Frontiers of Network Science
Fall 2023

Class 12: Degree Correlations part II
(Chapter 7 in Textbook)

Boleslaw Szymanski

based on slides by
Albert-László Barabási
and Roberta Sinatra

www.BarabasiLab.com

Structural cut-off

High assortativity→ high number of links between the hubs.

If we allow only one link between two nodes, we can simply run out of hubs to connect to each other to satisfy the assortativity criteria.

Number of edges between the set of nodes with degree *k* and degree *k'*:

$$E_{kk'} = e_{kk'} \underline{\langle k \rangle N}$$

This represents the number of edges

This is frequency of edges with degree kk'

Maximum number of edges between the two groups:

$$m_{kk'} = \min\{kN_k, k'N_{k'}, N_k N_{k'}\}$$

There cannot be more links between the two groups, than the overall number of edges joining the nodes with degree k, or k'

If we only have **simple edges**, we cannot have more links between the two groups, than if we connect every node with degree k to every node with degree k' **once**.

This is true even if we allow multiple edges.

M. Boguñá, R. Pastor-Satorras, A. Vespignani, EPJ B **38,** 205 (2004)

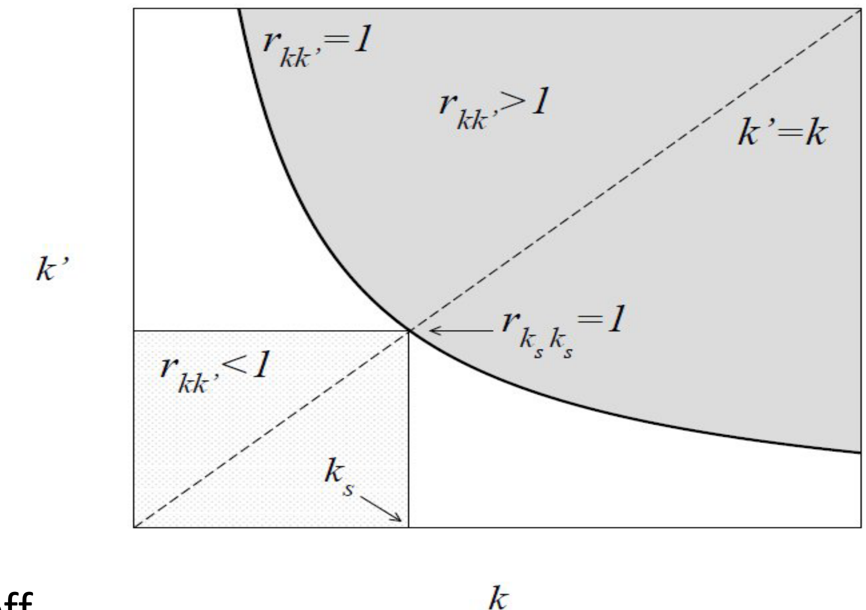
Structural cut-off

$$E_{kk'} = e_{kk'} \langle k \rangle N$$

$$m_{kk'} = \min\{kN_k, k'N_{k'}, N_k N_{k'}\}$$

The ratio of $E_{kk'}$ and $m_{kk'}$ has to be ≤ 1 in the physical region!

$$r_{kk'} = \frac{E_{kk'}}{m_{kk'}} \leq 1$$



$\longrightarrow$ $r_{k_s k_s} = 1$ defines the structural cut-off

M. Boguñá, R. Pastor-Satorras, A. Vespignani, EPJ B **38,** 205 (2004)

Structural cut-off for uncorrelated networks

Uncorrelated networks:

$$m_{kk'} = \min\{kN_k, k'N_{k'}, N_kN_{k'}\}$$

$$m_{k_sk_s} = k_sN_{k_s} = k_sNp_{k_s}$$

$$m_{k_sk_s} = N_{k_s}^2 = N^2p_{k_s}^2$$

$$e_{kk'} = q_kq_{k'} = \frac{kk'p_kp_{k'}}{\langle k\rangle^2} \longrightarrow r_{kk'} = \frac{E_{kk'}}{m_{kk'}} = \frac{\langle k\rangle N e_{kk'}}{m_{kk'}}$$

$$r_{k_sk_s} = \frac{\langle k\rangle N\cdot k_s^2\cdot p_{k_s}^2}{\langle k\rangle^2 k_sp_{k_s}N} = \frac{k_sp_{k_s}}{\langle k\rangle} = q_{k_s} < 1 \quad \forall k_s$$

$$r_{k_sk_s} = \frac{\langle k\rangle N\cdot k_s^2\cdot p_{k_s}^2}{\langle k\rangle^2 N^2\cdot p_{k_s}^2} = \frac{k_s^2}{\langle k\rangle N} \longrightarrow \boxed{k_s(N) = (\langle k\rangle N)^{1/2}}$$

***$k_s(N)$* represents a *structural cutoff*:**

one cannot have nodes with degree larger than $k_s(N)$,

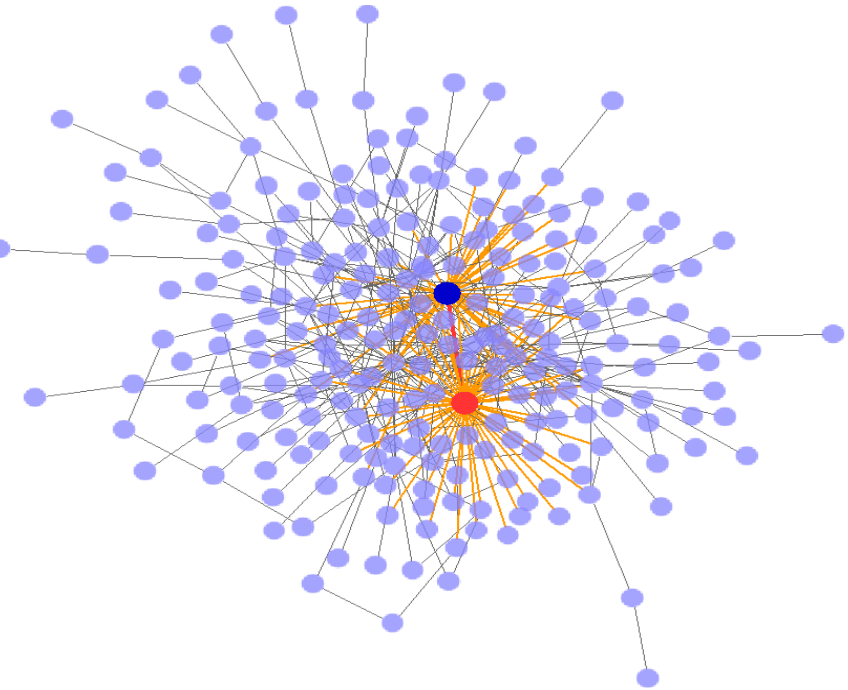
→*if there are nodes with $k > k_s(N)$ we cannot find sufficient links between the highly connected nodes to maintain the neutral nature of the network.*

Solution:

(a) Introduce a structural cutoff (i.e. do not allow nodes with $k > k_s(N)$)

(b) Let the network become more dissasortative, having fewer links between hubs.

Example: Degree sequence introduces disassortativity



Scale-free network generated with the configuration model (N=300, L=450, γ=2.2).

The measured r=-0.19! → Dissasortative!

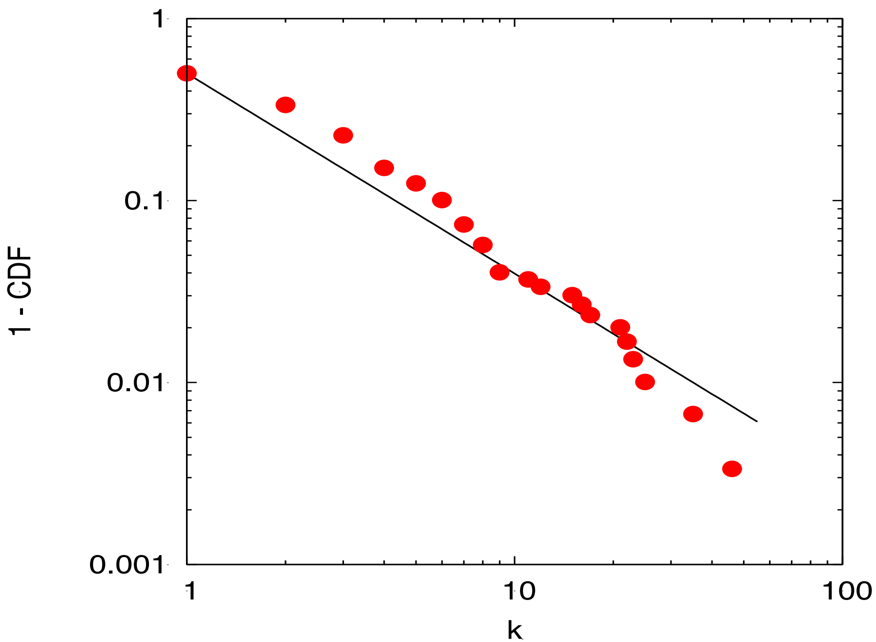
Red hub: 55 neighbors.
Blue hub: 46 neighbors.

Let's calculate the expectation number of links between red node (k=55) and blue node (k=46) for uncorrelated networks!

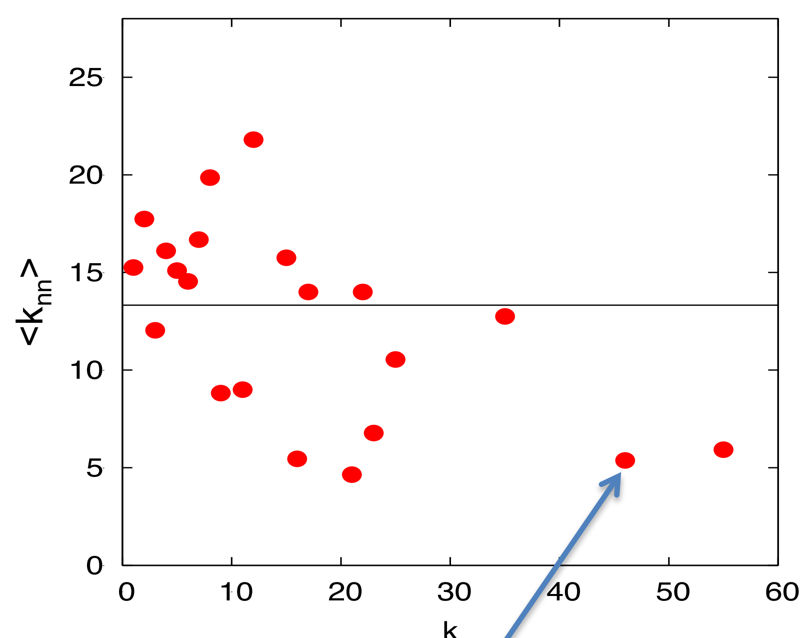
Here $N_{55}=N_{46}=1$, hence $m_{55,46}=1$ so $r_{55,46}=E_{55,46}$

$$E_{55,46} = \langle k \rangle N \cdot e_{55,46} = 900 \cdot \frac{\overset{k}{55}\,\overset{p_k}{\frac{1}{300}} \cdot \overset{k'}{46}\,\overset{p_{k'}}{\frac{1}{300}}}{\underset{\langle k \rangle}{3^2}} \approx 2.8 > 1$$

In order for the network to be neutral, we need 2.8 links between these two hubs.

