \frac{k}{2t} N_k(t) \ \ for\ k>1$$

$$N_1(t+1) = N_1(t) - \frac{1}{2t} N_1(t) + 1$$

- In the limit $t>>1$ $N_k(t)=tP_k$, substituting this expression in the master equation we get

$$P_k = \frac{4}{(k+2)(k+1)k}$$

Dorogovtsev et al (2000)

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

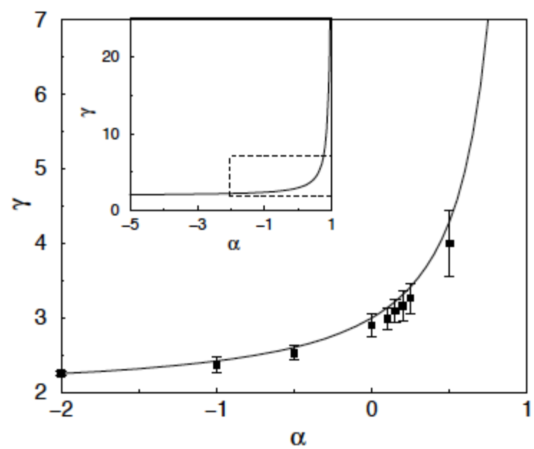
Growing network with aging of the nodes

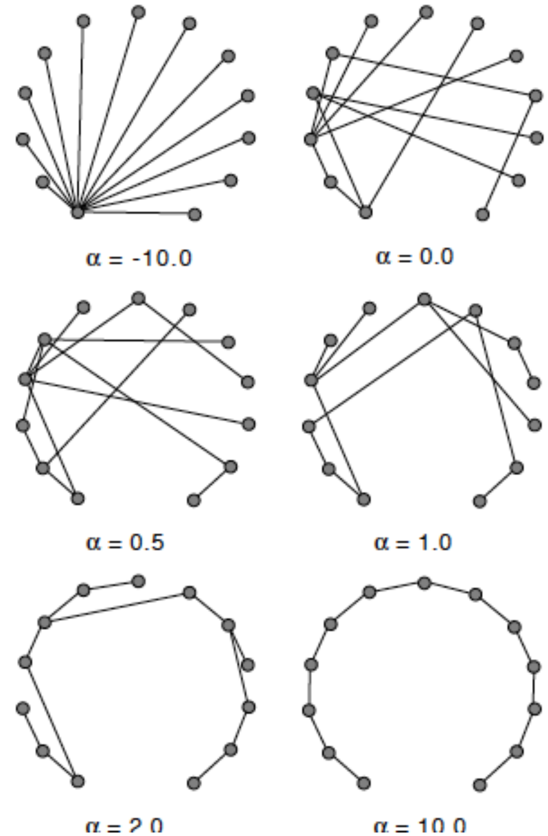
- Generalized preferential attachment

$$\Pi_i \propto k_i (t - t_i)^{-\alpha}$$

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

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





Dorogovtsev Mendes (2000)

Non-linear preferential attachment

The probability to link to a node i is given by

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

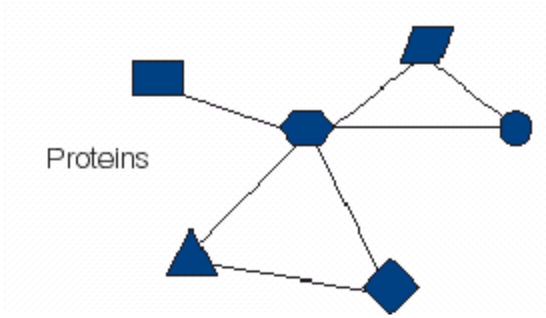
- $\alpha<1$ **Absence of power-law degree distribution**
- $\alpha=1$ **Power-law degree distribution**
- $\alpha>1$ **Gelation phenomena**

