

Frontiers of Network Science

Fall 2021

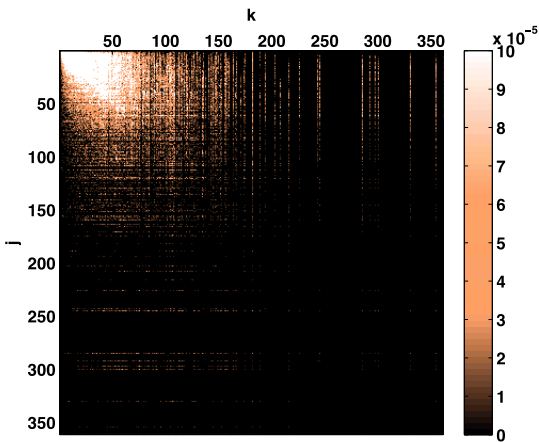
Class 17: Degree Correlations part II **(Chapter 7 in Textbook)**

Boleslaw Szymanski

based on slides by
Albert-László Barabási
and Roberta Sinatra

PROBLEM WITH THE FULL STATISTICAL DESCRIPTION

(1) Difficult to extract information from a visual inspection of a matrix.



(2) Based on e_{jk} and hence requires a large number of elements to inspect:

$$\frac{k_{\max}(k_{\max}-1)}{2} - 1 - k_{\max}$$

Nr. of independent elements

Undirected network:
 $k_{\max} \times k_{\max}$ matrix

$\sum_{j,k} e_{jk} = 1$

$\sum_{j=1, k_{\max}} e_{jk} = q_k$

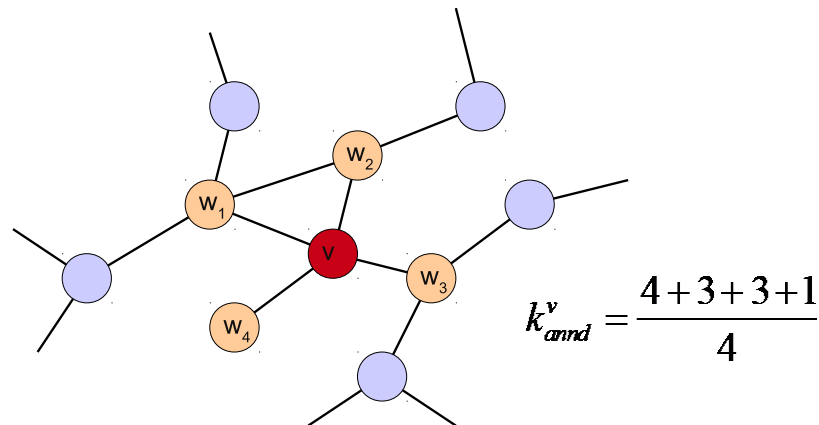
Constraints

We need to find a way to reduce the information contained in e_{jk}

Average next neighbor degree

$k_{ann}(k)$: average degree of the first neighbors of nodes with degree k .

$$k_{ann}(k) = \sum_{k'} k' P(k' | k) = \frac{\sum_{k'} k' e_{kk'}}{\sum_{k'} e_{kk'}}$$

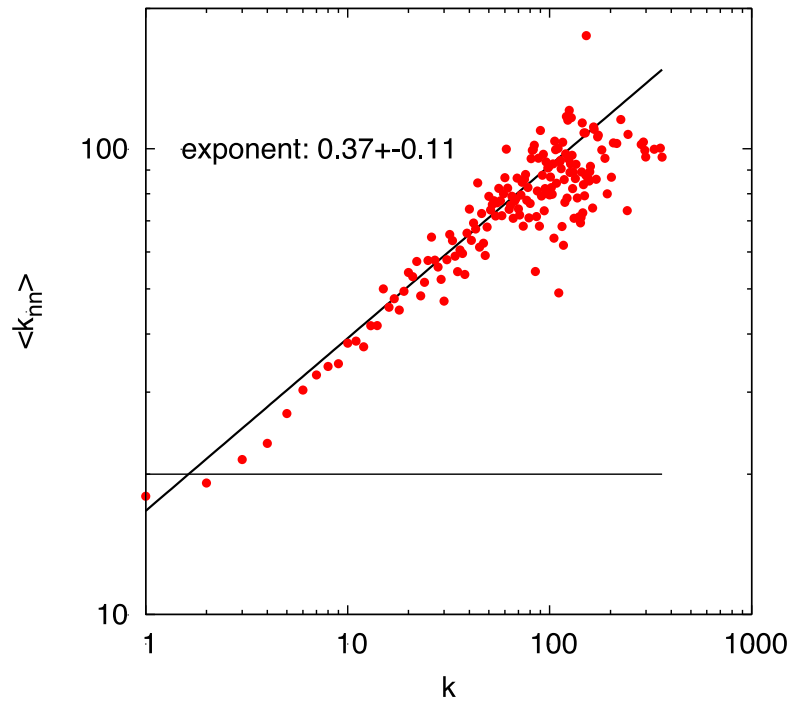


No degree correlations:

$$k_{ann}(k) = \frac{\sum_{k'} k' e_{kk'}}{\sum_{k'} e_{kk'}} = \frac{\sum_{k'} k' q_k q_{k'}}{q_k} = \sum_{k'} k' q_{k'} = \sum_{k'} k' \frac{k' p(k')}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

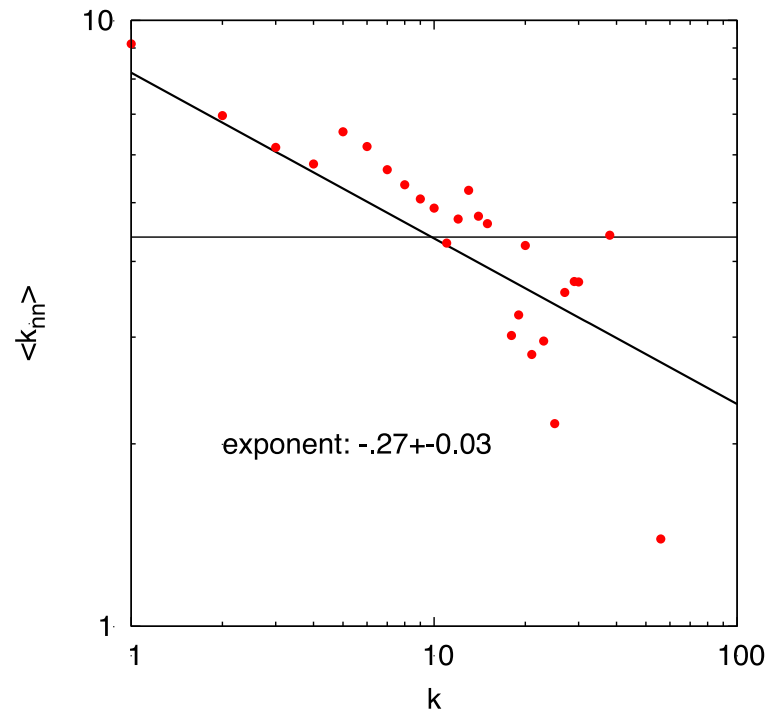
If there are no degree correlations, $k_{ann}(k)$ is independent of k .

$k_{\text{ann}}(k)$ FOR REAL NETWORKS



Astrophysics co-authorship network

Assortative



Yeast PPI

Disassortative

Average next neighbor degree

$k_{annd}(k)$: average degree of the first neighbors of nodes with degree k .

constraint:
$$\sum_k k_{annd}(k) \cdot k N p_k = \sum_k k^2 \cdot N p_k$$
$$\langle k_{annd}(k) k \rangle = \langle k^2 \rangle$$
 $\longrightarrow$ $k_{max}-1$ independent elements

$k_{annd}(k)$ is a k -dependent function, hence it has much fewer parameters,
and it is easier to interpret/read.

PEARSON CORRELATION

If there are degree correlations, e_{jk} will differ from $q_j q_k$. The magnitude of the correlation is captured by $\langle jk \rangle - \langle j \rangle \langle k \rangle$ difference, which is:

$$\sum_{jk} jk(e_{jk} - q_j q_k)$$

$\langle jk \rangle - \langle j \rangle \langle k \rangle$ is expected to be:

- positive for *assortative* networks,
- zero for *neutral* networks,
- negative for *dissortative* networks

To compare different networks, we should normalize it with its maximum value; the maximum is reached for a *perfectly assortative network*, i.e. $e_{jk} = q_k \delta_{jk}$

normalization:

$$\sigma_r^2 = \max \sum_{jk} jk(e_{jk} - q_j q_k) = \sum_{jk} jk(q_k \delta_{jk} - q_j q_k)$$
$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_r^2} \quad -1 \leq r \leq 1$$

$r \leq 0$	<i>disassortative</i>
$r = 0$	<i>neutral</i>
$r \geq 0$	<i>assortative</i>

REAL NETWORKS

Social networks are <i>assortative</i>	Physics coauthorship (a)	52 909	0.363	Biological, technological networks are <i>disassortative</i>
	Biology coauthorship (a)	1 520 251	0.127	
	Mathematics coauthorship (b)	253 339	0.120	
	Film actor collaborations (c)	449 913	0.208	
	Company directors (d)	7 673	0.276	
Internet (e)	10 697	−0.189		
World-Wide Web (f)	269 504	−0.065		
Protein interactions (g)	2 115	−0.156		
Neural network (h)	307	−0.163		
Marine food web (i)	134	−0.247		
Freshwater food web (j)	92	−0.276		
Random graph (u)		0		
Callaway <i>et al.</i> (v)		$\delta/(1 + 2\delta)$		
Barabási and Albert (w)		0		

$r > 0$: assortative network:

Hubs tend to connect to other hubs.

$r < 0$: disassortative network:

Hubs tend to connect to small nodes.

RELATIONSHIP BETWEEN r AND k_{annd}

$$r = \frac{\sum_{kj} kj(e_{kj} - q_k q_j)}{\sigma_r^2} = \frac{\sum_k k q_k \sum_j \frac{j e_{kj}}{q_k} - \left(\sum_k k q_k \right)^2}{\sigma_r^2} = \frac{\sum_k k k_{annd}(k) q_k - \frac{\langle k^2 \rangle^2}{\langle k \rangle^2}}{\sigma_r^2}$$

$$k_{annd}(k) = \sum_{k'} k' P(k' | k) = \frac{\sum_{k'} k' e_{kk'}}{\sum_{k'} e_{kk'}} = \frac{\sum_{k'} k' e_{kk'}}{q_k}$$

In general case we need to know q_k and $k_{annd}(k)$ to calculate r .