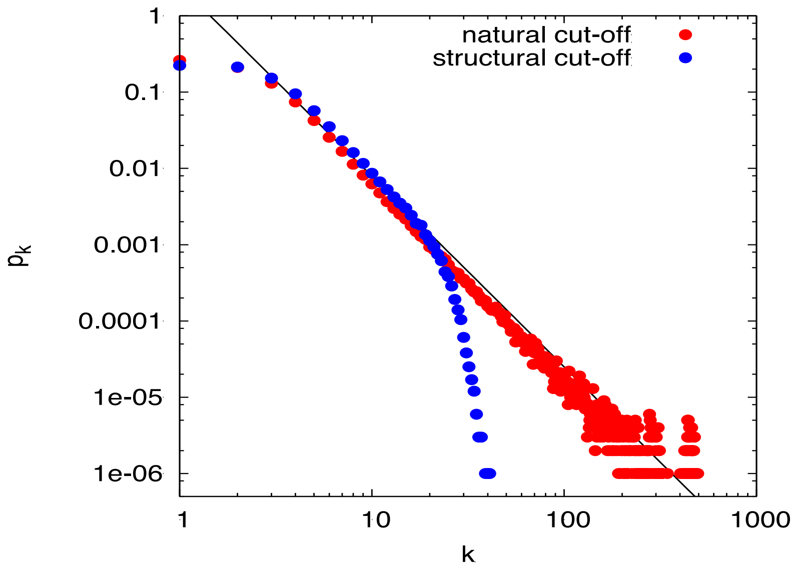


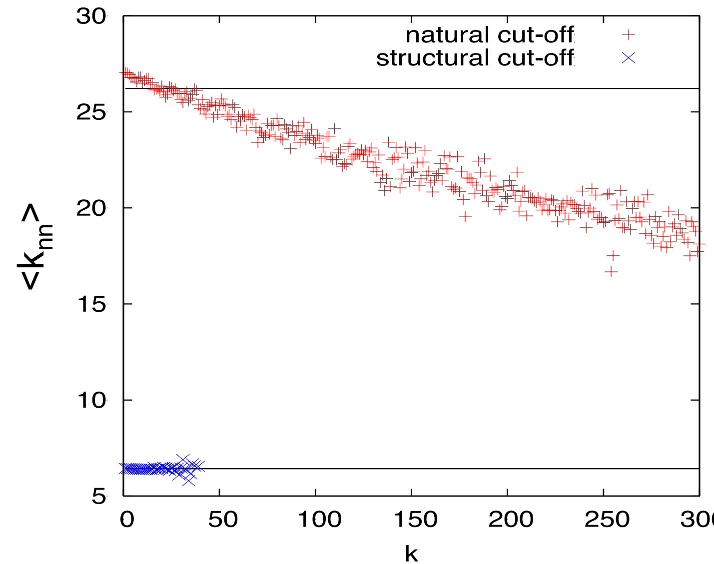
$$1 - CDF = P(k' > k) = 1 - \sum_{k'}^{k} p_{k'}$$



The largest nodes have $k_{nn} < \langle k_{nn} \rangle$

The effect is particularly clear for N=10,000:





The red curves are those of interest to us: one can see that a clear dissasortativity property is visible in this case.

Natural cutoffs in scale-free networks

All real networks are finite → let us explore its consequences.
→ We have an expected maximum degree, K_{max}

Estimating K_{max}

$$\int_{K_{\max}}^{\infty} P(k)dk \approx \frac{1}{N}$$

Why: the probability to have a node larger than K_{max} should not exceed the prob. to have one node, i.e. *1/N* fraction of all nodes

$$\int_{K_{\max}}^{\infty} P(k)dk = (\gamma-1)K_{\min}^{\gamma-1}\int_{K_{\max}}^{\infty} k^{-\gamma}dk = \frac{(\gamma-1)}{(-\gamma+1)}K_{\min}^{\gamma-1}\left[k^{-\gamma+1}\right]_{K_{\max}}^{\infty} = \frac{K_{\min}^{\gamma-1}}{K_{\max}^{\gamma-1}} \approx \frac{1}{N}$$

Natural cutoff: $$K_{\max} = K_{\min} N^{\frac{1}{\gamma-1}}$$

Structural cut-off for uncorrelated networks

Structural cutoff: $\boxed{k_s(N) \sim (\langle k \rangle N)^{1/2}}$

$$e_{kk'} = q_k q_{k'} = \frac{kk' p_k p_{k'}}{\langle k \rangle^2}$$

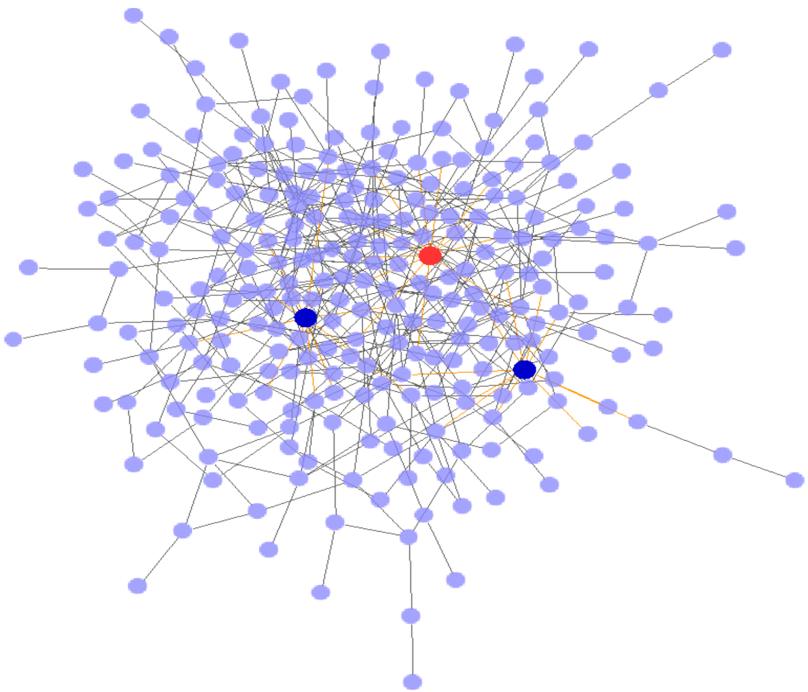
Natural cut-off: $k_{\max}(N) \sim N^{\frac{1}{\gamma-1}}$

γ=3: $k_s(N)$ and $k_{max}(N)$ scale the same way, i.e. $\sim N^{1/2}$.

γ<3: $k_{\max} > k_s \longrightarrow$ The size of the largest hub is above the structural cutoff, which means that it cannot have enough links to the other hubs to maintain its neutral status.
→ disassortative mixing

→a randomly wired network with γ<3 will be
(a) dissasortative
(b) Or will have to have a cutoff at $k_s(N) < k_{max}(N)$

Example: introducing a structural cut-off



Scale-free network generated with the configuration model (N=300, L=450, γ=2.2) with structural cut-off ~ N½.

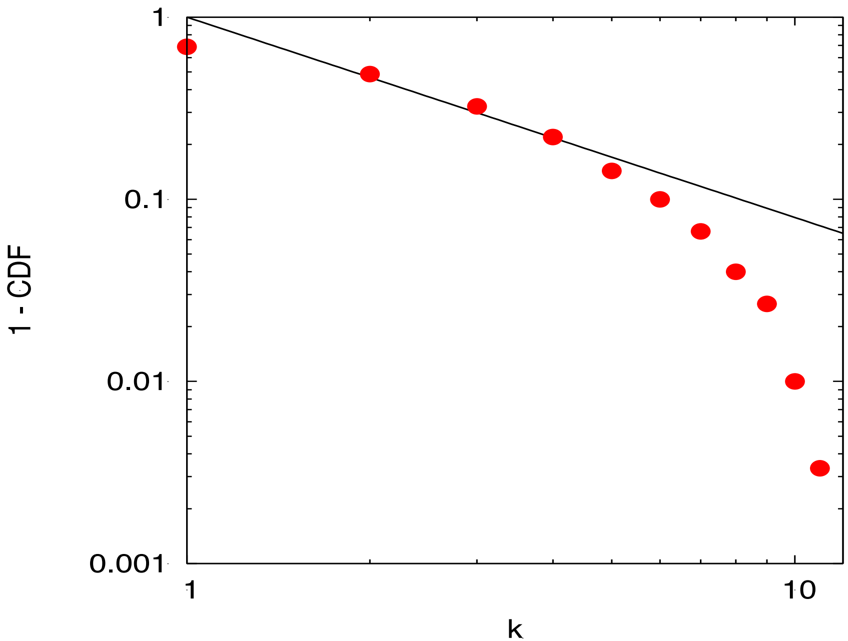
r=0.005 → neutral

Red hub: 12 neighbors.

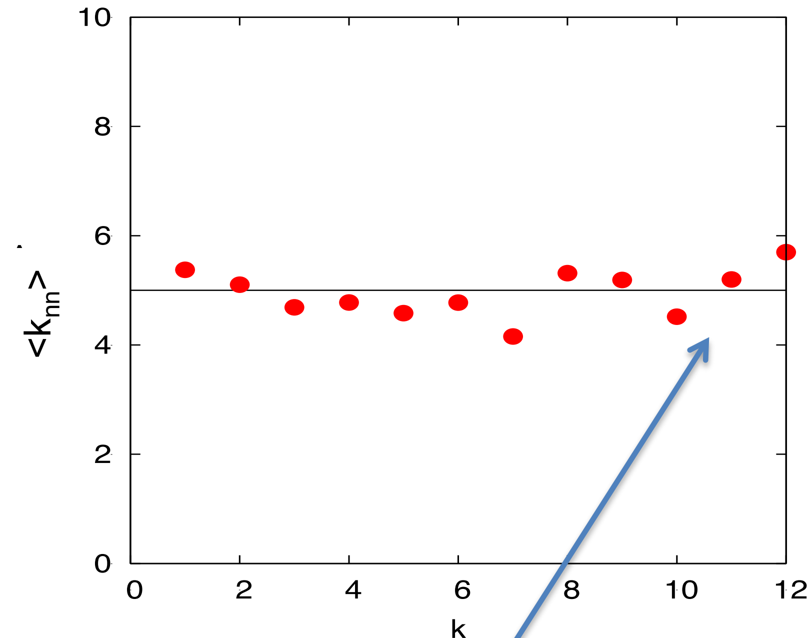
Blue hubs: 11 neighbors.