  The oldest node acquire most of the links

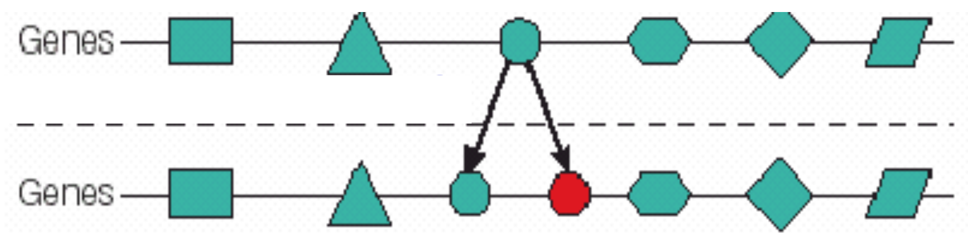
  First-mover-advantage

Krapivsky et al 2000

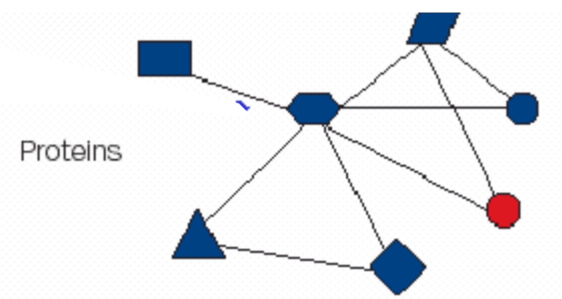
Gene duplication model



- **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

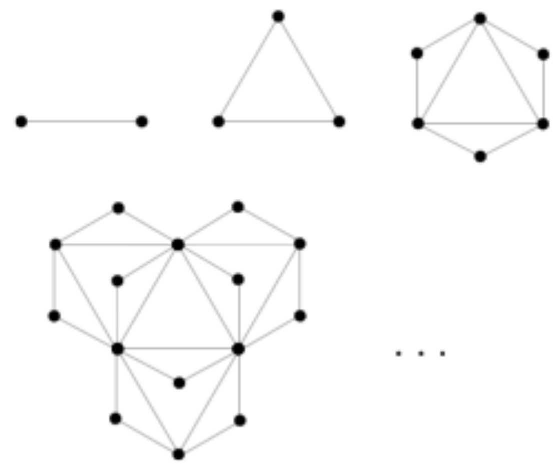
➢**Scale-free networks with high-clustering coefficient**

Dorogovtsev et al. 2001

Eguiluz & Klemm 2002

➢**Pseudofractal scale-free network**

Dorogovtsev et al 2002



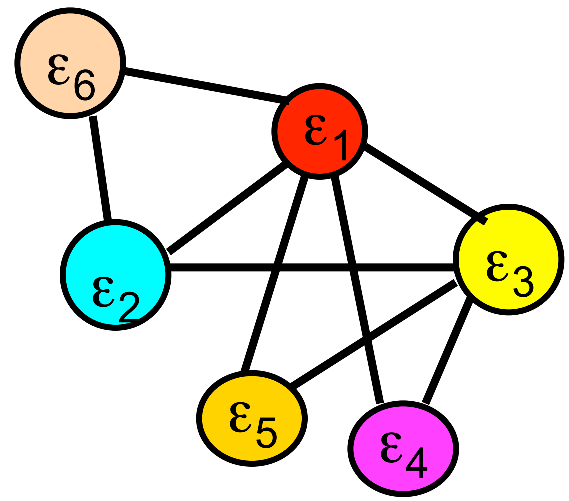
Fitness the nodes

Not all the nodes are the same!

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

that describes the capacity of a node to attract new links

In the limit β=0 all the nodes have same fitness



The fitness model

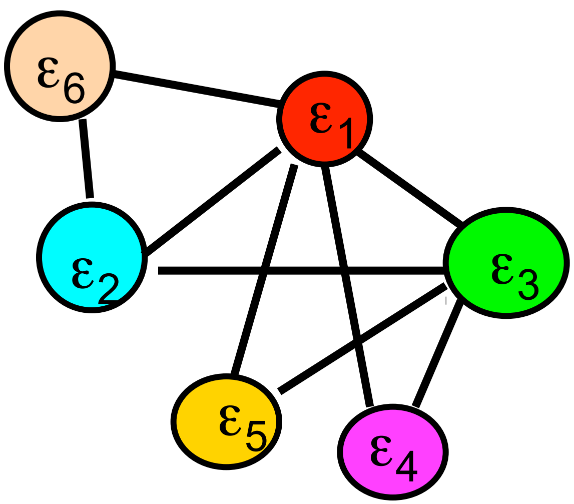
Growth:

–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**

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

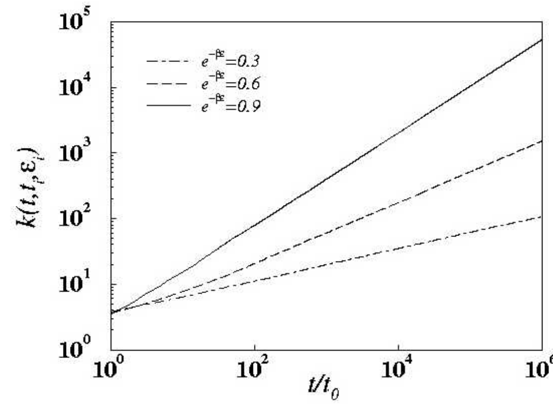
- **Fit-get-rich mechanism**

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

Fit-get rich mechanism

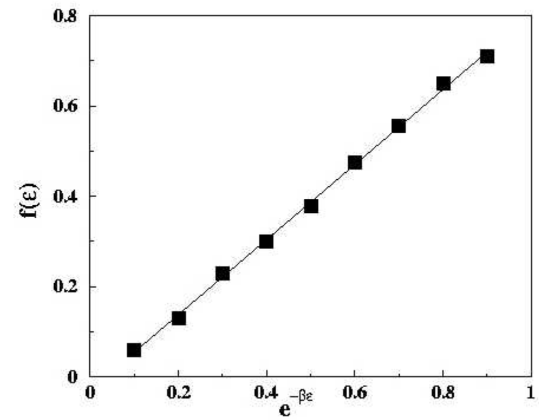
The nodes with higher fitness increases the connectivity faster