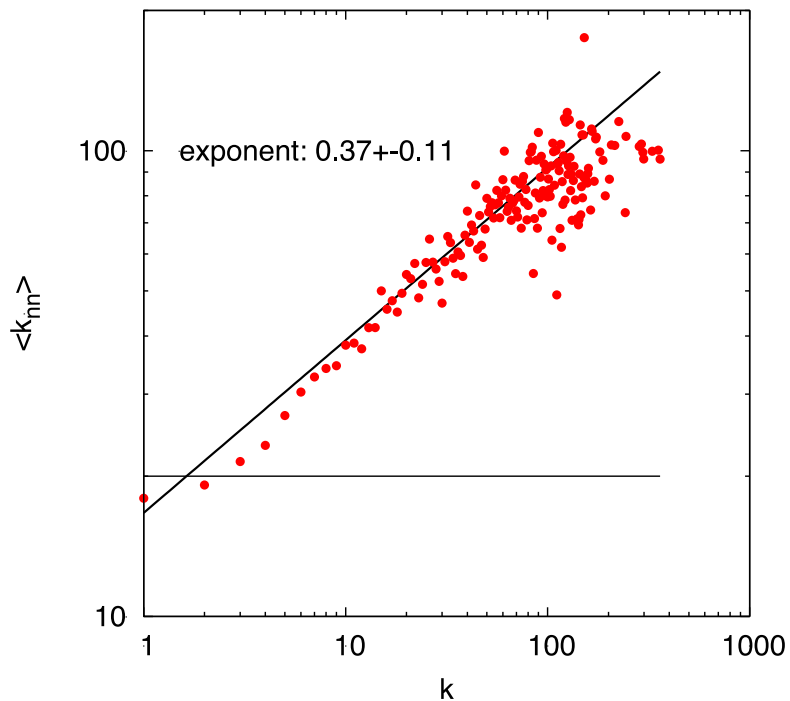
Assuming: $k_{annd}(k) = a \cdot k + b$

Using the constraint for **ANND**:

$$\langle k^2 \rangle = \langle k_{annd}(k) k \rangle = \sum_{k'} a \cdot k^2 p_k + b \cdot k p_k = a \langle k^2 \rangle + b \langle k \rangle \longrightarrow b = \frac{(1-a) \langle k^2 \rangle}{\langle k \rangle}$$

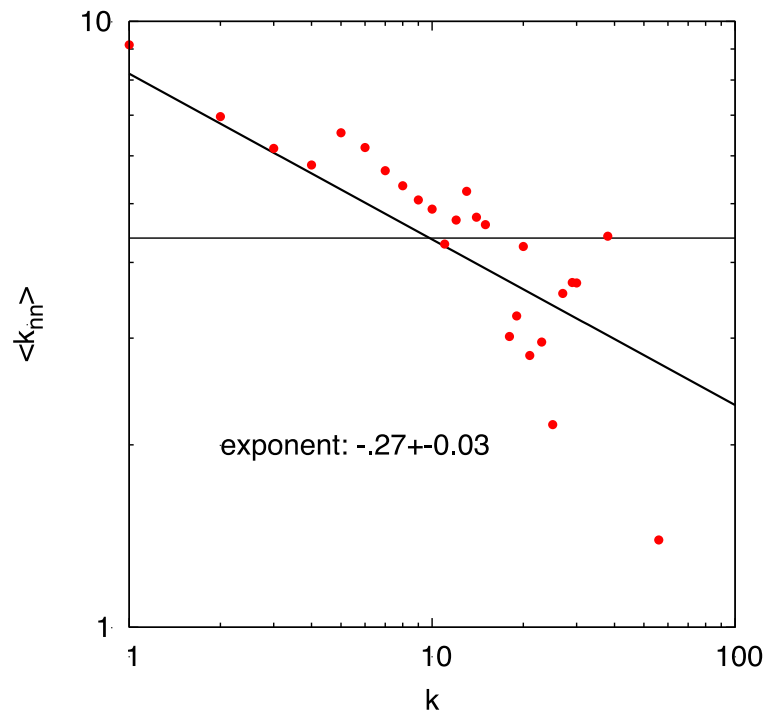
$$\begin{aligned} r &= \frac{\sum_k k \cdot (a \cdot k + b) q_k - \frac{\langle k^2 \rangle^2}{\langle k \rangle^2}}{\sigma_r^2} = \frac{\sum_k k \cdot \left(a \cdot k + \frac{(1-a) \langle k^2 \rangle}{\langle k \rangle} \right) k p_k - \frac{\langle k^2 \rangle^2}{\langle k \rangle^2}}{\sigma_r^2} = \\ &= \frac{a \left(\sum_k k^3 \frac{p_k}{\langle k \rangle} - \frac{\langle k^2 \rangle^2}{\langle k \rangle^2} \right) + \frac{\langle k^2 \rangle^2}{\langle k \rangle^2} - \frac{\langle k^2 \rangle^2}{\langle k \rangle^2}}{\sigma_r^2} = a \end{aligned}$$

PROBLEM WITH THE PREVIOUS DEVIATION: $k_{annd}(k) \sim k^\beta$



Astrophysics co-authorship network

Assortative



Yeast PPI

Disassortative

CONNECTION WITH ANND

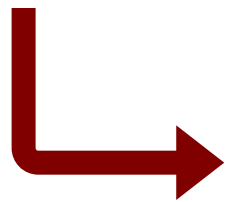
Assuming: $k_{annd}(k) = a \cdot k^\beta$

Using the constraint for **ANND**: $\langle k^2 \rangle = \langle k_{annd}(k)k \rangle = \sum_k a \cdot k^{\beta+1} p_k = a \langle k^{\beta+1} \rangle \longrightarrow a = \frac{\langle k^2 \rangle}{\langle k^{\beta+1} \rangle}$

$$r = \frac{\sum_k k \cdot a k^\beta \cdot q_k - \frac{\langle k^2 \rangle^2}{\langle k \rangle^2}}{\sigma_r^2} = \frac{\sum_k a \cdot k^{\beta+2} \frac{p_k}{\langle k \rangle} - \frac{\langle k^2 \rangle^2}{\langle k \rangle^2}}{\sigma_r^2} = \frac{\frac{\langle k^2 \rangle}{\langle k^{\beta+1} \rangle} \frac{\langle k^{\beta+2} \rangle}{\langle k \rangle} - \frac{\langle k^2 \rangle^2}{\langle k \rangle^2}}{\sigma_r^2} =$$

$$= \frac{1}{\sigma_r^2} \frac{\langle k^2 \rangle}{\langle k \rangle} \left(\frac{\langle k^{\beta+2} \rangle}{\langle k^{\beta+1} \rangle} - \frac{\langle k^2 \rangle}{\langle k \rangle} \right)$$

$$\sigma_r^2 = \sum_{jk} jk (q_k \delta_{jk} - q_j q_k) = \frac{\langle k^3 \rangle}{\langle k \rangle} - \frac{\langle k^2 \rangle^2}{\langle k \rangle^2}$$



$$\beta < 0 \rightarrow r < 0$$

$$\beta = 0 \rightarrow r = 0$$

$$\beta > 0 \rightarrow r > 0$$

CONNECTION BETWEEN R AND k_{ANND}

$$\beta=0: \quad \frac{\langle k^{\beta+2} \rangle}{\langle k^{\beta+1} \rangle} - \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle} - \frac{\langle k^2 \rangle}{\langle k \rangle} = 0 \Rightarrow r=0$$

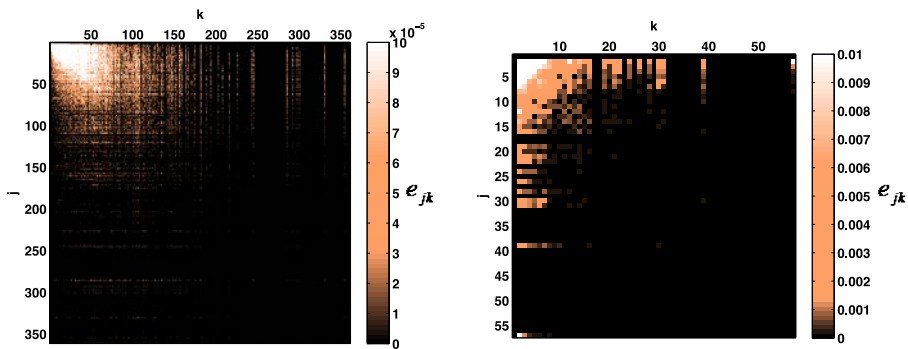
$$\begin{aligned} \langle k^{\alpha+\beta} \rangle &= \sum_{k_{\min}}^{k_{\max}} k^{\alpha+\beta} p_k \\ &< k_{\max}^{\beta} \sum_{k_{\min}}^{k_{\max}} k^{\alpha} p_k = k_{\max}^{\beta} \langle k^{\alpha} \rangle \\ &> k_{\min}^{\beta} \sum_{k_{\min}}^{k_{\max}} k^{\alpha} p_k = k_{\min}^{\beta} \langle k^{\alpha} \rangle \end{aligned}$$

$$0 > \beta > -1: \quad \frac{\langle k^2 \rangle}{\langle k \rangle} > \left(\frac{k_{\min}}{k_{\max}} \right)^{-\beta} \frac{\langle k^{\beta+2} \rangle}{\langle k^{\beta+1} \rangle} > \frac{\langle k^{\beta+2} \rangle}{\langle k^{\beta+1} \rangle} \Rightarrow r < 0$$

$$+1 > \beta > 0: \quad \frac{\langle k^{\beta+2} \rangle}{\langle k^{\beta+1} \rangle} > \left(\frac{k_{\min}}{k_{\max}} \right)^{\beta} \frac{\langle k^2 \rangle}{\langle k \rangle} > \frac{\langle k^2 \rangle}{\langle k \rangle} \Rightarrow r > 0$$

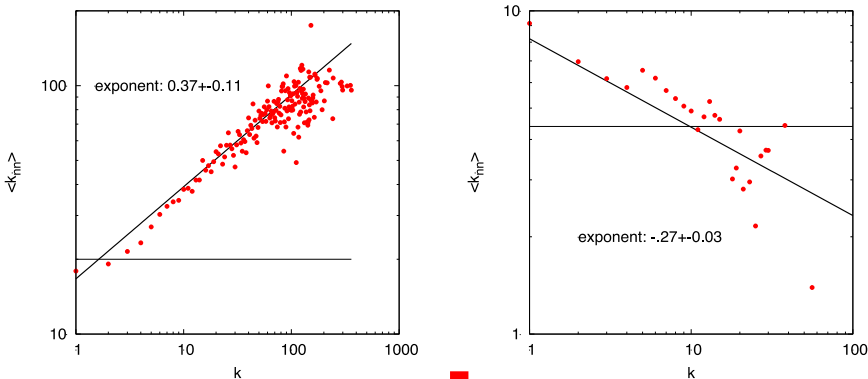
DEGREE CORRELATION IN NETWORKS

$$e_{jk}$$



$$\frac{k_{\max}(k_{\max}-1)}{2}-k_{\max}-1$$

$$k_{annd}(k)$$



$$k_{\max}-1$$

$$r$$

$$0.31$$

$$-0.16$$

$$1$$

GENERATING NETWORK WITH GIVEN ASSORTATIVITY

We have a desired e_{jk} distribution, which also specifies p_k .

1. Generate a network with the desired degree distribution using the configuration model.
2. Choose two links at random from the network: (v_1, w_1) and (v_2, w_2) .
3. Measure the degrees j_1, k_1, j_2, k_2 of nodes v_1, w_1, v_2, w_2 . Replace the two selected links with two new ones (v_1, v_2) and (w_1, w_2) with probability

$$P = \begin{cases} \frac{e_{j_1 j_2} e_{k_1 k_2}}{e_{j_1 k_1} e_{k_2 j_2}} & \text{if } e_{j_1 j_2} e_{k_1 k_2} < e_{j_1 k_1} e_{k_2 j_2} \\ 1 & \text{otherwise} \end{cases}$$

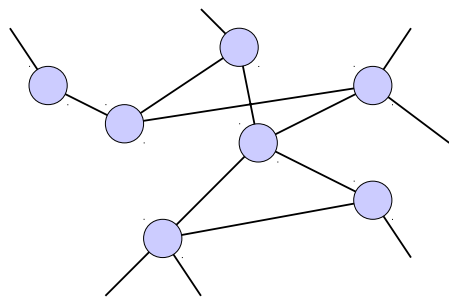
1. Repeat from step 2.

The algorithm is ergodic and satisfies detailed balance, therefore in the long time limit it samples the desired network ensemble correctly.