Again we can calculate the expectation number of edges between the hubs.

$$E_{11,12} = \langle k \rangle N \cdot e_{11,12} = 900 \cdot \frac{\overset{k}{12}\,\overset{p_k}{\frac{1}{300}} \cdot \overset{k'}{11}\,\overset{p_{k'}}{\frac{2}{300}}}{\underset{\langle k \rangle}{3^2}} \approx 0.3 < 1$$

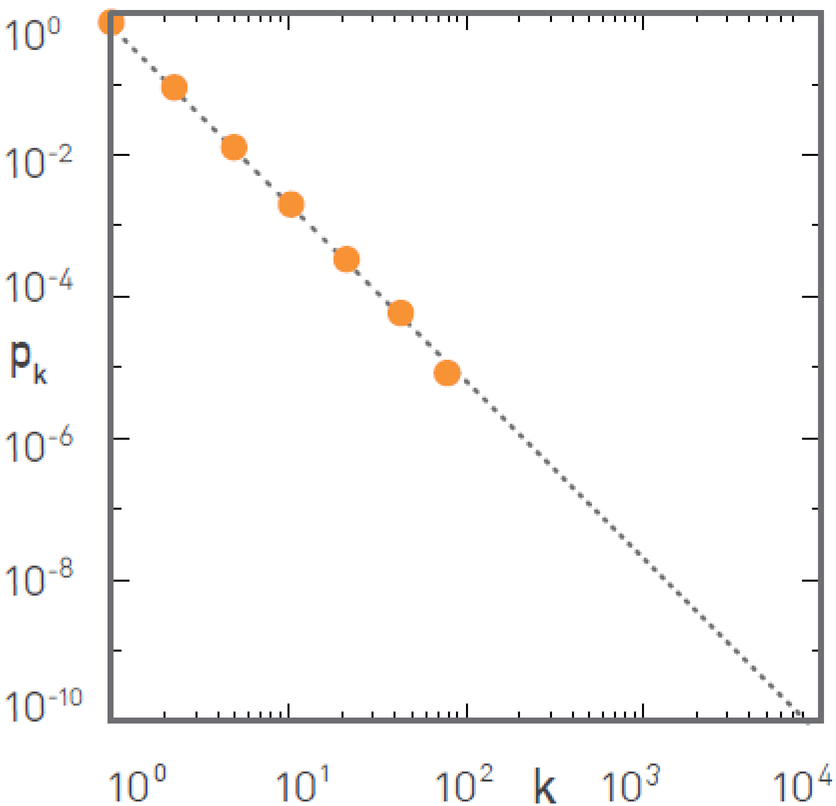


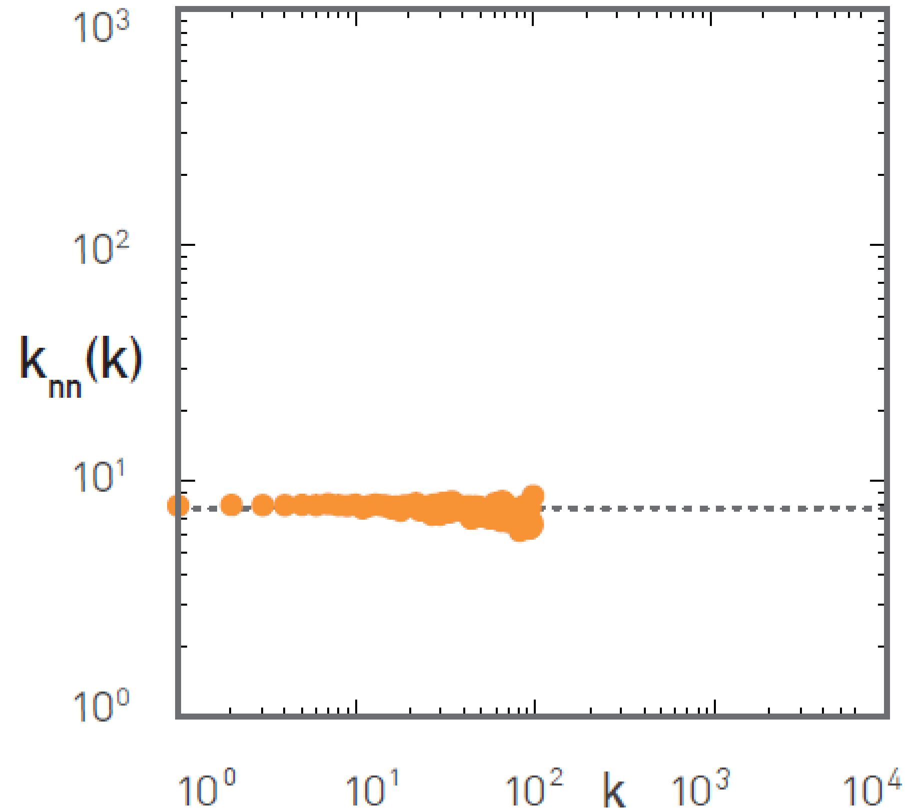
$$1 - CDF = P(k' > k) = 1 - \sum_{k'}^{k} p_{k'}$$



The largest nodes have $k_{nn} \sim \langle k_{nn} \rangle$

The effect is particularly clear for N=10,000:



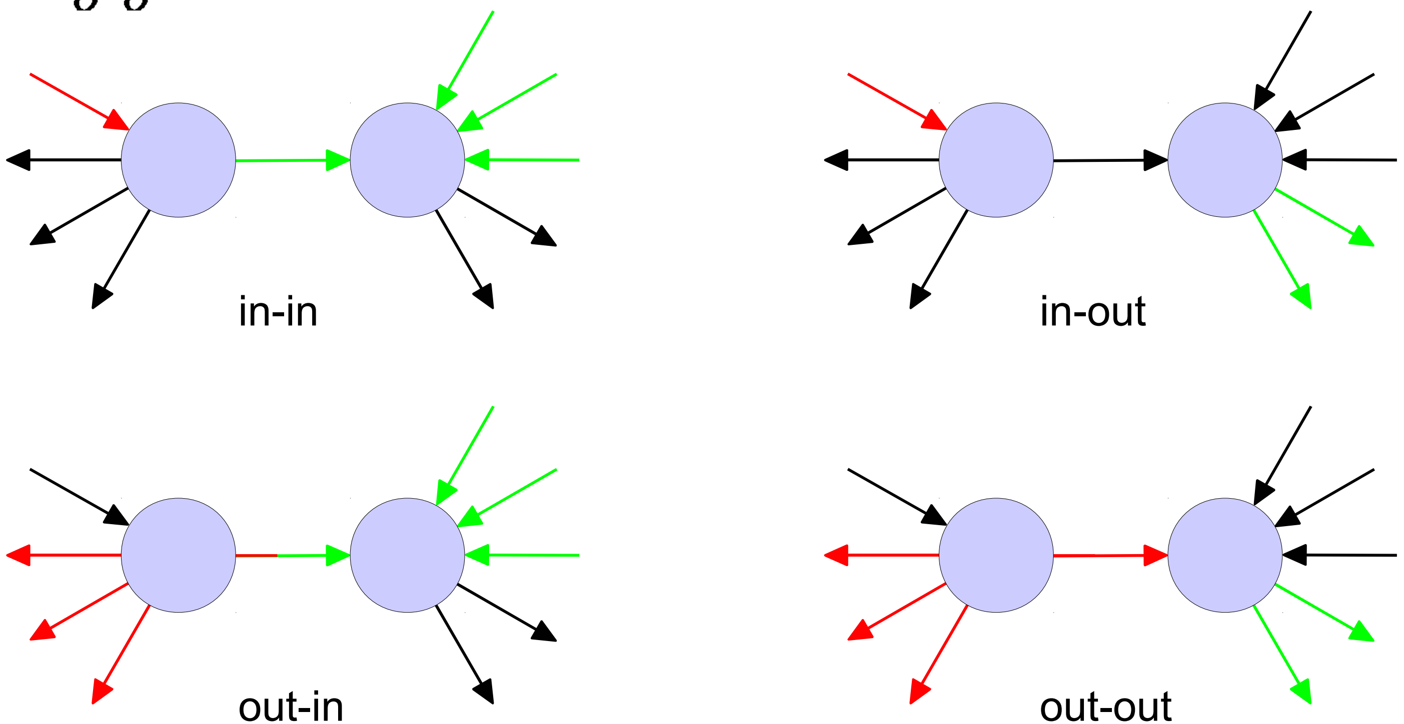


A clear case of neutral assortativity property is visible in this case thanks to imposing structural cut-off.

DIRECTED NETWORKS

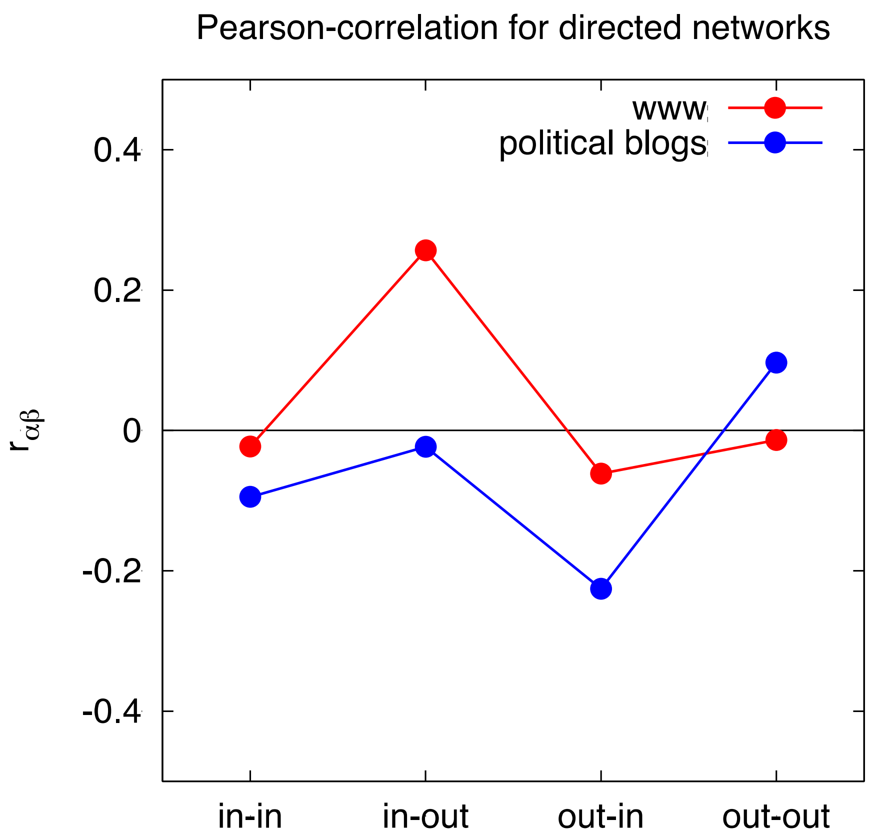
$$r_{\alpha\beta} = \frac{\sum_{jk} jk\left(e_{jk}^{\alpha\beta} - q_j^{\alpha} q_k^{\beta}\right)}{\sigma^{\alpha}\sigma^{\beta}}$$