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



μ satisfies the condition

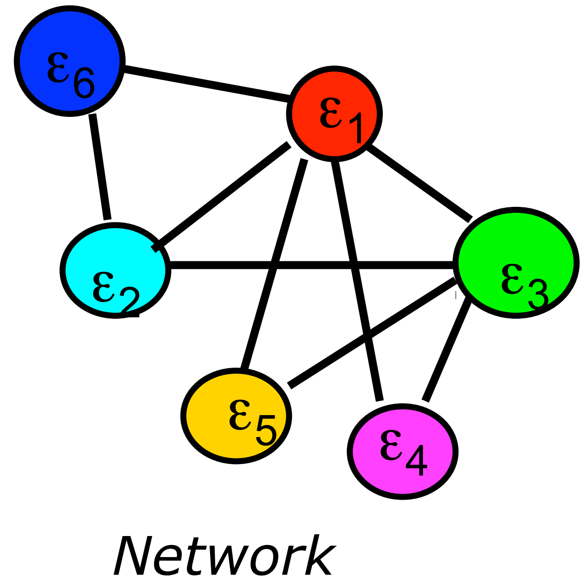
$$1 = \int d\varepsilon\, p(\varepsilon)\frac{1}{e^{\beta(\varepsilon-\mu)}-1}.$$



Mapping to a Bose gas

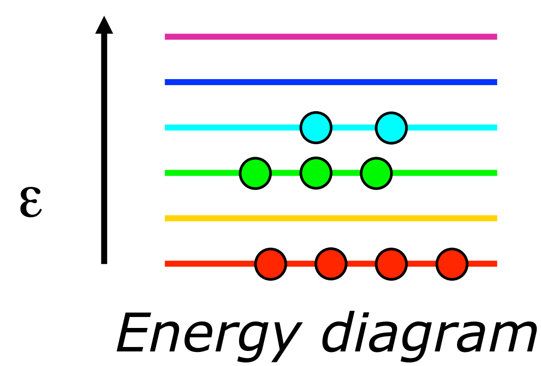
We can map the fitness model to a Bose gas with

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



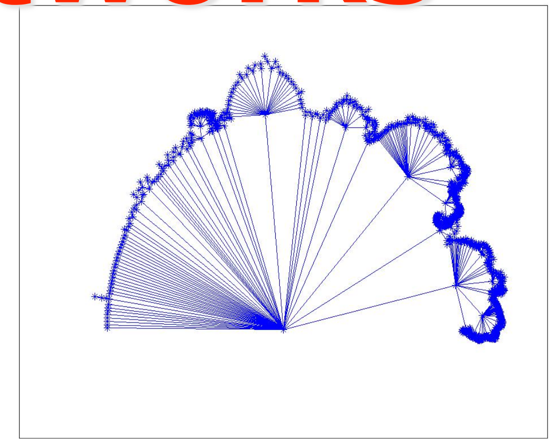
In this mapping,

- 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.