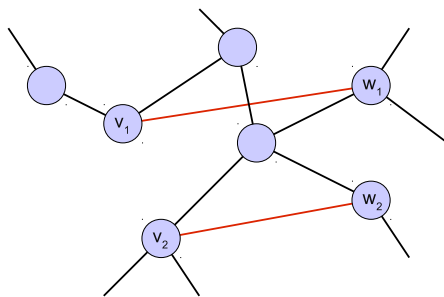
GENERATING NETWORK WITH GIVEN ASSORTATIVITY

2. Choose two edges random from the network: (v_1, w_1) and (v_2, w_2) .
3. Measure the degrees j_1, k_1, j_2, k_2 of vertices v_1, w_1, v_2, w_2 . Replace the two selected edges with two new ones (v_1, v_2) and (w_1, w_2) with probability

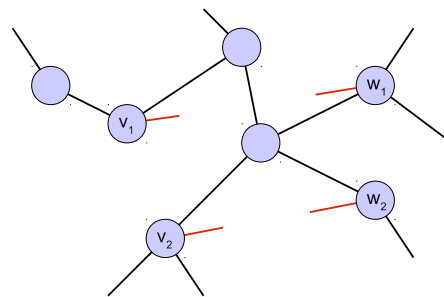
$$P = \begin{cases} \frac{e_{j_1 j_2} e_{k_1 k_2}}{e_{j_1 k_1} e_{k_2 j_2}} & \text{if } e_{j_1 j_2} e_{k_1 k_2} < e_{j_1 k_1} e_{k_2 j_2} \\ 1 & \text{otherwise} \end{cases}$$



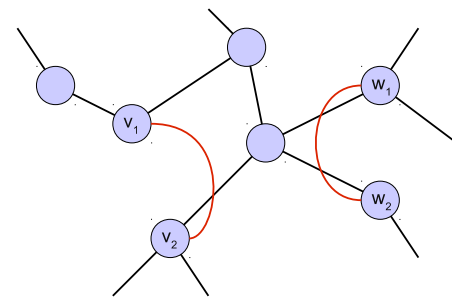
1



2



3



4

GENERATING NETWORK WITH GIVEN ASSORTATIVITY

If we only specify r we have great degree of freedom in choosing e_{jk} .

Possible choice for disassortative case:

$$e_{jk}^{(d)} = q_j x_k + x_j q_k - x_j x_k \quad \text{Where } x_k \text{ is any normalized distribution.}$$

This form satisfies the constraints on e_{jk} :

$$\sum_{jk} e_{jk} = \sum_{jk} q_k x_j + x_k q_j - x_k x_j = 1 + 1 - 1 = 1$$

$$\sum_j e_{jk} = \sum_j q_k x_j + x_k q_j - x_k x_j = q_k + x_k - x_k = q_k$$

The r value can be easily calculated:

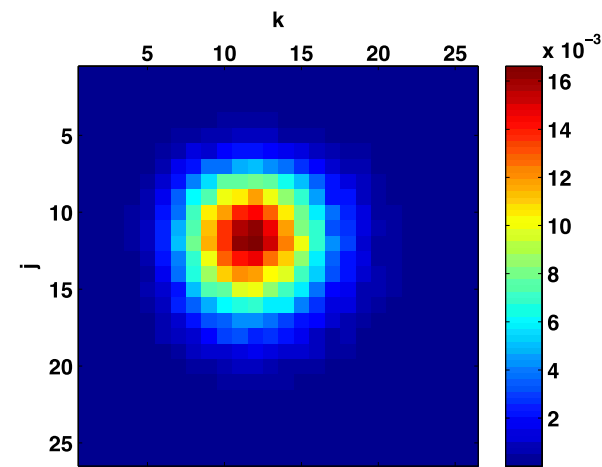
$$r_d = \frac{\sum_{jk} jk (q_k x_j + x_k q_j - x_k x_j - q_k q_j)}{\sigma_r^2} = \frac{2 \langle k \rangle_q \langle k \rangle_x - \langle k \rangle_x^2 - \langle k \rangle_q^2}{\sigma_r^2} = - \frac{(\langle k \rangle_x - \langle k \rangle_q)^2}{\sigma_r^2}$$

Assortative case:

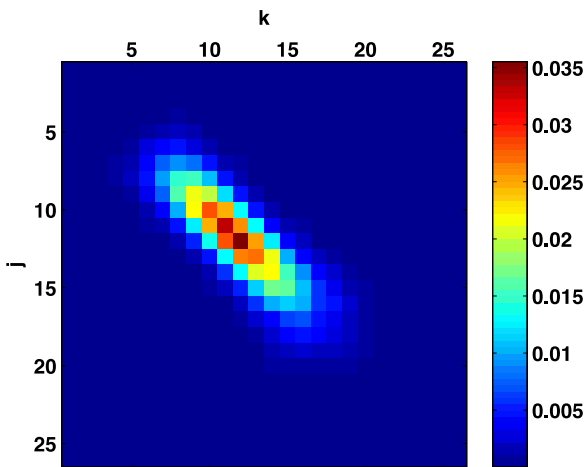
$$e_{jk}^{(a)} = q_j q_k - e_{jk}^{(d)} \longrightarrow r_a = -r_d$$

EXAMPLE: Erdős-Rényi

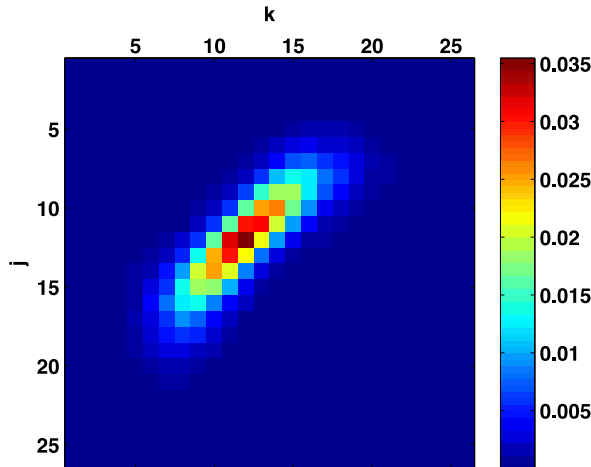
ER neutral



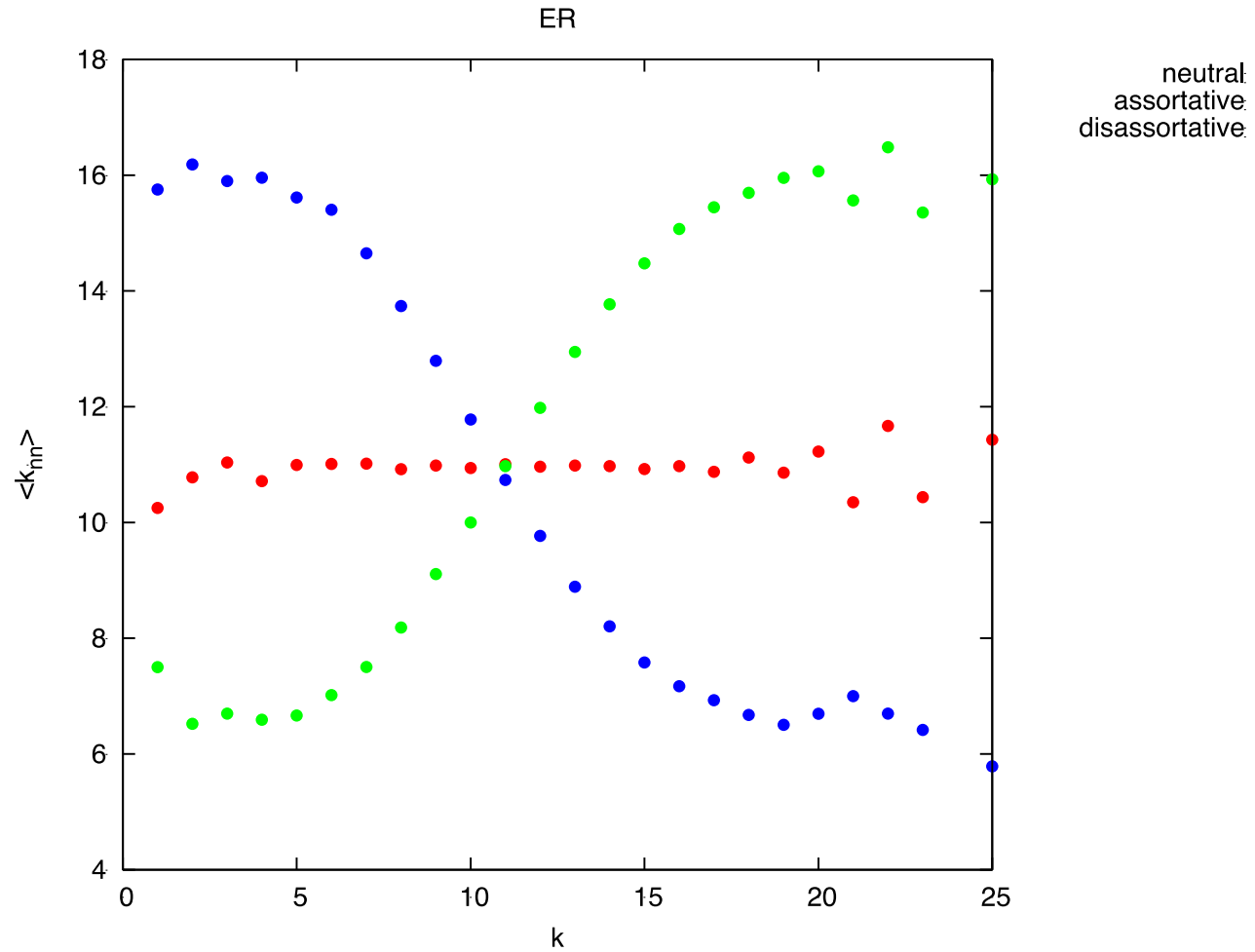
ER assortative



ER disassortative



EXAMPLE: Erdős-Rényi



Structural cut-off

High assortativity $\rightarrow$ 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' :

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

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 .

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.