α,β: {in,out}



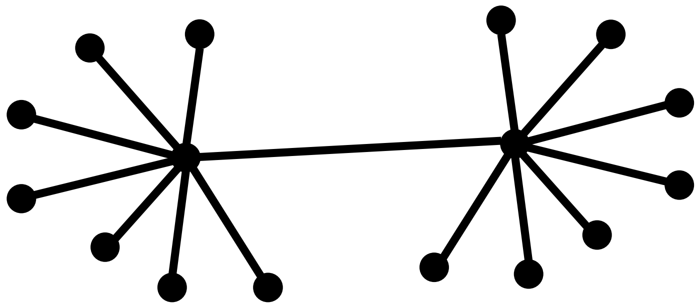
J. G. Foster, D. V. Foster, P. Grassberger, M. Paczuski, PNAS **107,** 10815 (2010)

DIRECTED NETWORKS

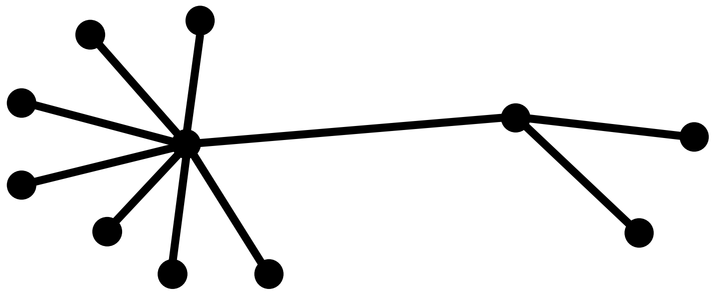


MULTIPOINT DEGREE CORRELATIONS

P(k): not enough to characterize a network



Large degree nodes tend to connect to large degree nodes
Ex: social networks



Large degree nodes tend to connect to small degree nodes
Ex: technological networks

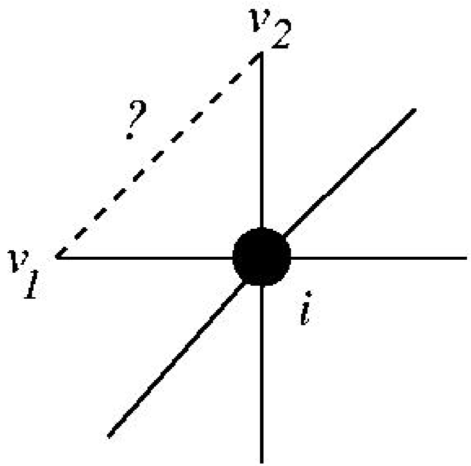
MULTIPOINT DEGREE CORRELATIONS

Measure of correlations:
$P(k',k'',\ldots k^{(n)}|k)$: conditional probability that a node of degree k is connected to nodes of degree k', k'',…

Simplest case:
$P(k'|k)$: conditional probability that a node of degree k' is connected to a node of degree k

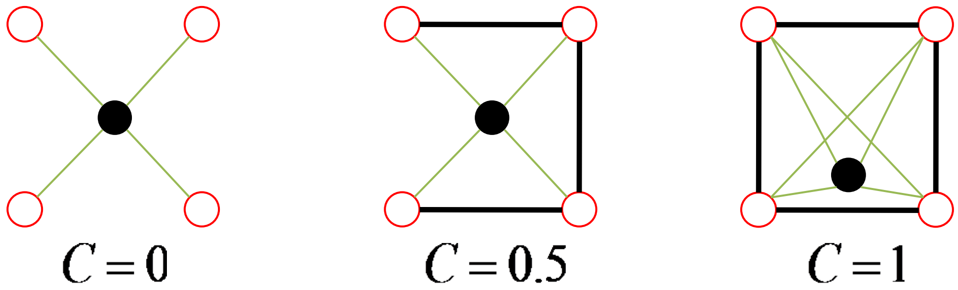
2-POINTS: *CLUSTERING COEFFICIENT*

- P(k',k''|k): cumbersome, difficult to estimate from data



Do your friends know each other ?

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



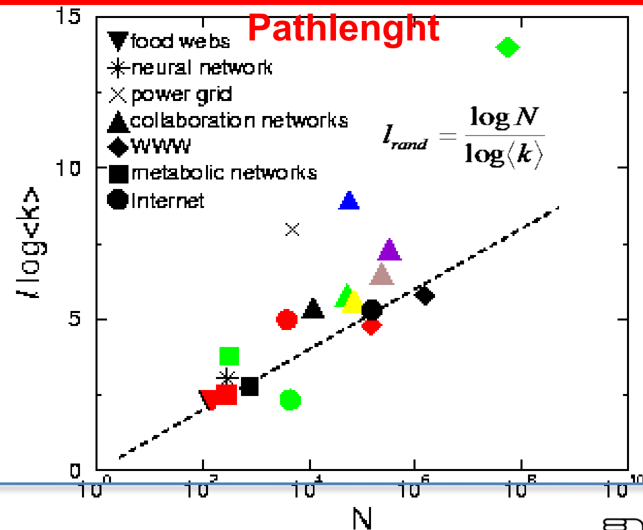
CORRELATIONS: CLUSTER SPECTRUM

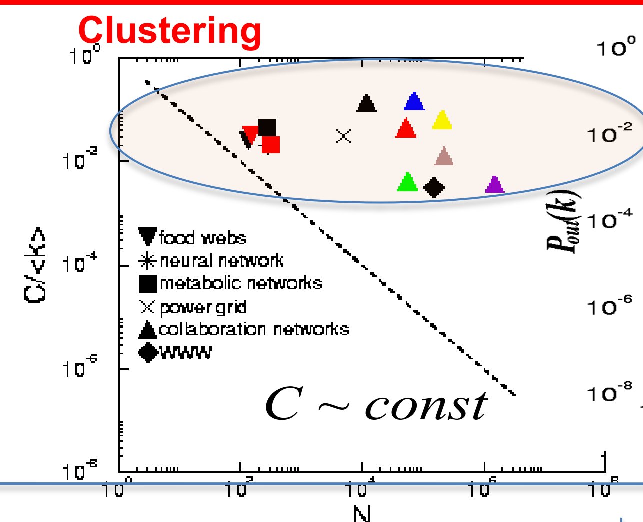
• Average clustering coefficient

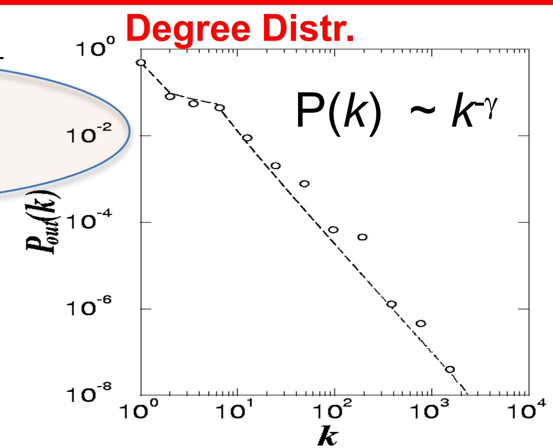
= average over nodes with very different characteristics

$$\overline{C} = \frac{1}{N}\sum_i C(i)$$

EMPIRICAL DATA FOR REAL NETWORKS







	Pathlength	Clustering	Degree Distr.
Regular network	$l \approx N^{1/D}$	$C \sim const$	$P(k)=\delta(k-k_d)$
Erdos-Renyi	$l_{rand} \approx \frac{\log N}{\log\langle k\rangle}$	$C_{rand} = p = \frac{\langle k\rangle}{N}$	$P(k) = e^{-\langle k\rangle}\frac{\langle k\rangle^k}{k!}$
Watts-Strogatz	$l_{rand} \approx \frac{\log N}{\log\langle k\rangle}$	$C \sim const$	*Exponential*
Barabasi-Albert	$l \approx \frac{\ln N}{\ln \ln N}$	$C \sim \frac{(\ln N)^2}{N}$	$P(k) \sim k^{-\gamma}$

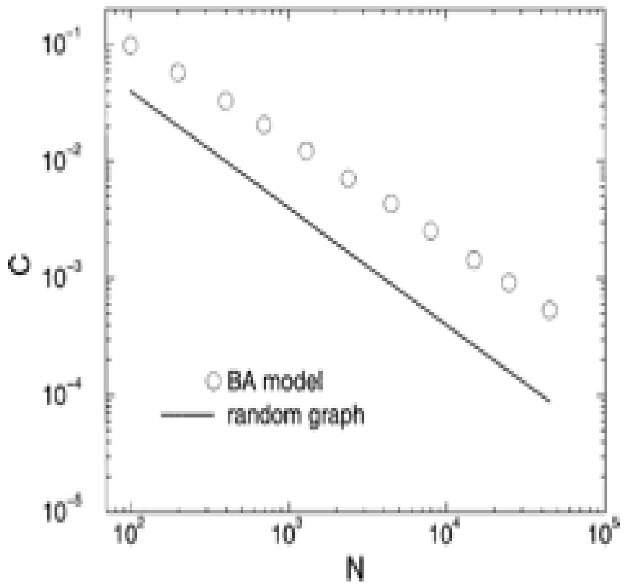
CLUSTERING COEFFICIENT OF THE BA MODEL

Reminder: for a random graph we have:

$$C_{rand} = \frac{<k>}{N} \sim N^{-1}$$

The numerical results indicate a *slightly* slower decay for BA network than for random networks.

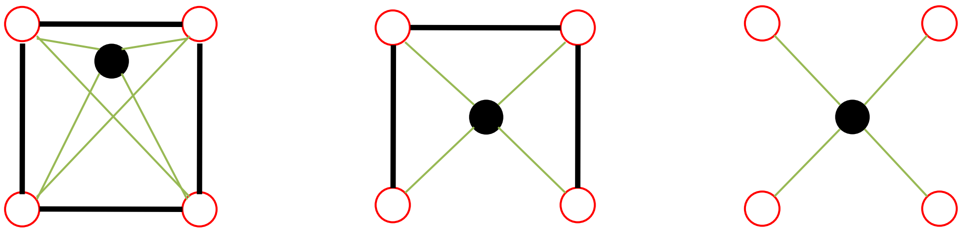
But not slow *enough*...