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

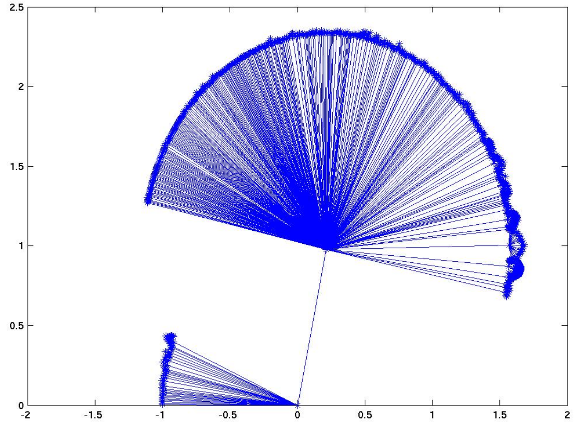
fit-get-rich phase

•for $T<T_c$ the network is in the

winner-takes-all

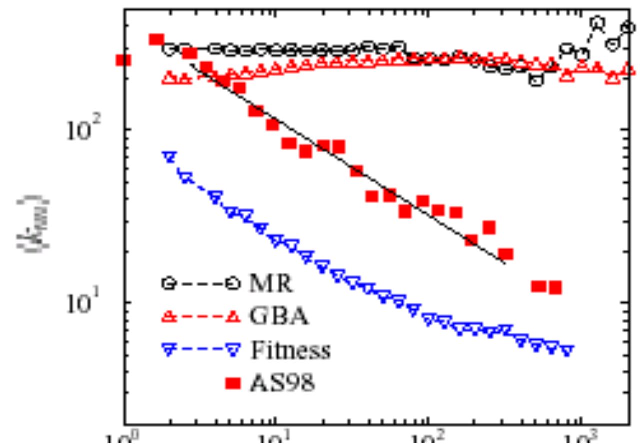
or Bose-condensate

phase

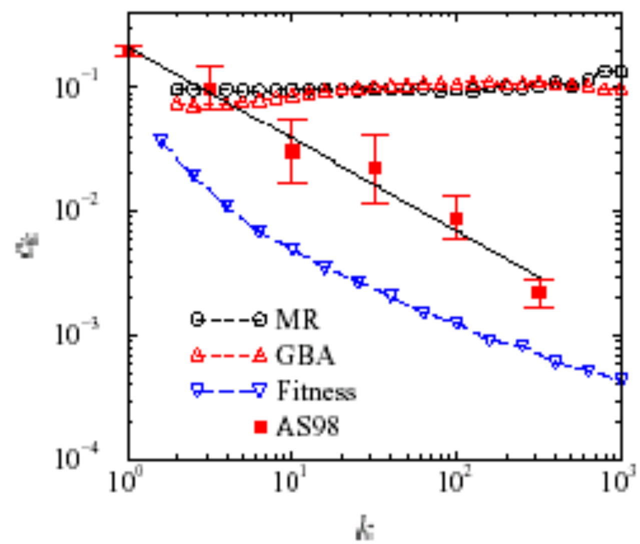


Correlations in the Internet and the fitness model

- $k_{nn}(k)$ 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

Why this universality?

- Growing networks:
  - **Preferential attachment**

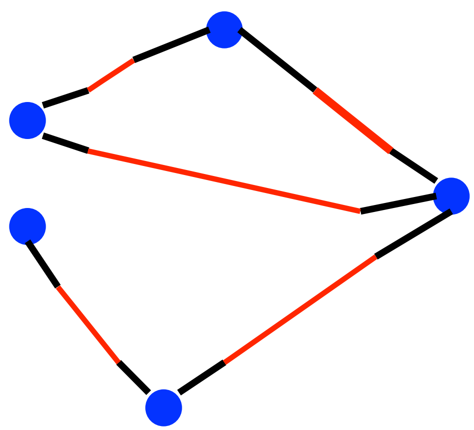
Barabasi & Albert 1999,
Dorogovtsev Mendes 2000, Bianconi & Barabasi 2001, etc.

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

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

Configuration model

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



Bollobas,
Molloy & Reed

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
- *Note that the degree sequence must be graphical!*