Structural cut-off

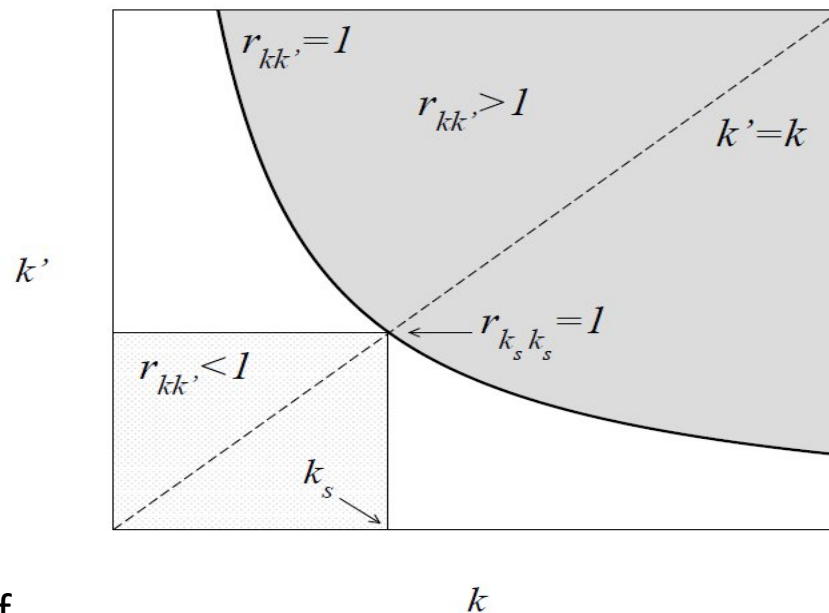
$$E_{kk'} = e_{kk'} \langle k \rangle N$$

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

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

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

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



Structural cut-off for uncorrelated networks

Uncorrelated networks:

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

$$m_{k_s k_s} = k_s N_{k_s} = k_s N p_{k_s}$$

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

$$e_{kk'} = q_k q_{k'} = \frac{kk' p_k p_{k'}}{\langle k \rangle^2} \longrightarrow r_{kk'} = \frac{E_{kk'}}{m_{kk'}} = \frac{\langle k \rangle N e_{kk'}}{m_{kk'}}$$

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

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

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

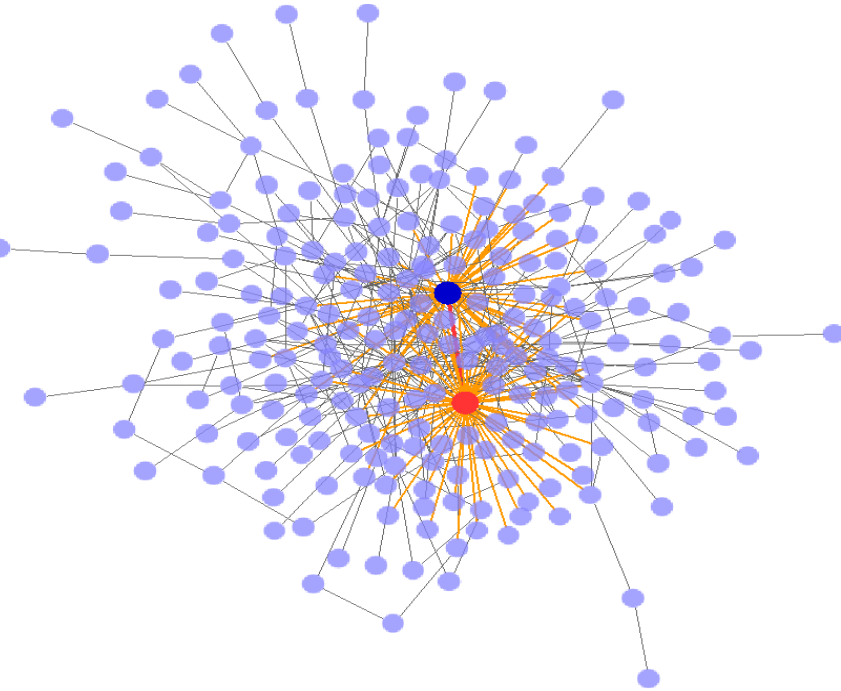
one cannot have nodes with degree larger than $k_s(N)$,

→ 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:

- Introduce a structural cutoff (i.e. do not allow nodes with $k > k_s(N)$)
- Let the network become more disassortative, having fewer links between hubs.

Example: Degree sequence introduces disassortativity



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

The measured $r=-0.19!$ → Dissortative!

Red hub: 55 neighbors.

Blue hub: 46 neighbors.

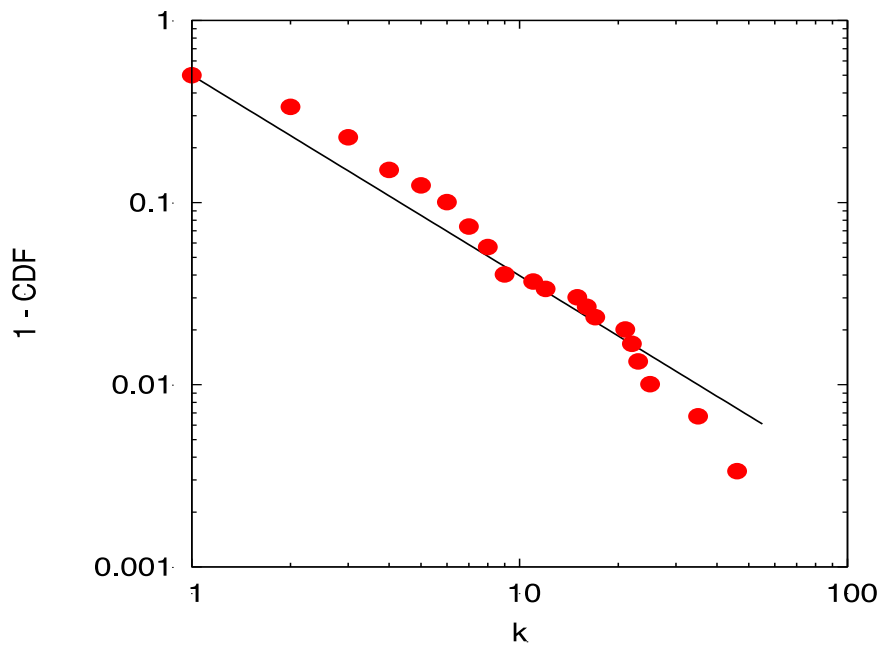
Let's calculate the expectation number of links between red node ($k=55$) and blue node ($k=46$) for uncorrelated networks!

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

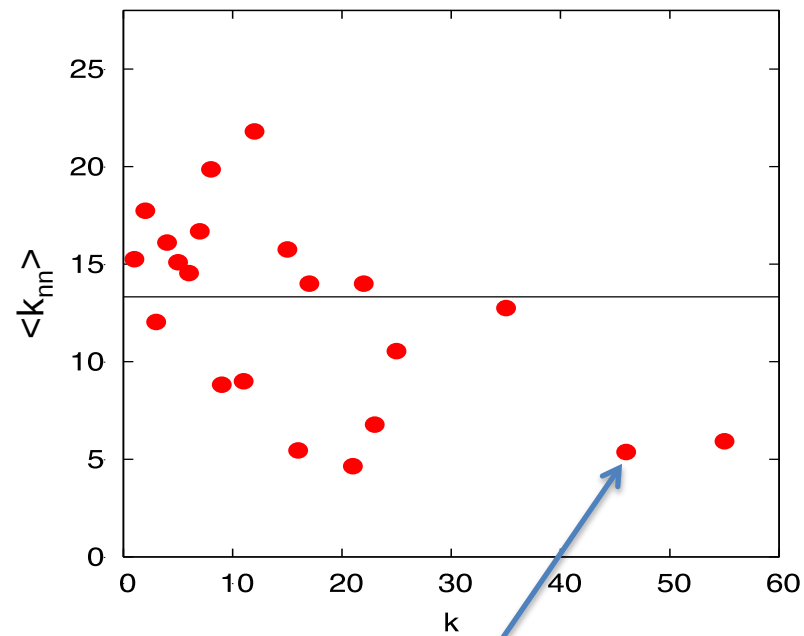
$$E_{55,46} = \langle k \rangle N \cdot e_{55,46} = 900 \cdot \frac{55 \frac{1}{300} \cdot 46 \frac{1}{300}}{3^2} \approx 2.8 > 1$$

Diagram illustrating the calculation of the expected number of links between two hubs in an uncorrelated network. The formula is shown with arrows indicating the variables: k points to 55, k' points to 46, P_k points to $\frac{1}{300}$, $P_{k'}$ points to $\frac{1}{300}$, and $\langle k \rangle$ points to 900.

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

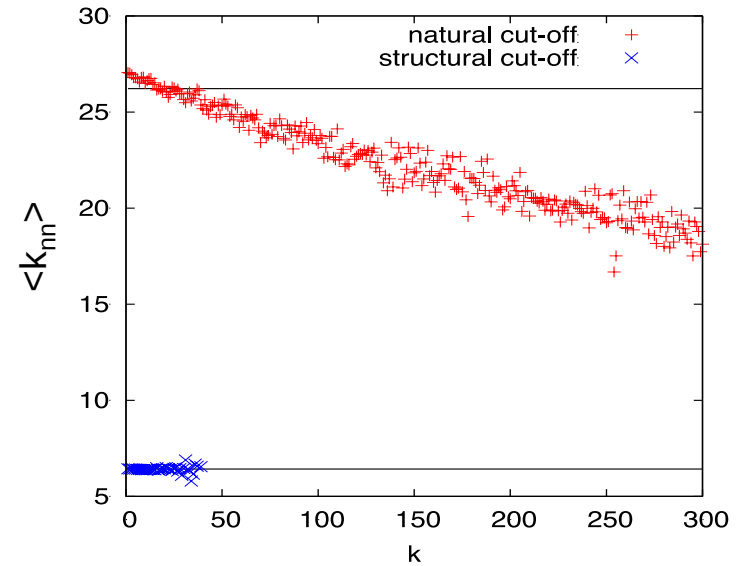
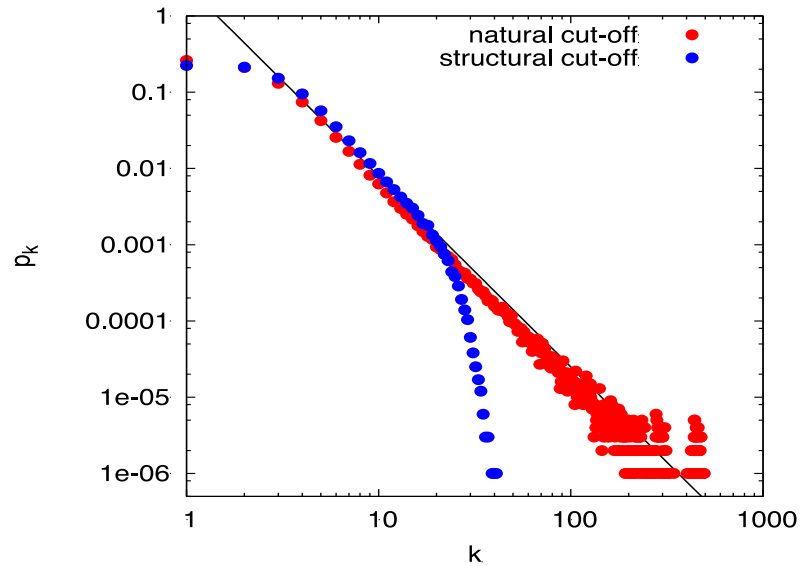


$$1 - \text{CDF} = P(k' > k) = 1 - \sum_{k'}^k p_{k'}$$



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 dissortativity property is visible in this case.

Natural cutoffs in scale-free networks

All real networks are finite $\rightarrow$ let us explore its consequences.

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

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

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

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

$\gamma < 3$: $k_{\max} > k_s \longrightarrow$

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.

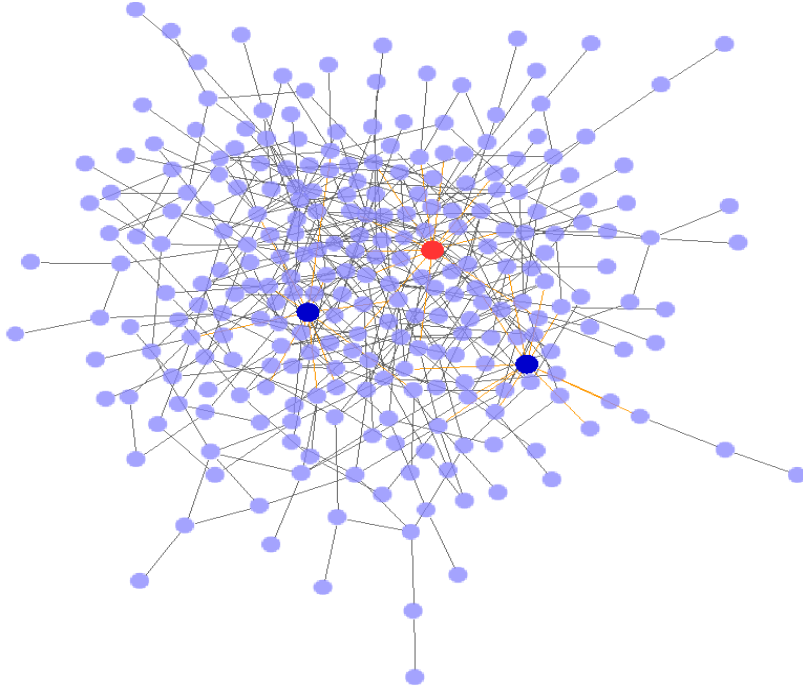
$\rightarrow$ *disassortative mixing*

$\rightarrow$ a randomly wired network with $\gamma < 3$ 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 ($N=300$, $L=450$, $\gamma=2.2$) with structural cut-off $\sim N^{1/2}$.