Clustering coefficient versus size of the Barabasi-Albert (BA) model with <k>=4, compared with clustering coefficient of random graph, $C_{rand} = \frac{<k>}{\mathrm{N}}$

Konstantin Klemm, Victor M. Eguiluz,
Growing scale-free networks with small-world behavior,
Phys. Rev. E 65, 057102 (2002), cond-mat/0107607

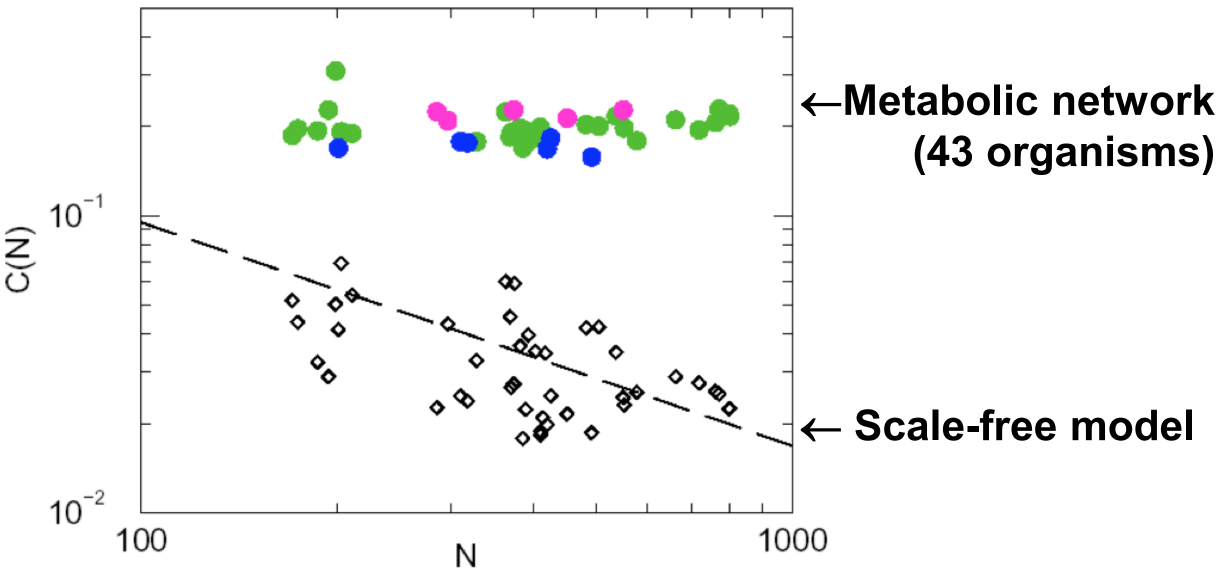
MODULARITY IN THE METABOLISM



$C = 1$ $\quad$ $C = \frac{1}{2}$ $\quad$ $C = 0$

Clustering Coefficient:

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



THE MEANING OF C(N)

Existence of a high degree of local modularity in real networks, that is not captured by the current models.

C(N)– the average number of triangles around each node in a system of size N.

The fact that C(N) does not decrease means that the relative number of triangles around a node remains constant as the system size increases—in contrast with the ER and BA models, where the relative number of triangles around a node decreases.
(here relative means relative to how many triangles we expected if all triangles that could be there would be there)

But C has some unexpected behavior, if we measure *C(k)*– the average clustering coefficient for nodes with degree *k*.

CORRELATIONS: CLUSTER SPECTRUM

- Average clustering coefficient

= average over nodes with very different characteristics

$$\overline{C} = \frac{1}{N}\sum_i C(i)$$

- Clustering spectrum:

putting together nodes which have the same degree

(link with hierarchical structures)

$$C(k) = \frac{1}{N_k}\sum_{i / k_i = k} C(i)$$

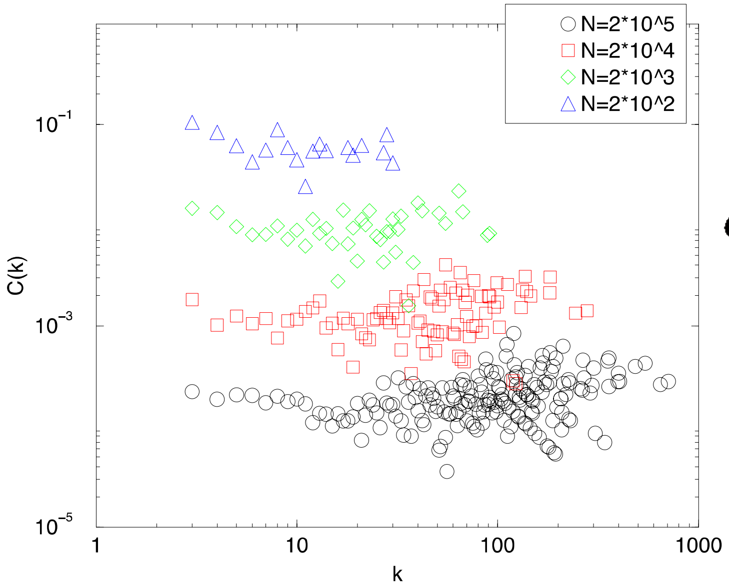
class of degree k

C(k) for the ER and BA models

Erdos-Renyi

$$C_{rand} = p = \frac{\langle k \rangle}{N}$$

Barabasi-Albert

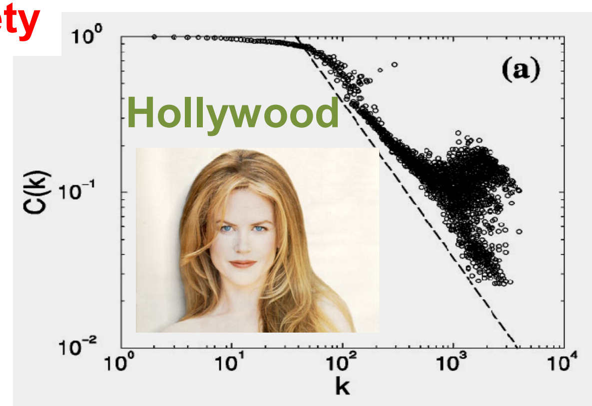


$$C \sim \frac{(\ln N)^2}{N}$$

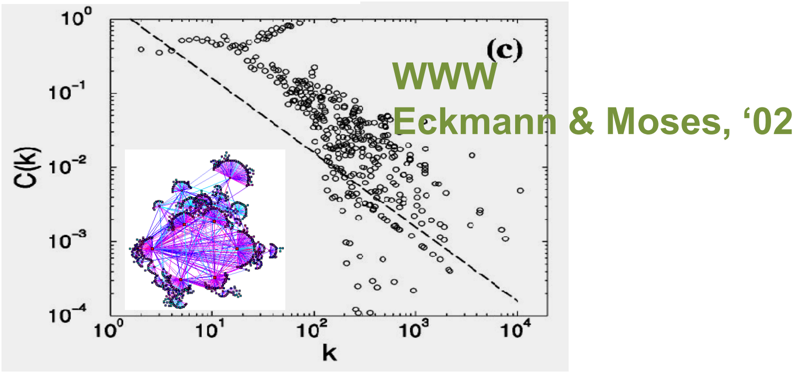
This is not true, however, for real networks. Let us look at some empirical data.