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**

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

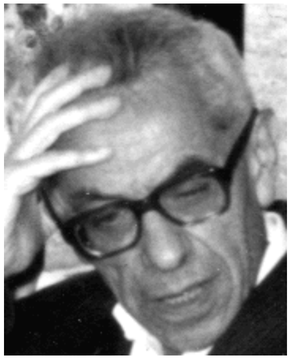
Random graphs

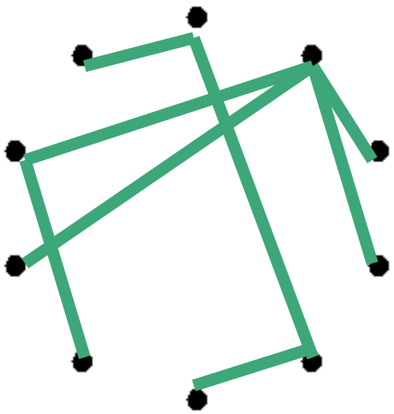
G(N,L) ensemble

Graphs with exactly N nodes and L links

G(N,p) ensemble

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



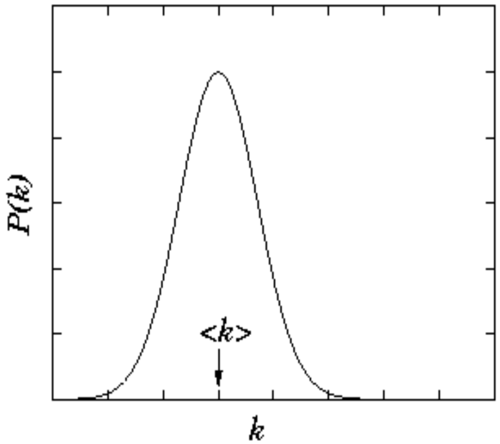


Binomial distribution

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

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

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

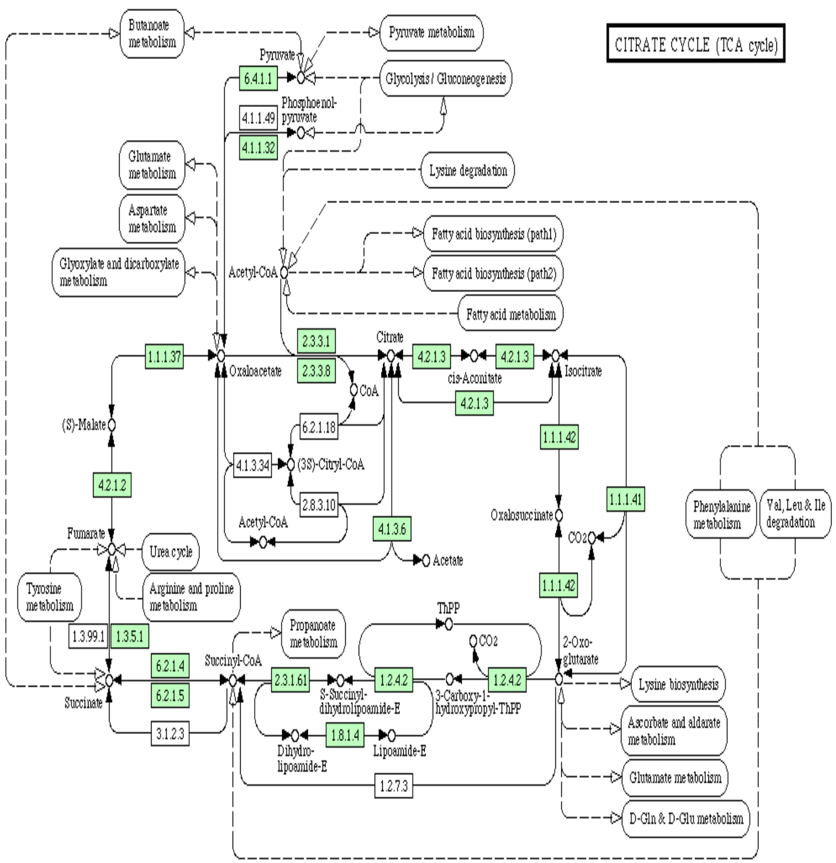
is a decreasing function

of the entropy of this ensemble

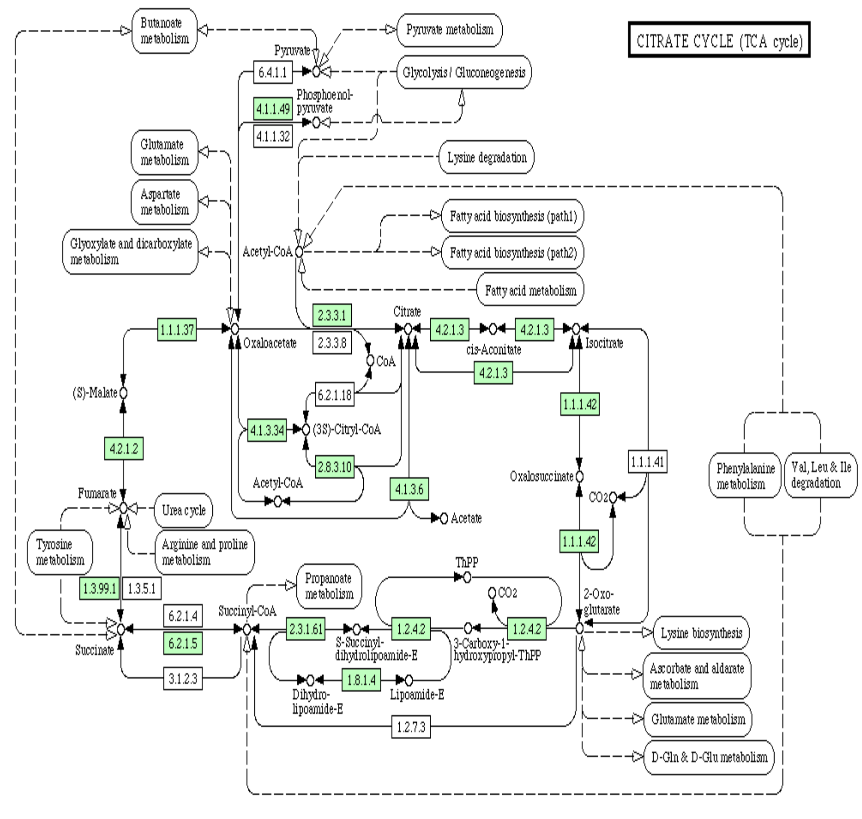
G. Bianconi EPL (2008)