$r=0.005 \rightarrow$ **neutral**

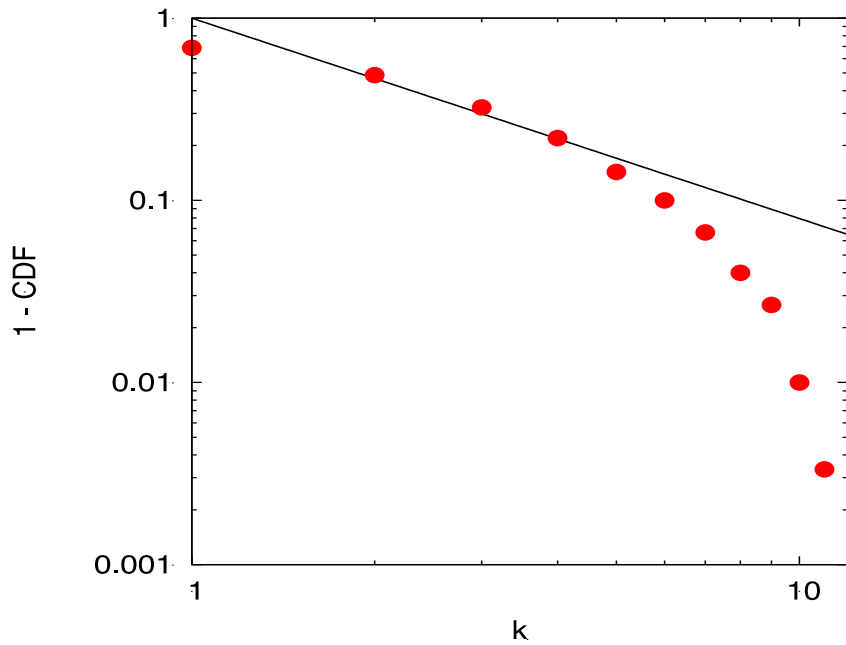
Red hub: 12 neighbors.

Blue hubs: 11 neighbors.

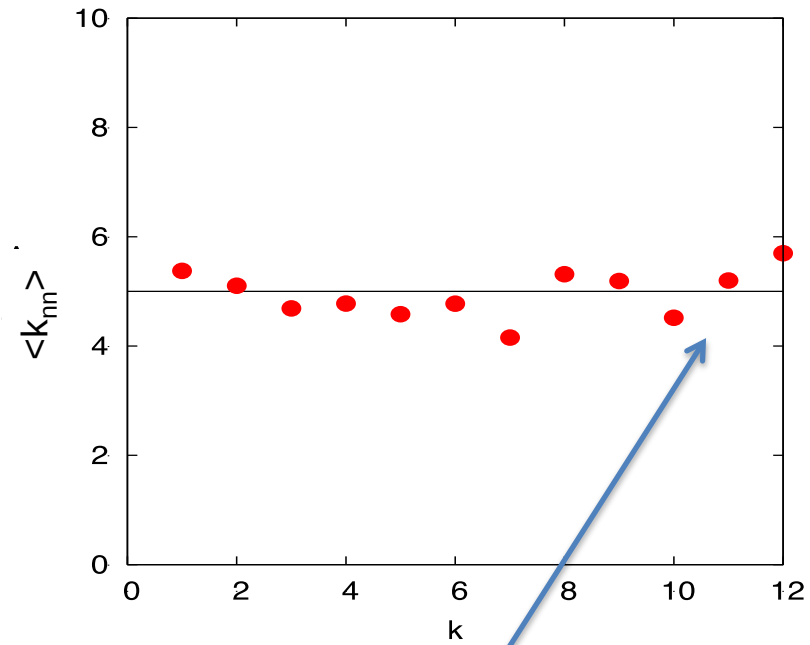
Again we can calculate the expectation number of edges between the hubs.

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

Diagram illustrating the calculation of the expectation number of edges between hubs. The formula is shown with arrows indicating the substitution of values: $\langle k \rangle$ points to 900, k points to 12, k' points to 11, and $p_{k'}$ points to 2.

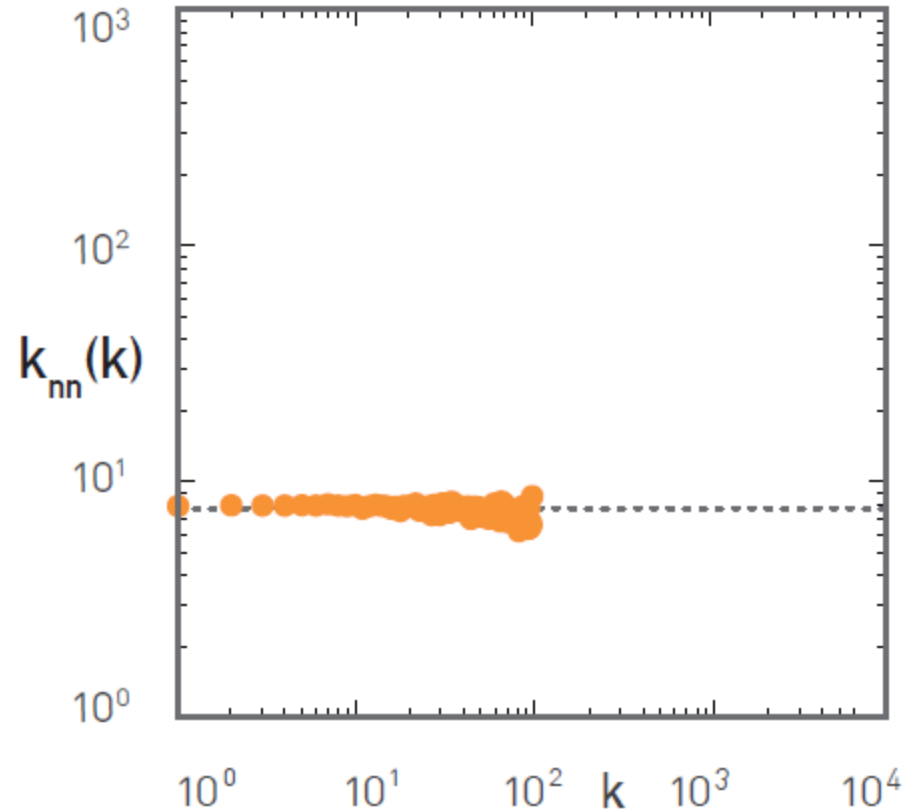
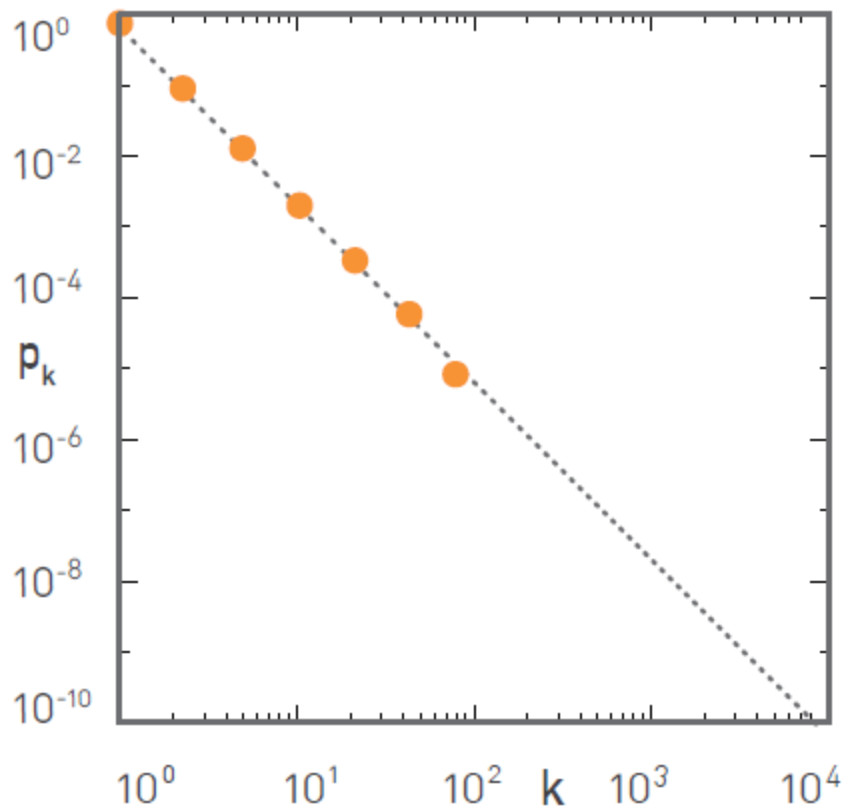


$$1 - CDF = P(k' > k) = 1 - \sum_{k'}^k p_{k'}$$



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

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

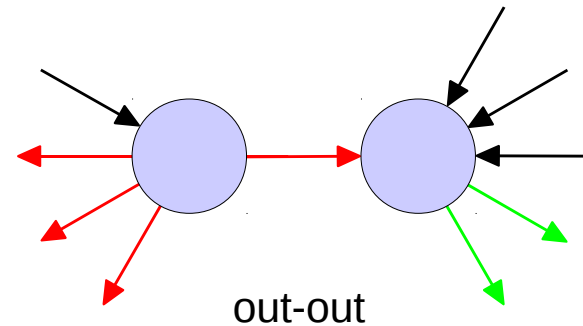
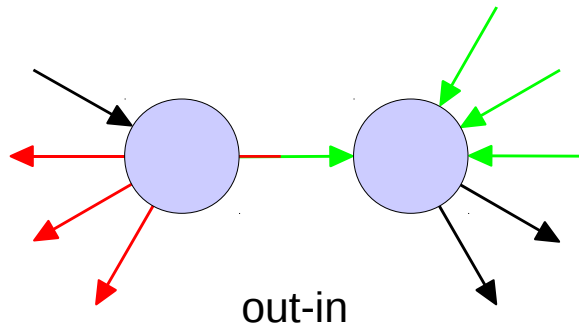
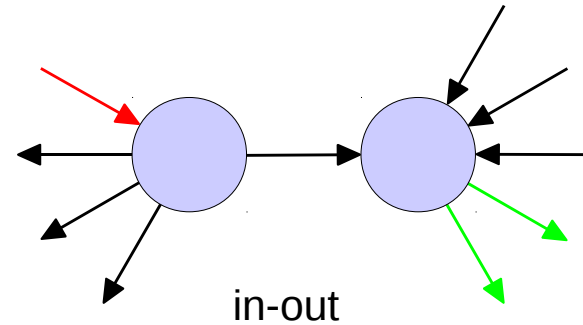
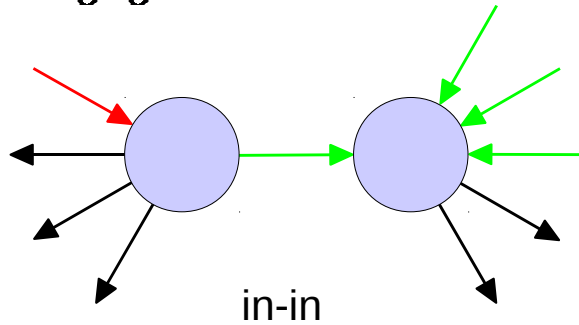