HIERARCHICAL NETWORKS

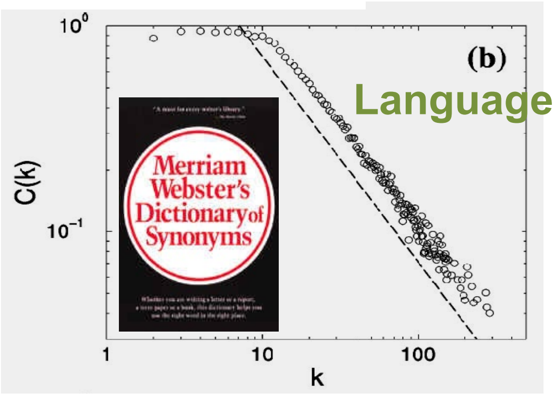
Society

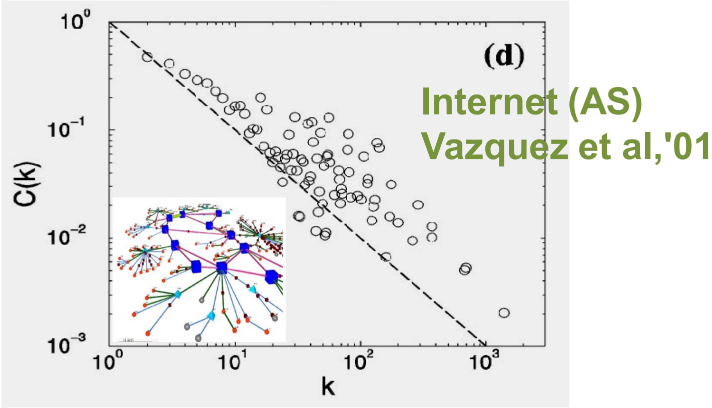


The electronic skin

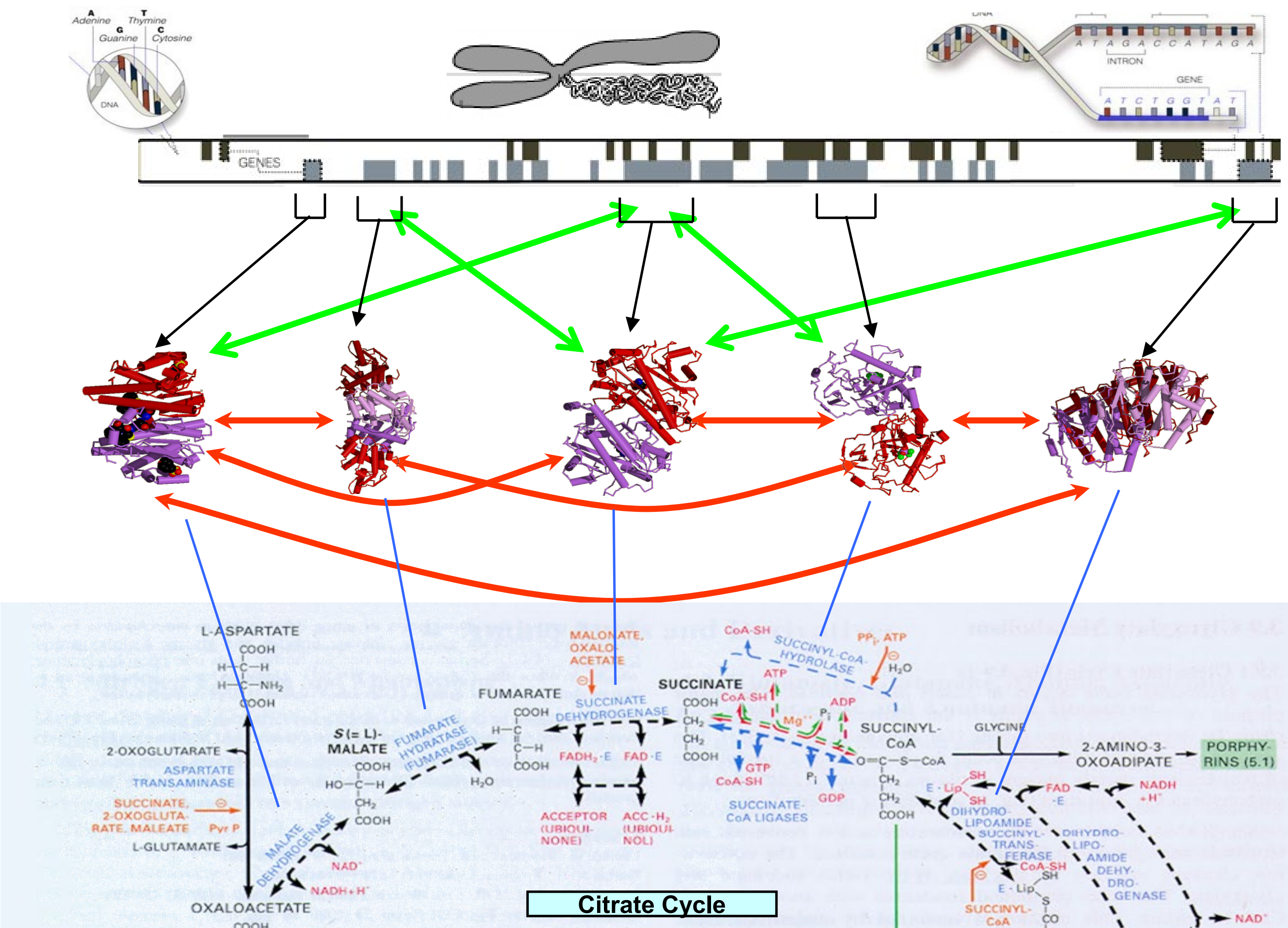


Human communication





Cellular networks:



GENOME

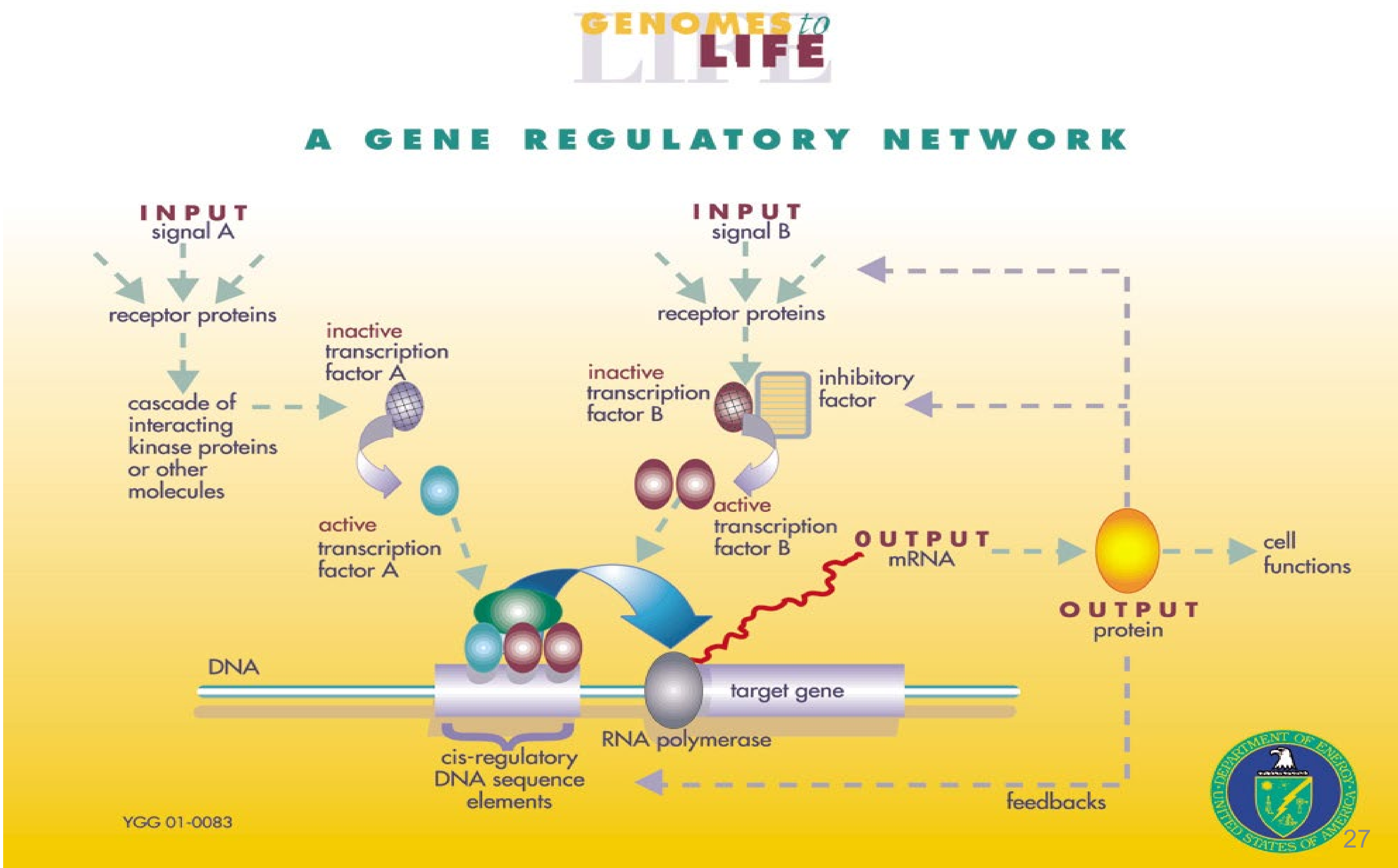
protein-gene interactions

PROTEOME

protein-protein interactions

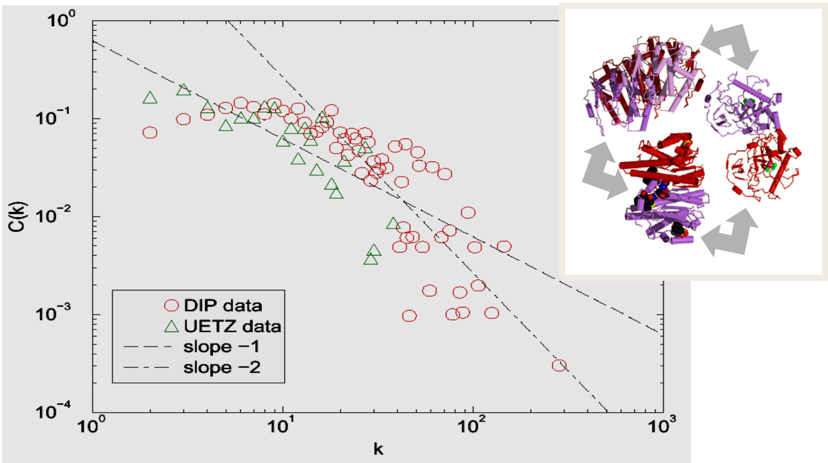
METABOLISM

Bio-chemical reactions

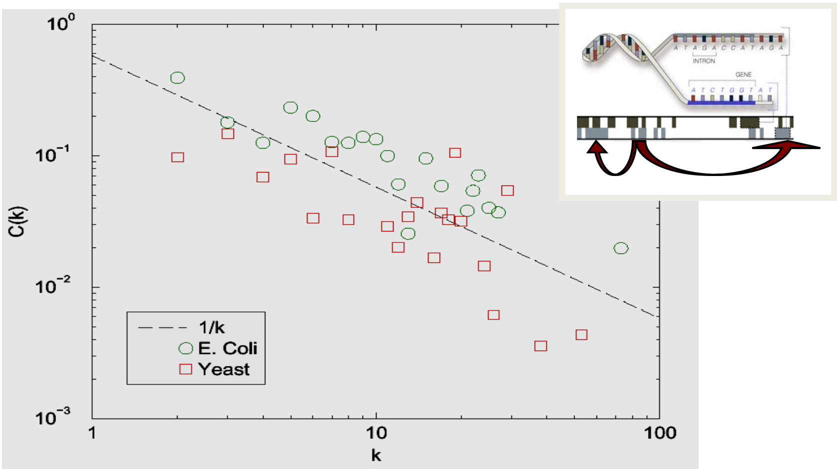

GENOMES to LIFE
A GENE REGULATORY NETWORK
INPUT signal A
receptor proteins
cascade of interacting kinase proteins or other molecules
inactive transcription factor A
active transcription factor A
INPUT signal B
receptor proteins
inactive transcription factor B
inhibitory factor
active transcription factor B
OUTPUT mRNA
OUTPUT protein
cell functions
DNA
target gene
RNA polymerase
cis-regulatory DNA sequence elements
feedbacks
YGG 01-0083