Citrate cycle:highly preserved

Escherichia coli



Homo Sapiens



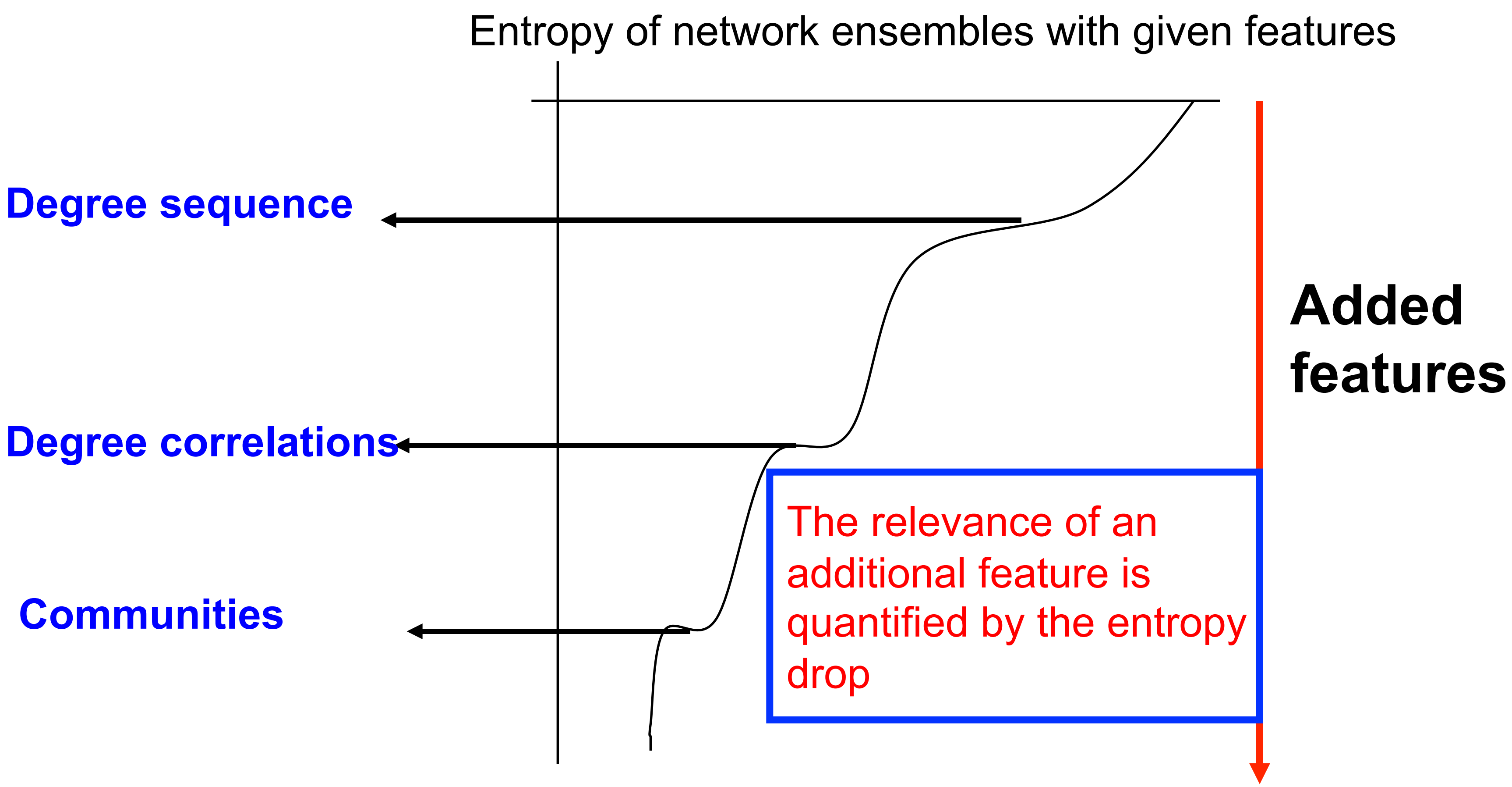
The complexity of networks is indicated by their organization at different levels

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

Relevance of a network characteristics

How many networks

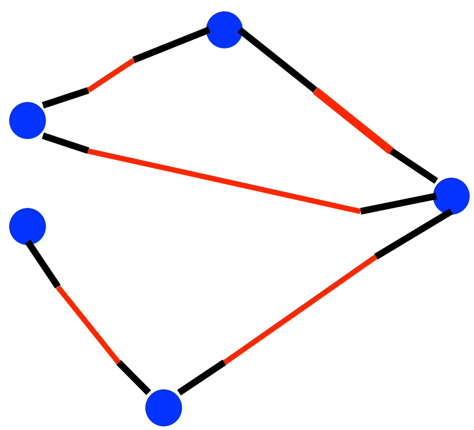
have the same:



Networks with given degree sequence

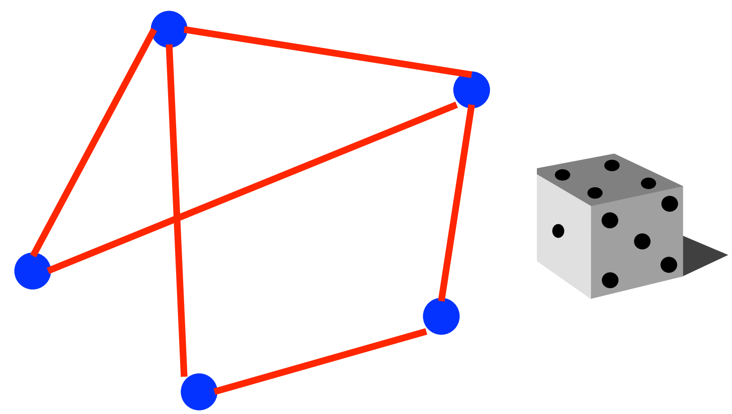
Microcanonical ensemble

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



Canonical ensemble

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



Ensemble of network with exactly M links

Molloy-Reed

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

- Total number of links $L=cN$
- Degree sequence $\{k_i\}$
- Degree sequence $\{k_i\}$ and number of links in within and in between communities $\{q_i\}$
- In spatial networks, degree sequence $\{k_i\}$ and number of links at distance d

Link probability

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

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

$$p_{ij} = \frac{\theta_i \theta_j W(q_i, q_j)}{1+\theta_i \theta_j W(q_i, q_j)}$$

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

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

$$a_{ij} = 1 \quad \textit{if } i \textit{ is linked to } j$$

$$a_{ij} = 0 \quad \textit{otherwise}$$

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.