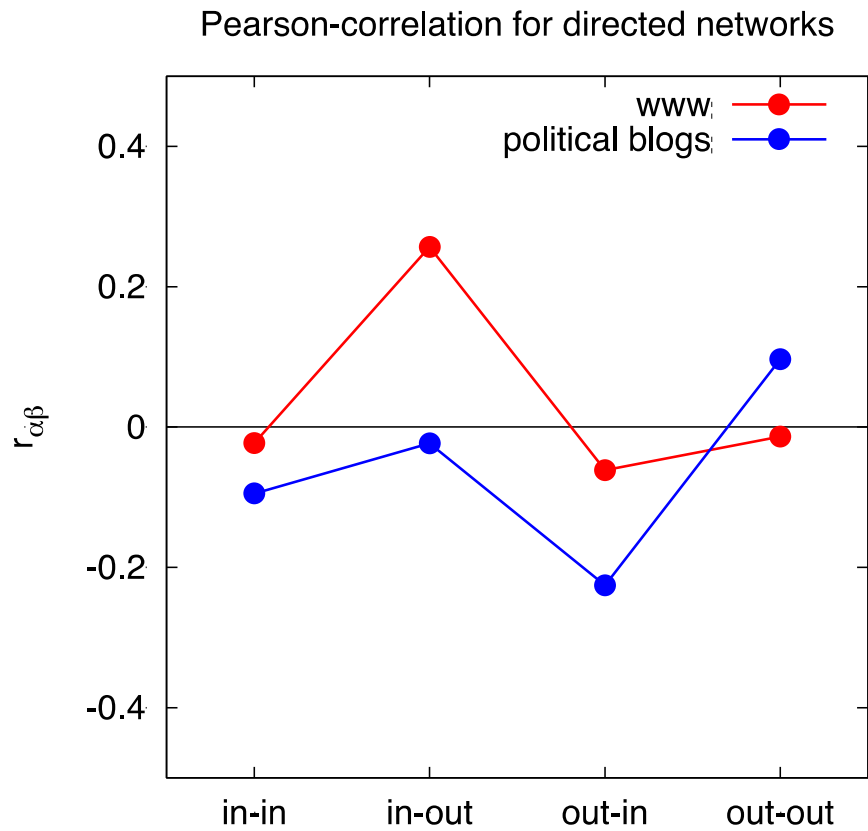
A clear case of neutral assortativity property is visible in this case thanks to imposing structural cut-off.

DIRECTED NETWORKS

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

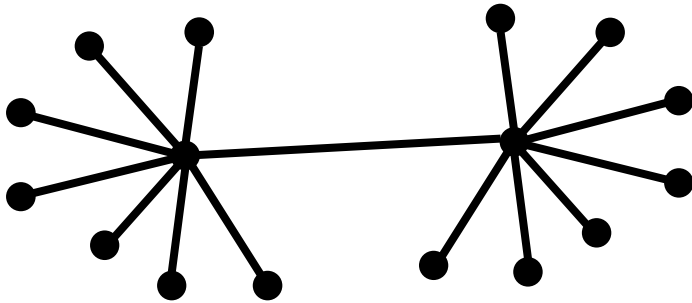
$\alpha, \beta: \{\text{in}, \text{out}\}$





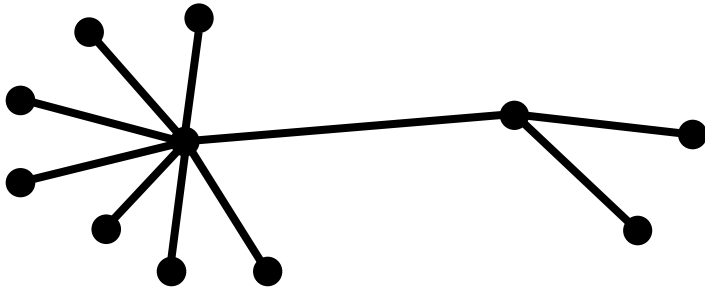
MULTIPOINT DEGREE CORRELATIONS

$P(k)$: 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:

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

Simplest case:

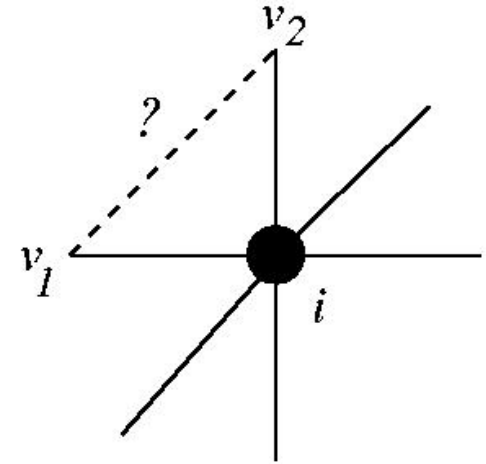
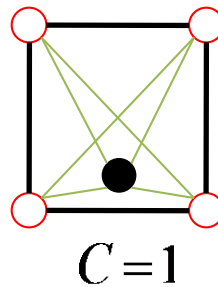
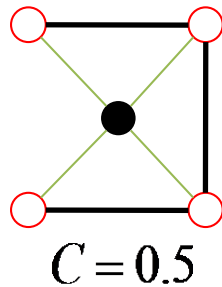
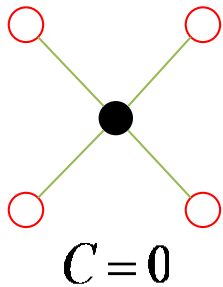
$P(k' | k)$: conditional probability that a node of degree k' is connected to a node of degree k

2-POINTS: CLUSTERING COEFFICIENT

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

Do your friends know each other ?

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

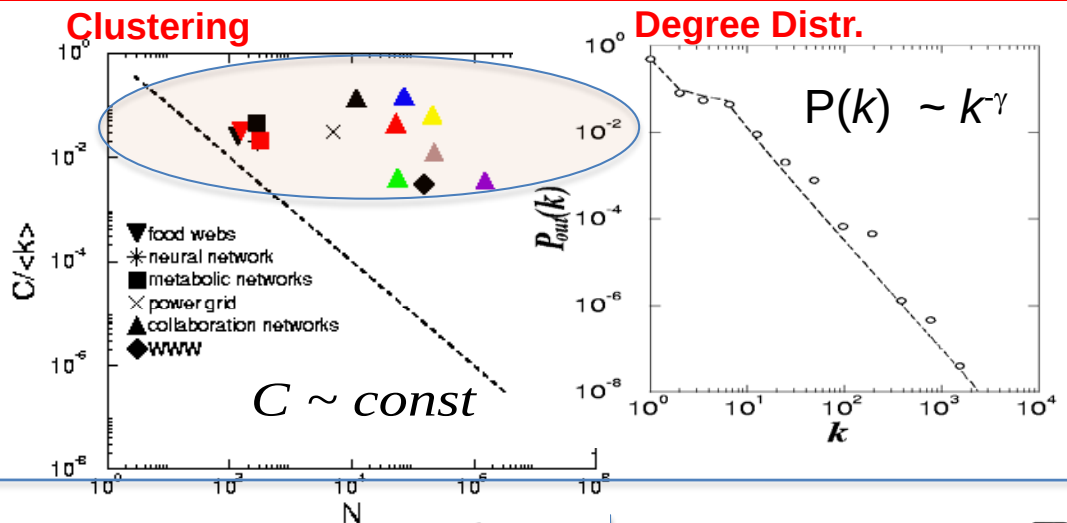
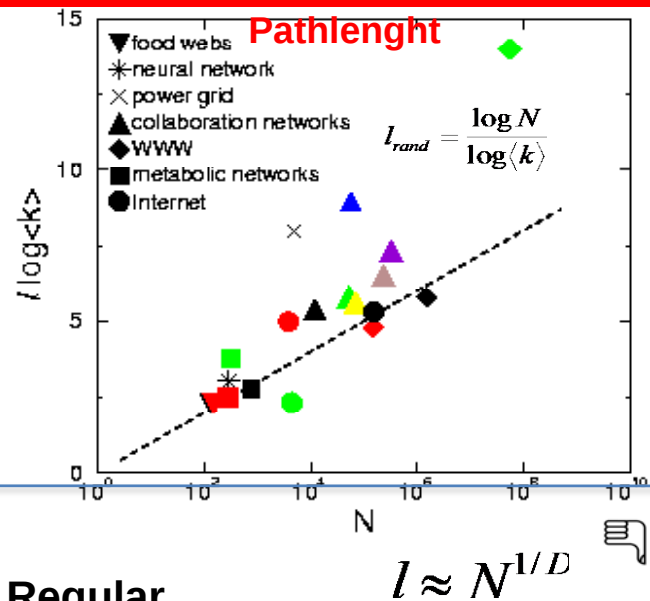


- Average clustering coefficient

= average over nodes with very different characteristics

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

EMPIRICAL DATA FOR REAL NETWORKS



Regular network		$C \sim const$	$P(k) = \delta(k - k_d)$
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	$l \approx \frac{\ln N}{\ln \ln N}$	$C \sim \frac{(\ln N)^2}{N}$	$P(k) \sim k^{-\gamma}$

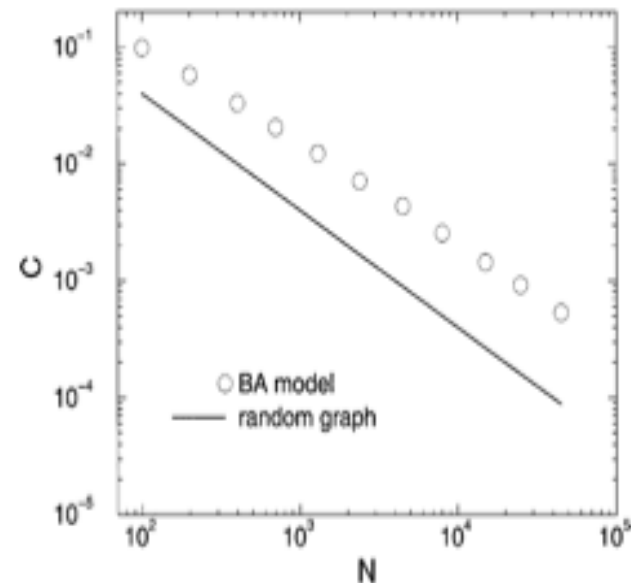
CLUSTERING COEFFICIENT OF THE BA MODEL

Reminder: for a random graph we have:

$$C_{rand} = \frac{\langle k \rangle}{N} \sim N^{-1}$$

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

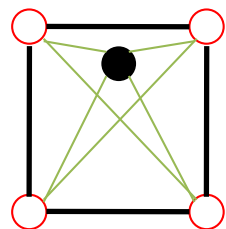
But not slow *enough*...