BIOLOGICAL SYSTEMS

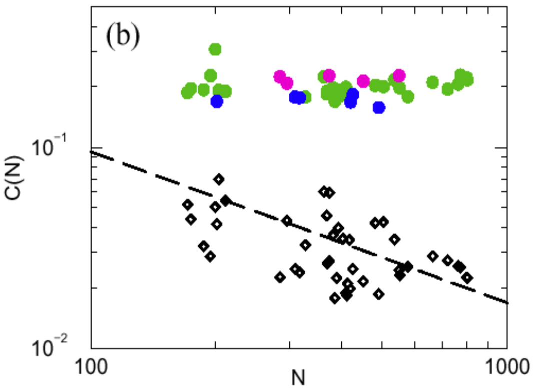
Protein-protein interaction

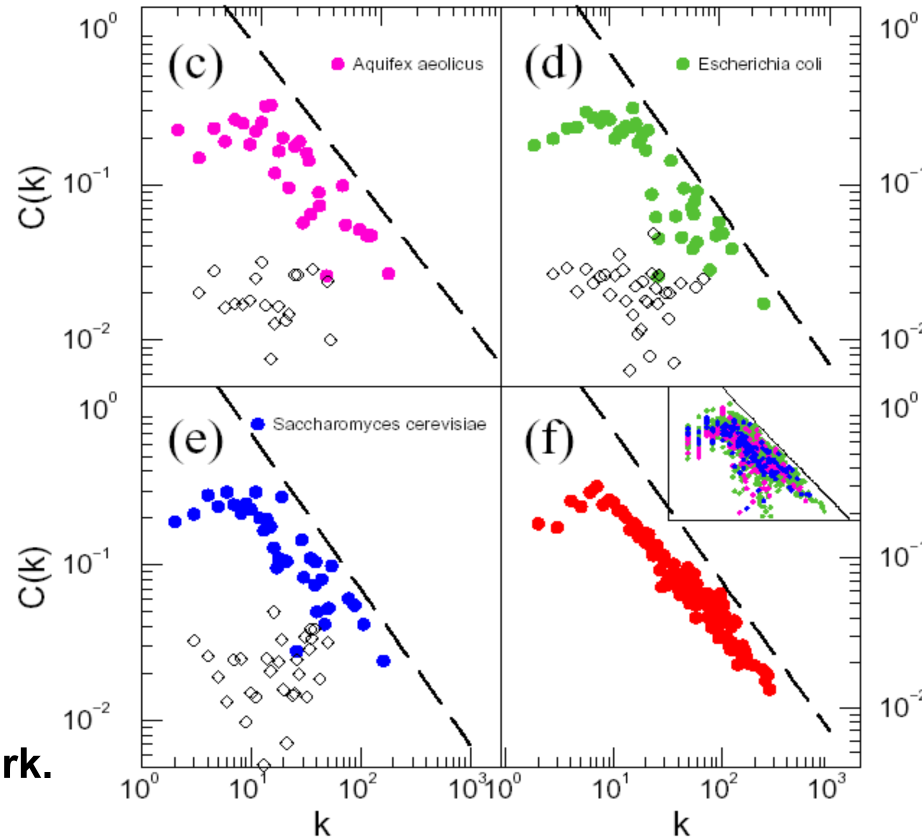


Regulatory networks



SCALING OF THE CLUSTERING COEFFICIENT C(k)



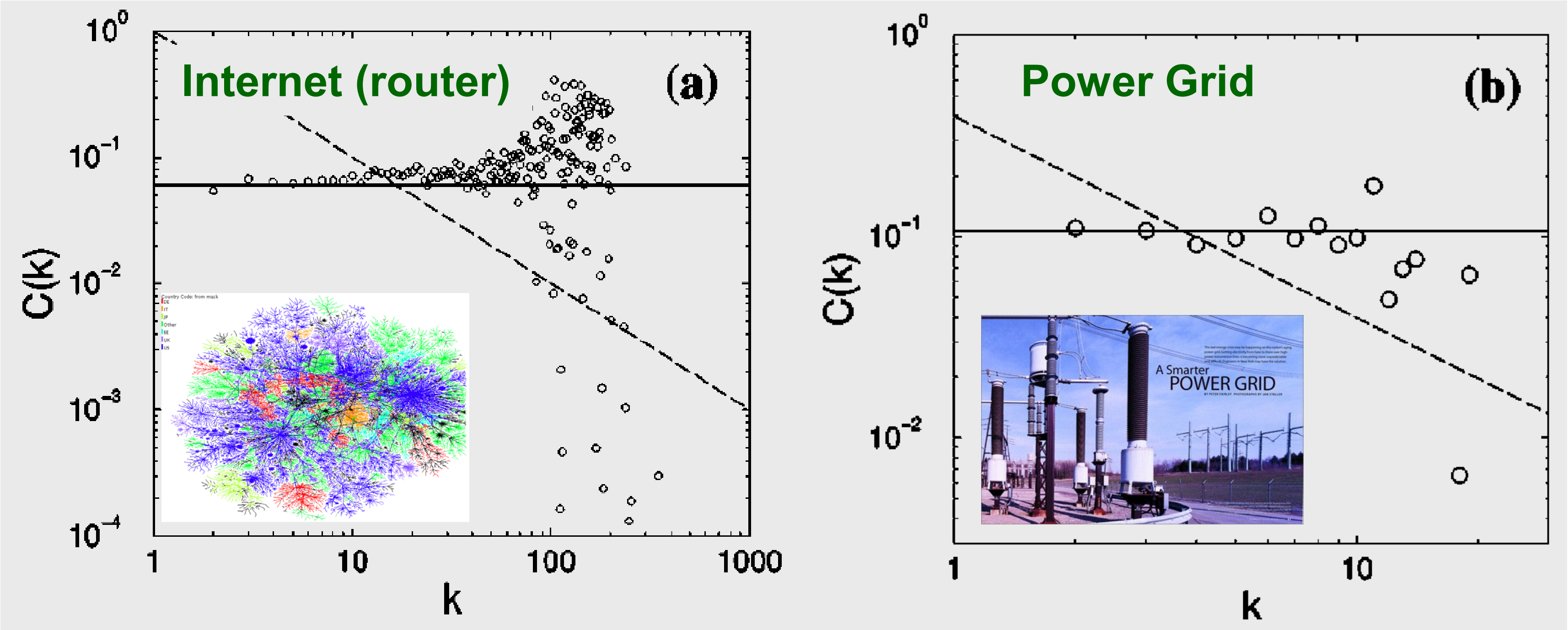


The metabolism forms a hierarchical network.

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

ABSENCE OF HIERARCHY

Geographically localized networks



SUMMARY OF EMPIRICAL RESULTS

	$C(k) \sim k^{-\beta}$	$C(k)$ indep. of k
Real systems	Internet (AS) WWW Metabolism Protein interaction network Regulatory network Language	Internet (router) Power grid
Models	?	ER model WS model BA model

But there is a deeper issue as stake, that need to consider– that of modularity.

HIERARCHICAL EXPONENT

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

Is the exponent universal?

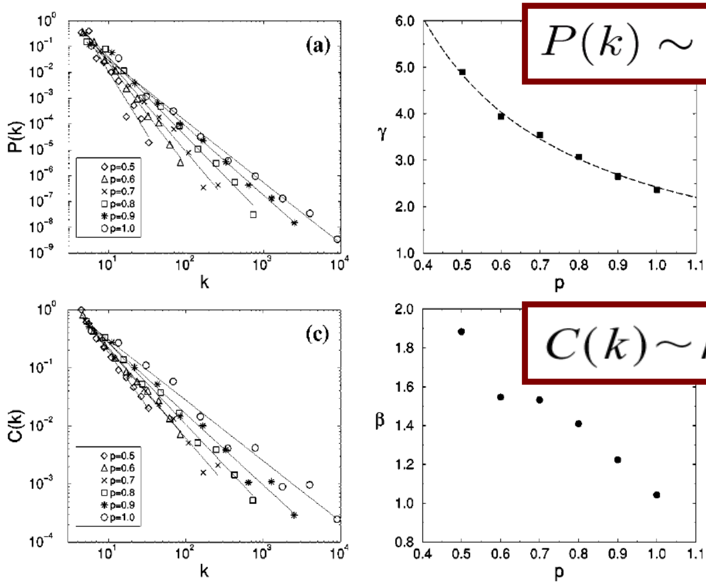
Or could we have for example: $C(k) \sim k^{-\beta}$

STOCHASTIC VERSION

Randomly pick a *p fraction of the newly added nodes and connect each of them independently to the nodes belonging to the central module.*

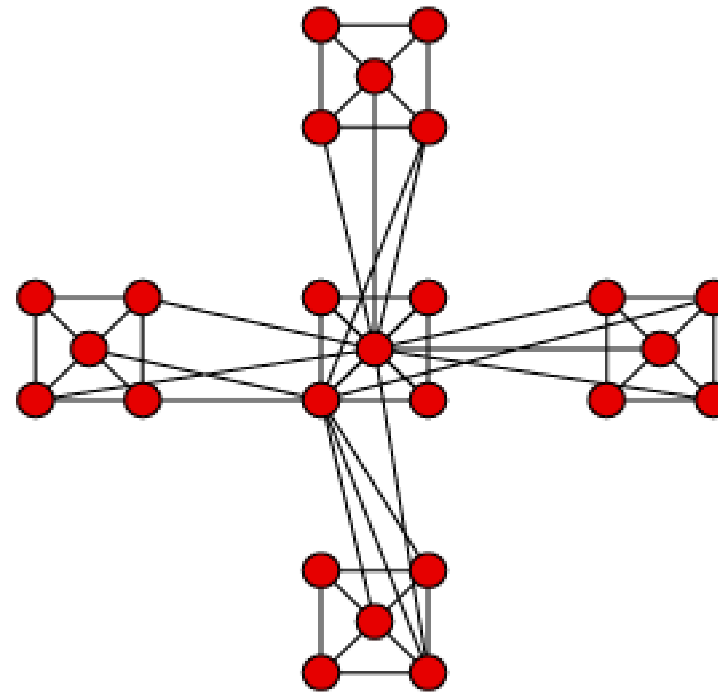
-use preferential attachment to decide, to which central node the selected nodes link to.

-at the next level p^2 fraction will link, back, then p^3, …p^i



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

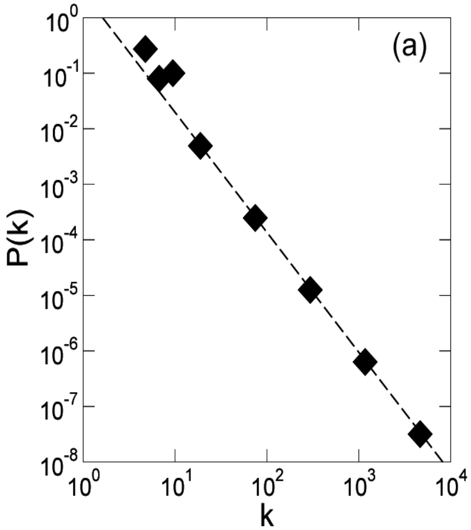
$$C(k) \sim k^{-\beta}$$



SUMMARY

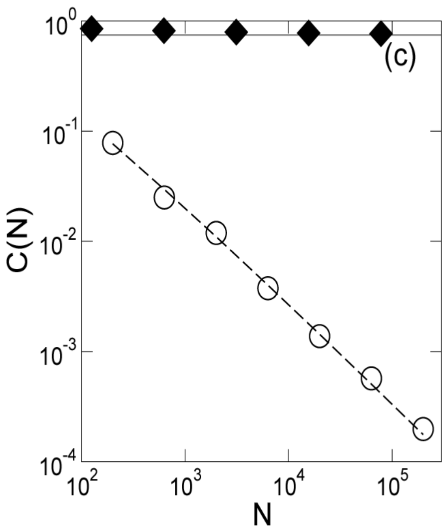
1. Scale-free

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



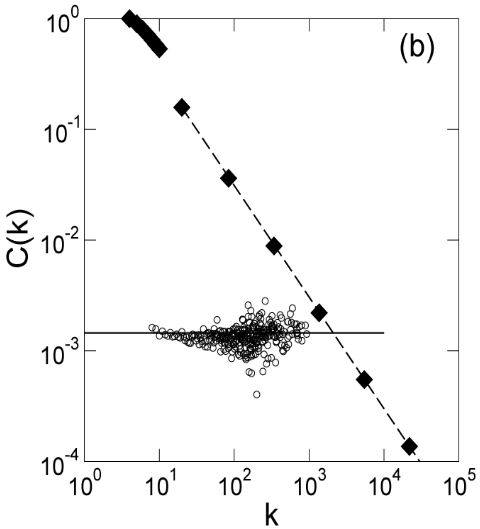
2. Clustering coefficient independent of N