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

Gibbs entropy

The Gibbs entropy is equal to the Shannon entropy of the corresponding canonical ensemble minus a correction

$$\Sigma = S/N - \Omega$$

$$\Omega = -\frac{1}{N}\log\left[\sum_{\{a\}}\prod_{i<j} p_{ij}^{\,a_{ij}}(1-p_{ij})^{1-a_{ij}}\prod_{\kappa}\delta(const_{\kappa})\right]$$

Ω is the normalized logarithm of the probability that a canonical network satisfy the hard constraints

K. Anand and G. Bianconi PRE 2010

Case of linear constraints

- *Degree sequence*
- *Degree sequence and number of links within communities*
- *Degree sequence an number of links at distance d*

$$K_\alpha = \sum_{\ell \in \partial\alpha} a_\ell$$

$$\Omega[P] = -\frac{1}{N}\sum_\alpha \log(\pi_{K_\alpha}(K_\alpha))$$

If the number of constraints is extensive there is no equivalence between Gibbs and Shannon entropy

G Bianconi, ACC Coolen and C. Perez-Vicente PRE 2008
K. Anand and G. Bianconi PRE 2010

The link probabilities in microcanonical and canonical ensembles are the same

-Example

Microcanonical esemble	**Canonical ensemble**
Regular networks	Poisson networks

$$p_{ij} = \frac{c}{N-1}$$

but

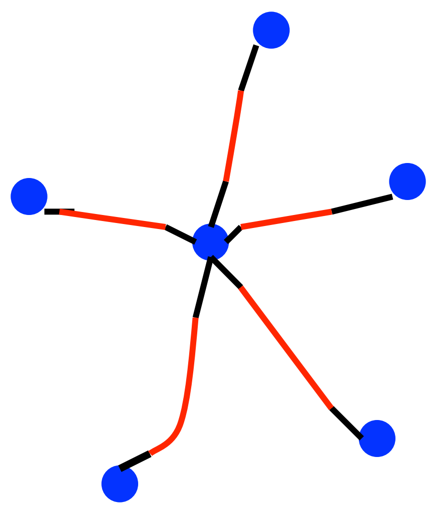
$$p_{ij} = \frac{c}{N-1}$$

$$N\Sigma < S$$

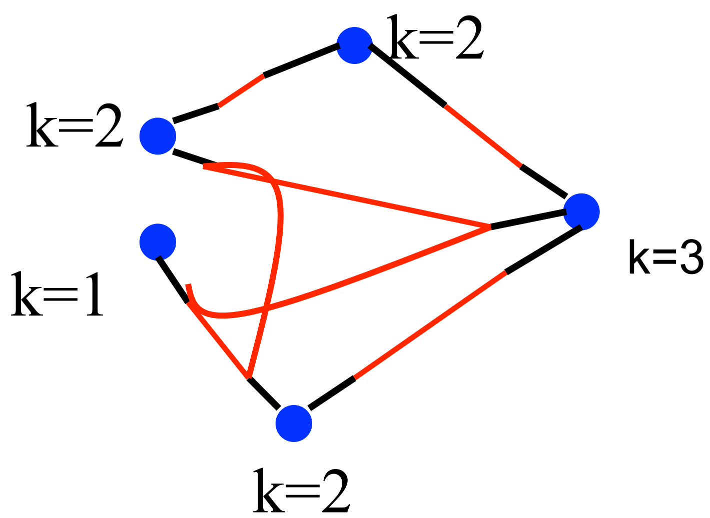
K. Anand, G. Bianconi PRE 2009

Two examples of given degree sequence

Zero entropy

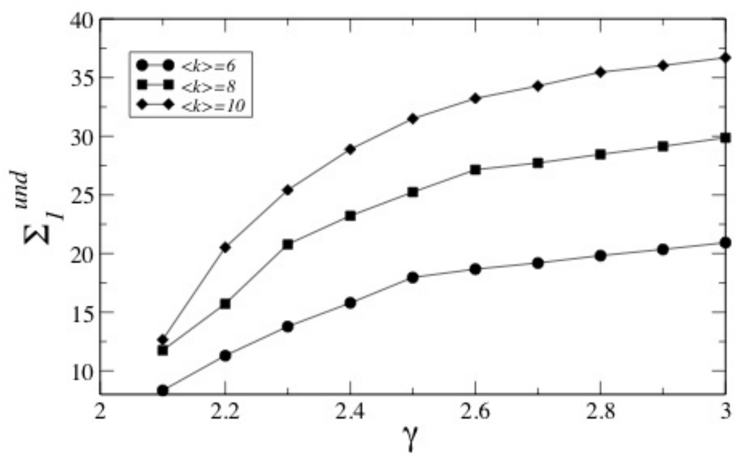


Non-zero entropy



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

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

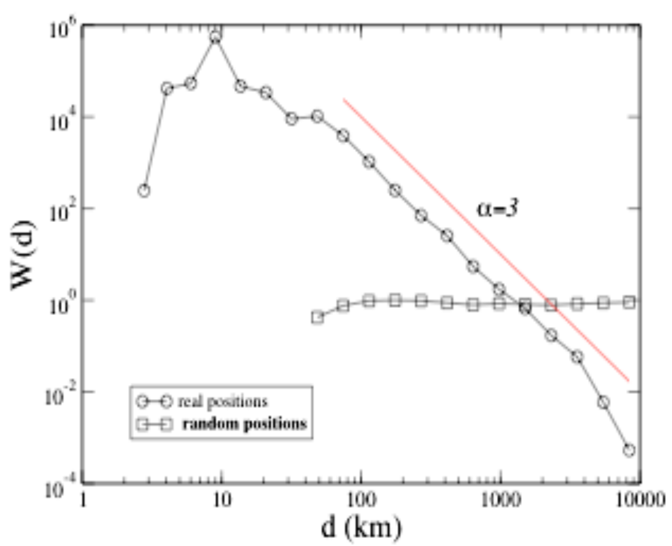
- Assessing the relevance of node features for network structure

Bianconi et al. PNAS 2009

The spatial structure of the airport network

Link probability

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

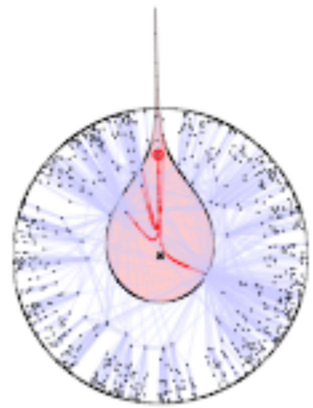


$$W(d) \approx d^{-\alpha}$$

$$\alpha \approx 3$$

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

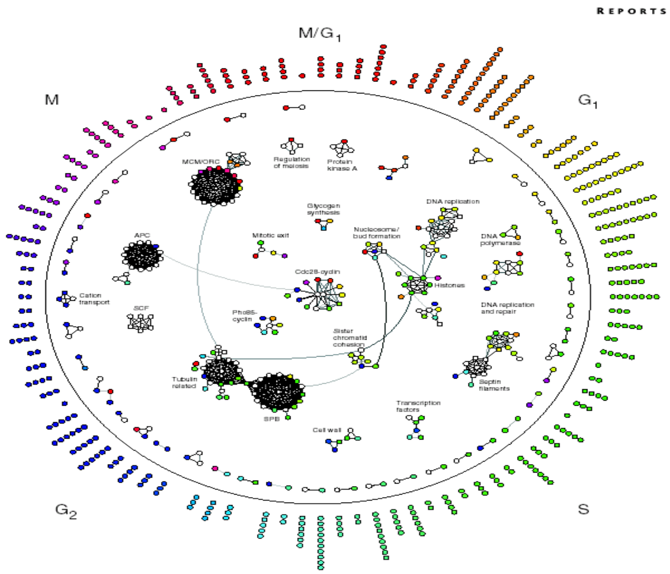
At any given time the network looks disconnected

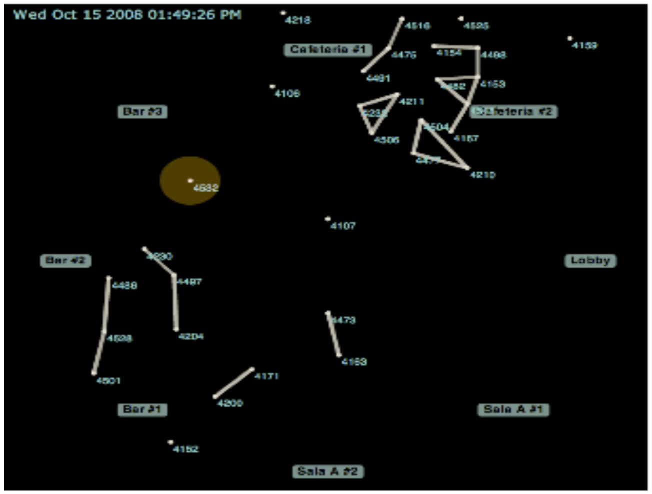
- **Protein complexes during the cell cycle of yeast**

  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 scale-free and small-world universalities in the network structure?**

- **How can we model a network at a given level of coarse-graining?**

And new challenging questions…

- **Iteracting networks: What happens is different networks interact with each other?**

- **Time: How can we model the dynamical behavior of complex social and biological networks?**