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



3. Clustering spectrum

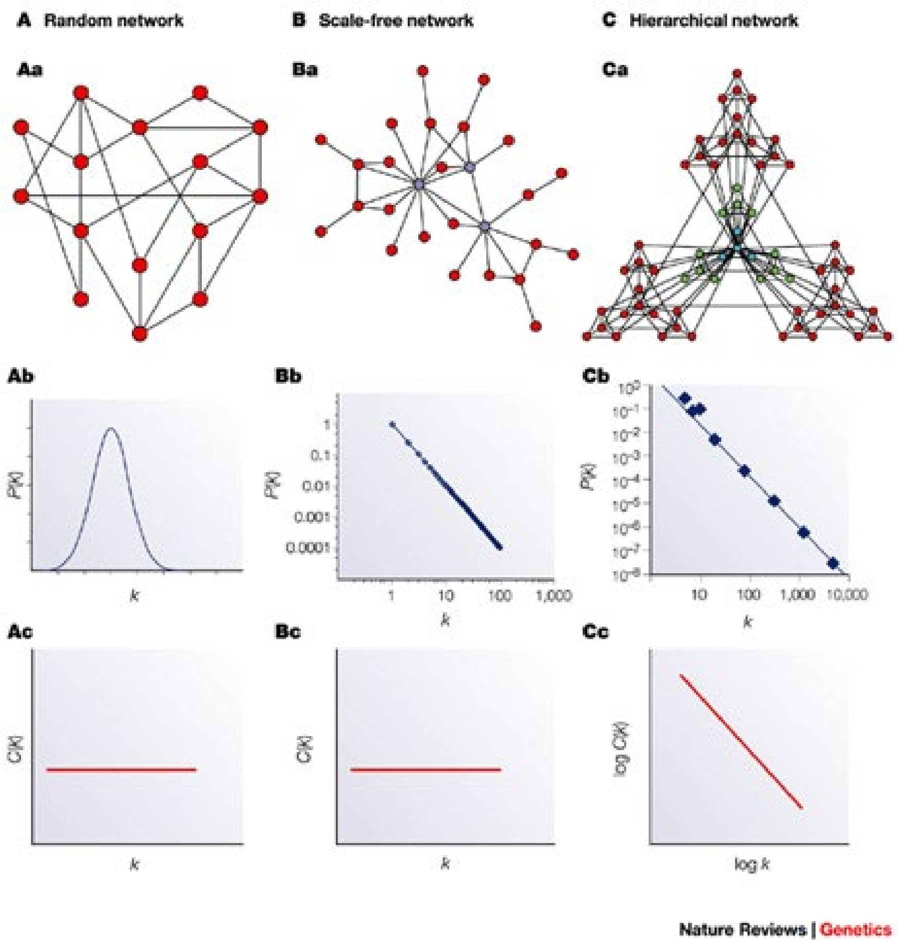
$$C(k) \sim k^{-1}$$



In real systems *C(k)* does not always decrease as a power law.
What matters, however, that it *decreases*, i.e. it *is not independent of k*.

THE BIG PICTURE

Hierarchy is a new rather generic network property.



A.-L. Barabasi and Z.N. Oltvai, Nat. Rev. Gen.(2004)

What does happen in real systems? Is a prediction that all systems with γ<3 should be automatically dissasortative, or have a cutoff – is this the case?
Let's see: www, γ=2.1, no cutoff, dissasortative NICE
Actor network, no cutoff, but it is ASSORTATIVE (how is this possible?).
Internet: γ=2.5, disassortative, cutoff , NICE

Networks with γ<3 don't have to be assortative:
Lets suppose we have a neutral network. High assortativity means a high degree nodes neighbors have high average degree. If we want to make it assortative we have to increase the degree of the neighbors of hubs. Even if the degree of the top neighbors cannot be increased because we used up all of the hubs, the low degree neighbors still can be replaced with higher ones, thus making the network assortative.
Anyway, the social networks checked (actor network, coauthorship network) have cut-offs according to Newman and Stanley.
http://samoa.santafe.edu/media/workingpapers/00-07-037.pdf
http://viseu.chem-eng.northwestern.edu/site_media/publication_pdfs/Amaral-2000-Proc.Natl.Acad.Sci.U.S.A.-97-11149.pdf

Static model used for examples

- Start with N unconnected nodes.
- Assign a w_i weight to each node i.
- Randomly select two nodes with probability proportional to w_i. Connect these nodes. Repeat L times.

If $w_i = \frac{1}{i^\alpha} \quad \rightarrow \quad p_k \sim k^{-1-1/\alpha}$

Upper cut-off may be added by introducing i_0: $w_i = \frac{1}{(i+i_0)^\alpha}$

For large N this should be equivalent to the configuration model.

Static model used for examples

- Start with N unconnected nodes.
- Assign a w_i weight to each node i.
- Randomly select two nodes with probability proportional to w_i. Connect these nodes. Repeat L times.

If $w_i = \frac{1}{i^\alpha} \quad \rightarrow \quad p_k \sim k^{-1-1/\alpha}$

Upper cut-off may be added by introducing i_0: $w_i = \frac{1}{(i+i_0)^\alpha}$

For large N this should be equivalent to the configuration model.

MALLOY-REED CRITERIA: THE EXISTENCE OF A GIANT COMPONENT

A giant cluster exists if each node is connected to at least two other nodes.

The average degree of a node i linked to the GC, must be 2, i.e.

$$< k_m \mid i \leftrightarrow j >= \sum_{k_m} k_m P(k_m \mid i \leftrightarrow j) = 2$$

$$P(k_m \mid i \leftrightarrow j) = \frac{P(k_m, i \leftrightarrow j)}{P(i \leftrightarrow j)} = \frac{P(i \leftrightarrow j \mid k_m) P(k_m)}{P(i \leftrightarrow j)}$$

Bayes' theorem

$P(k_m|i \leftrightarrow j)$: joint probability that a node has degree k_m and is connected to nodes i and j.

For a randomly connected network (does NOT mean random network!) with P(k):

$$P(i \leftrightarrow j) = \frac{2L}{N(N-1)} = \frac{<k>}{N-1} \qquad P(i \leftrightarrow j \mid k_m) = \frac{k_m}{N-1}$$

i **can choose between N-1 nodes to link to, each with probability 1/(N-1). I can try k_i times.**

$$\sum_{k_m} k_m P(k_m \mid i \leftrightarrow j) = \sum_{k_m} k_m \frac{P(i \leftrightarrow j \mid k_m) P(k_m)}{P(i \leftrightarrow j)} = \sum_{k_m} k_m \frac{k_m P(k_m)}{<k>} = \frac{\sum_{k_m} k_m^2 P(k_m)}{<k>}$$

$$\boxed{\kappa \equiv \frac{<k^2>}{<k>} = 2}$$

$\kappa>2$: a giant cluster exists

$\kappa<2$: many disconnected clusters

Malloy, Reed, Random Structures and Algorithms (1995); Cohen et al., *Phys. Rev. Lett*. **85**, 4626 (2000).

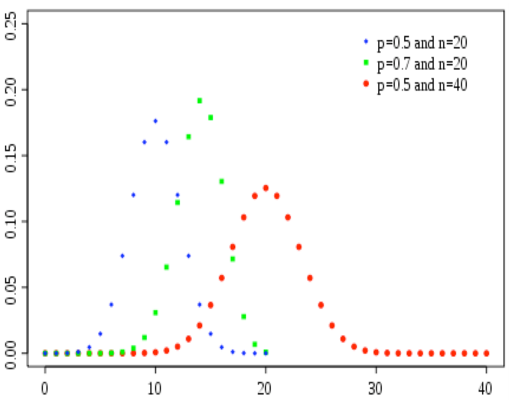
Apply the Malloy-Reed Criteria to an Erdos-Renyi Network

Probability Distribution Function (PDF)

Discrete Formulation

-binomial distribution-

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



$$<k> = (N-1)p$$

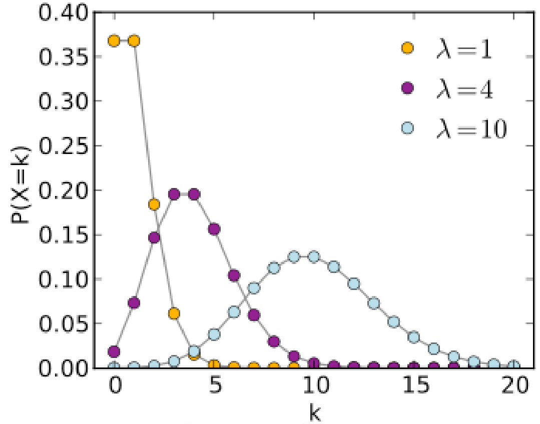
$$<k^2> = p(1-p)(N-1) + p^2(N-1)^2$$

$$\sigma_k = (<k^2> - <k>^2)^{1/2} = [p(1-p)(N-1)]^{1/2}$$

Continuum Formulation

-Poisson distribution-

$$P(k) = e^{-<k>} \frac{<k>^k}{k!}$$



$$<k> = <k>$$

$$<k^2> = <k>(1 + <k>)$$

$$\sigma_k = (<k^2> - <k>^2)^{1/2} = <k>^{1/2}$$

Apply the Malloy-Reed Criteria to an Erdos-Renyi Network

A giant cluster exists if each node is connected to at least two other nodes.

$$\kappa \equiv \frac{<k^2>}{<k>} = 2$$

κ>2: a giant cluster exists;

κ<2: many disconnected clusters;

$$<k> = <k>$$

$$<k^2> = <k>(1+<k>)$$

$$\sigma_k = (<k^2> - <k>^2)^{1/2} = <k>^{1/2}$$

$$\kappa \equiv \frac{<k^2>}{<k>} = \frac{<k>(1+<k>)}{<k>} = 1+<k> = 2$$

$$<k> = 1$$

Malloy-Reed; Cohen et al., Phys. Rev. Lett. 85, 4626 (2000).