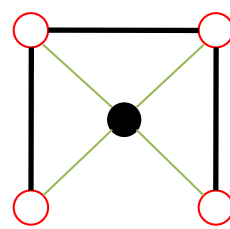
Clustering coefficient versus size of the Barabasi-Albert (BA) model with $\langle k \rangle = 4$, compared with clustering coefficient of random

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

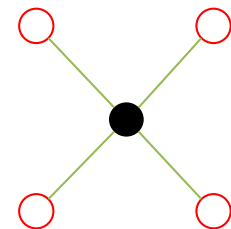
MODULARITY IN THE METABOLISM



$C=1$



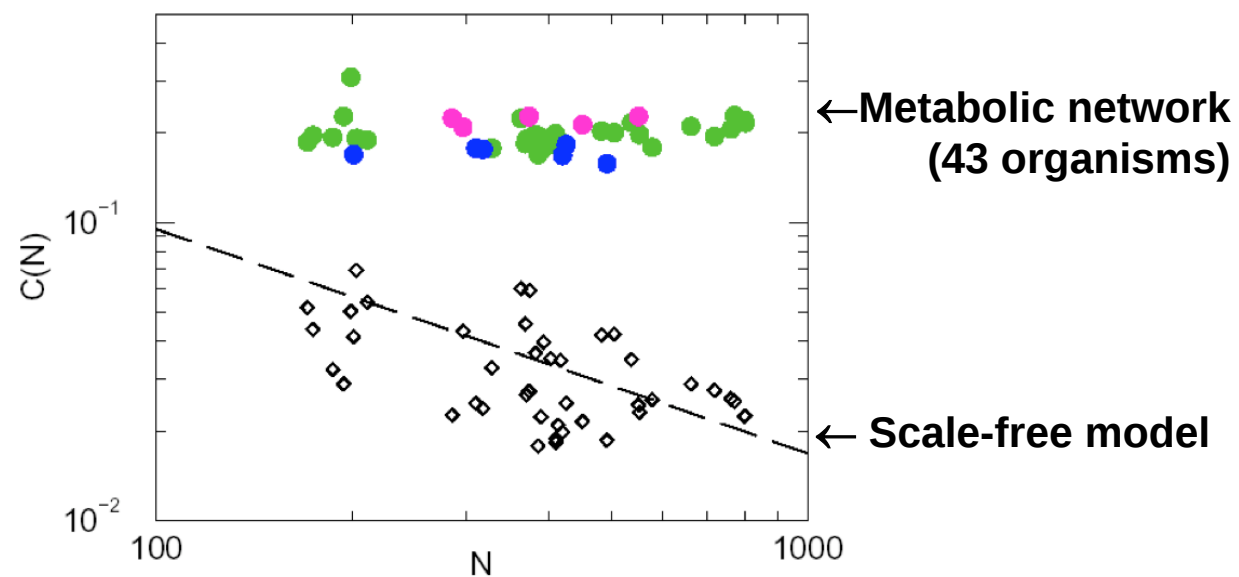
$C=\frac{1}{2}$



$C=0$

Clustering Coefficient:

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



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

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

- Clustering **spectrum**:

putting together nodes which have the same degree

(link with hierarchical structures)

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

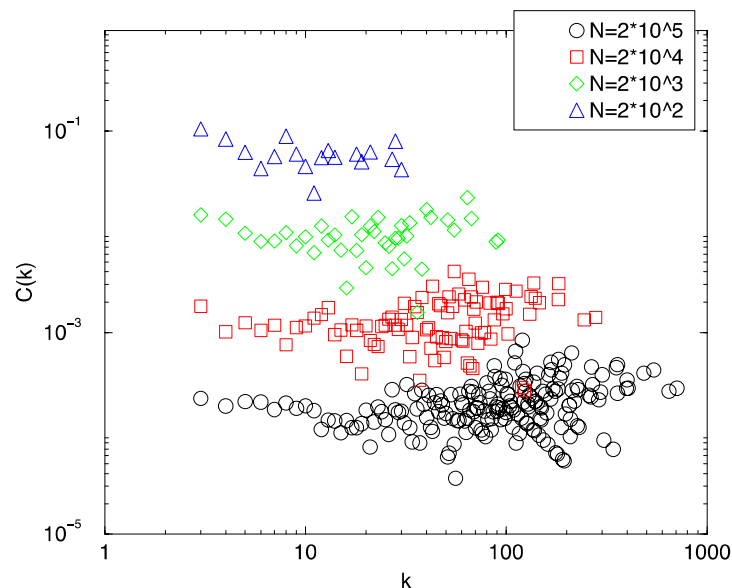
↑
class of degree k

C(k) for the ER and BA models

Erdos-Renyi

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

Barabasi-Albert

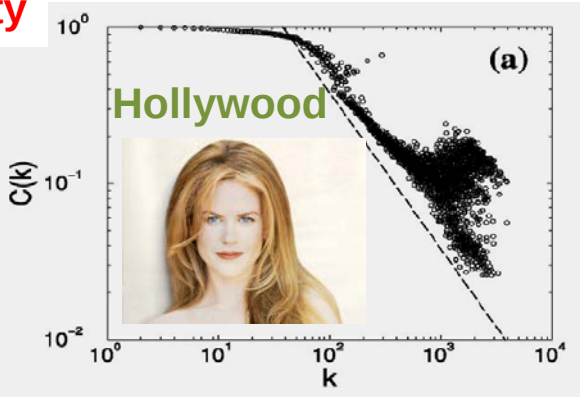


$$C \sim \frac{(\ln N)^2}{N}$$

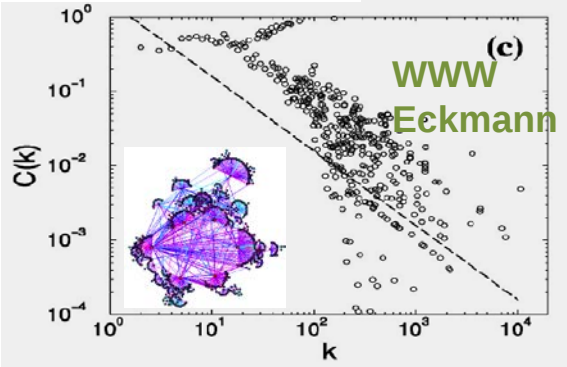
This is not true, however, for real networks. Let us look at some empirical data.

HIERARCHICAL NETWORKS

Society

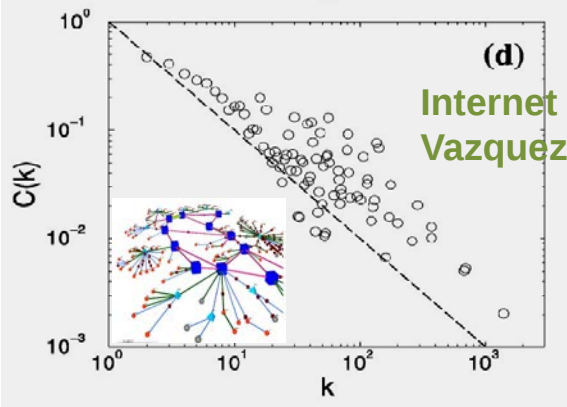
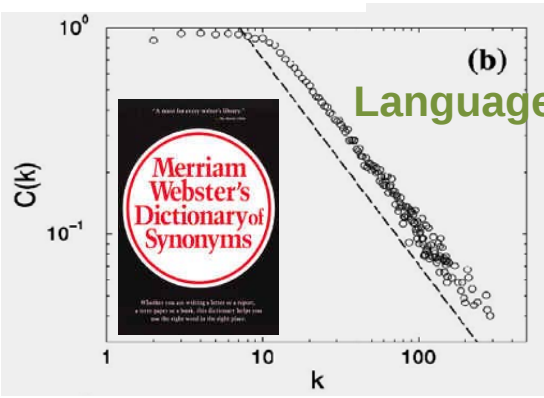


The electronic skin



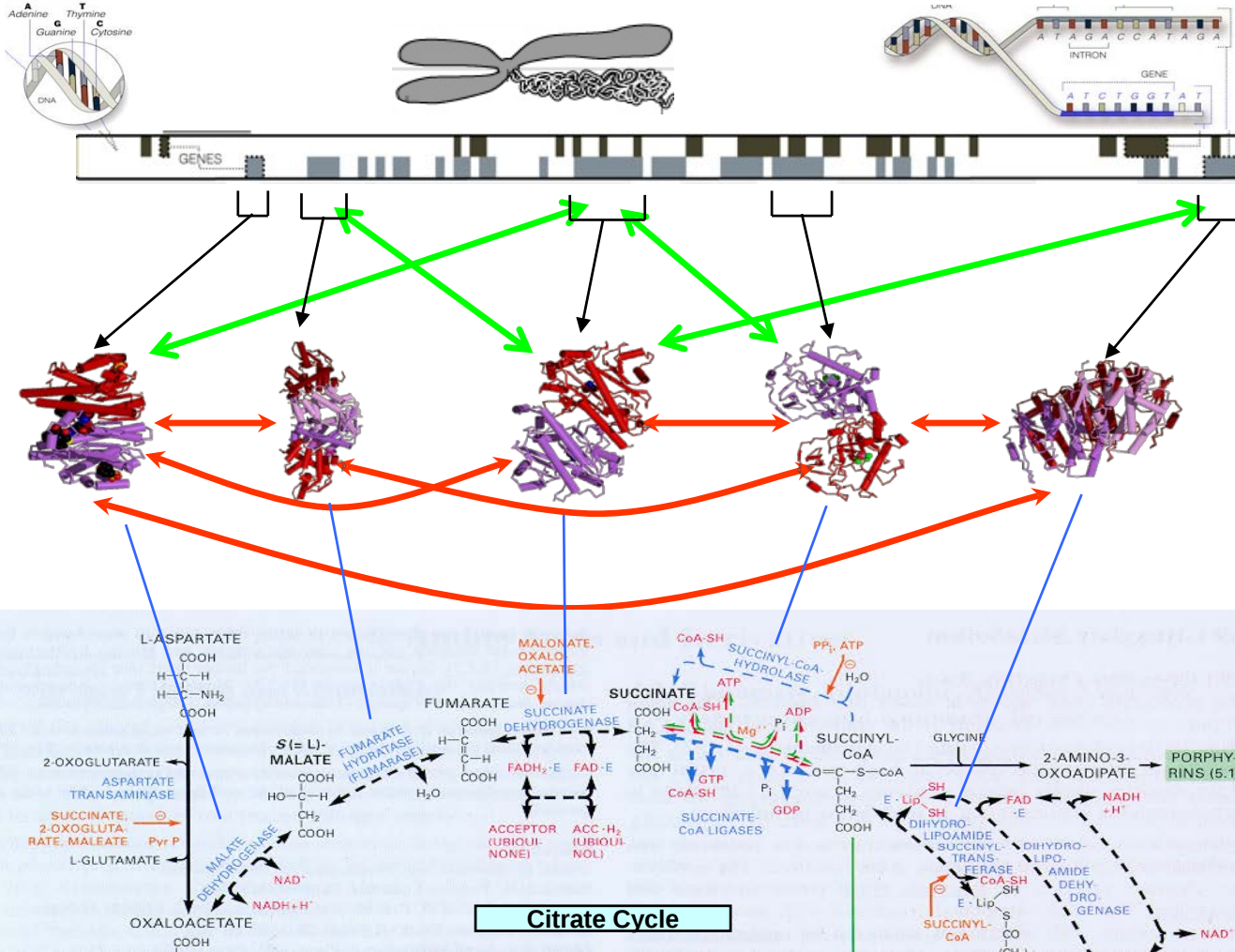
Eckmann & Moses, '02

Human communication



Internet (AS)
Vazquez et al, '01

Cellular networks:



GENOME

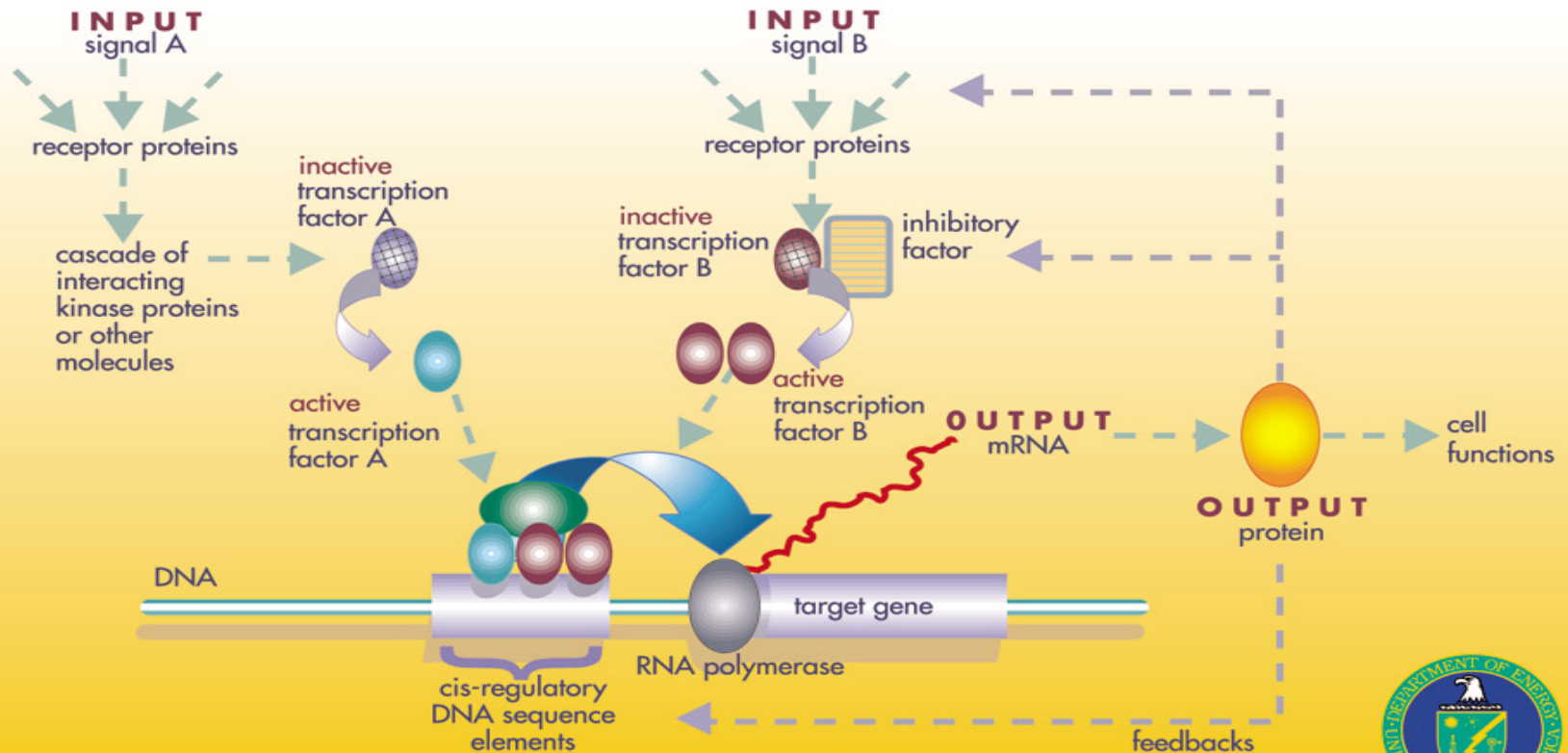
**protein-gene
interactions**

**PROTEOME
protein-protein
interactions**

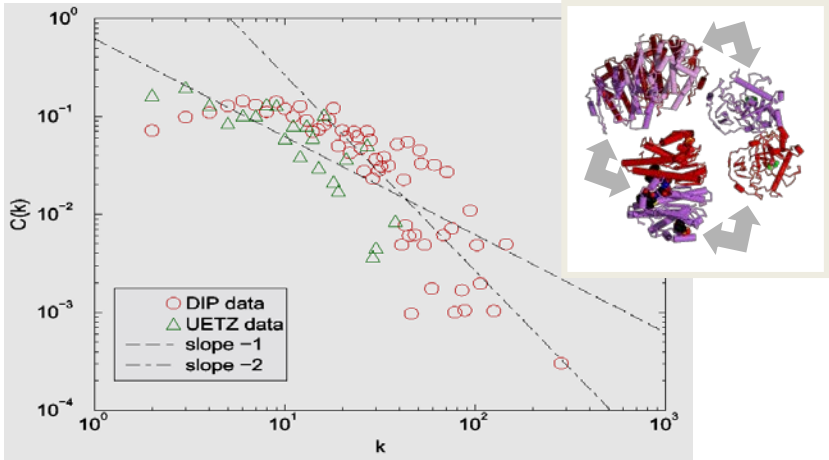
METABOLISM

**Bio-chemical
reactions**

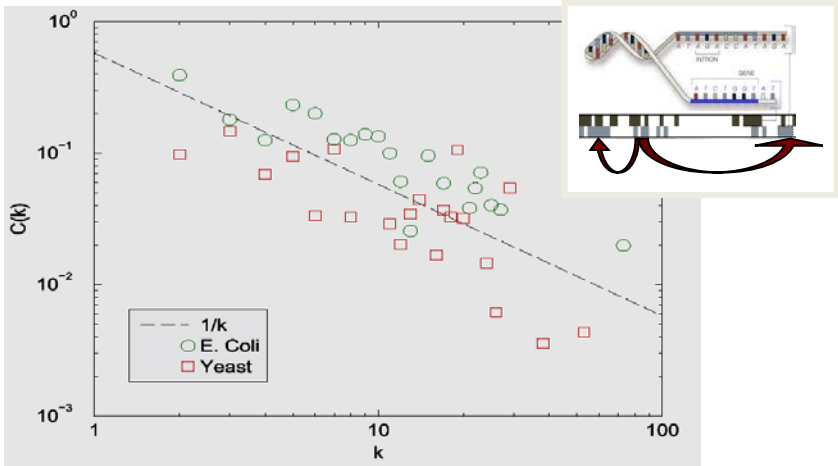
A GENE REGULATORY NETWORK



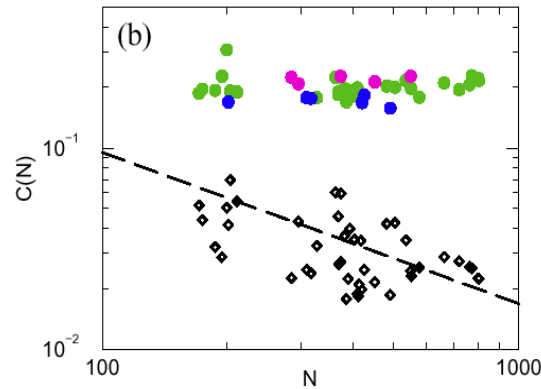
Protein-protein interaction



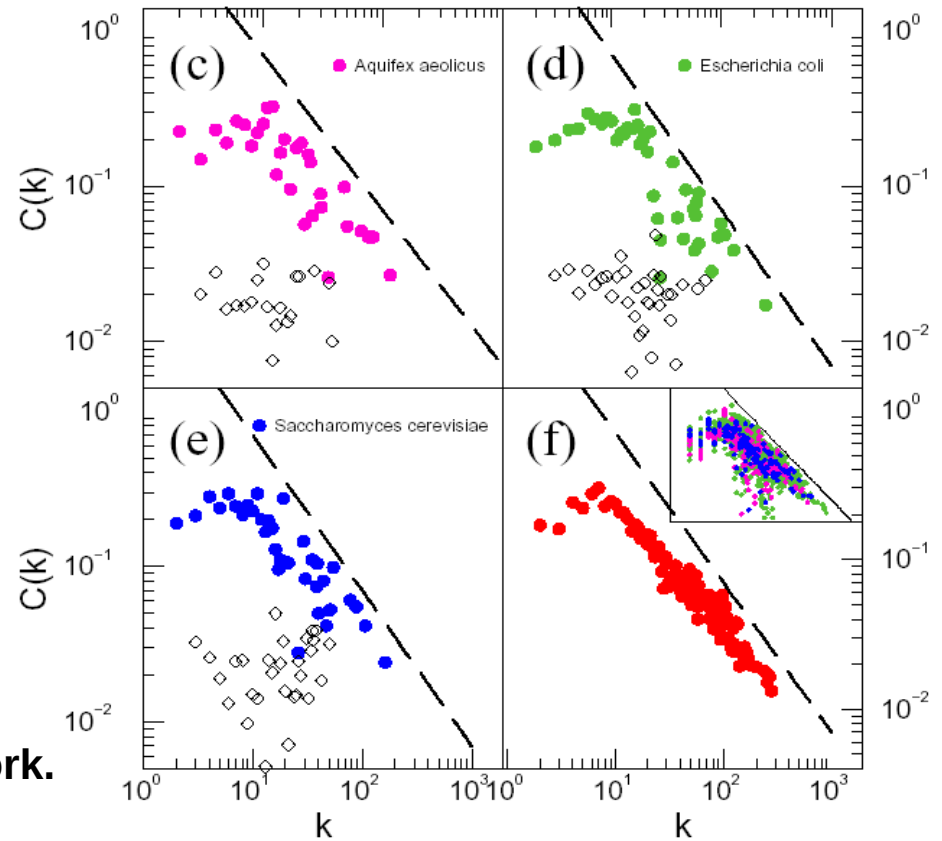
Regulatory networks



SCALING OF THE CLUSTERING COEFFICIENT $C(k)$

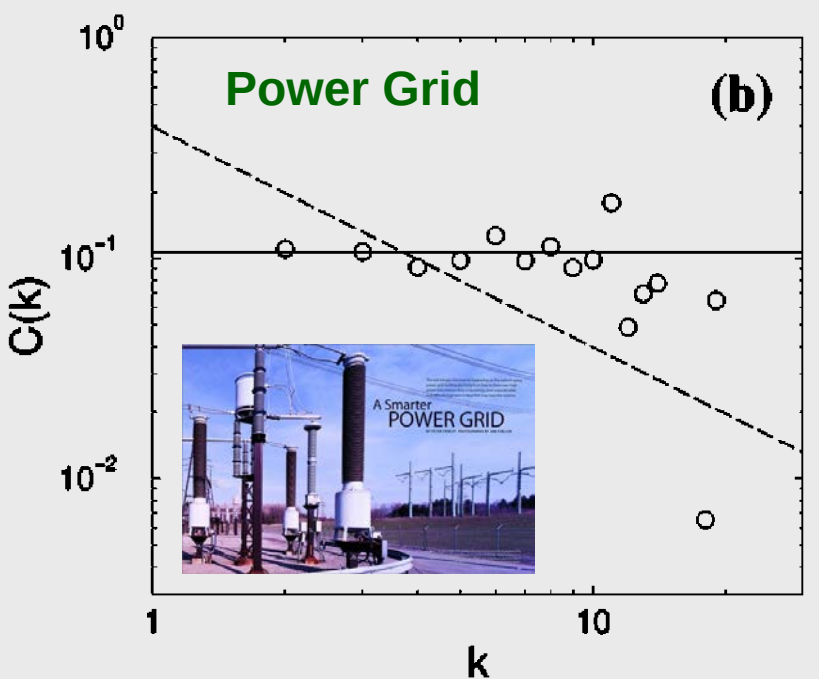
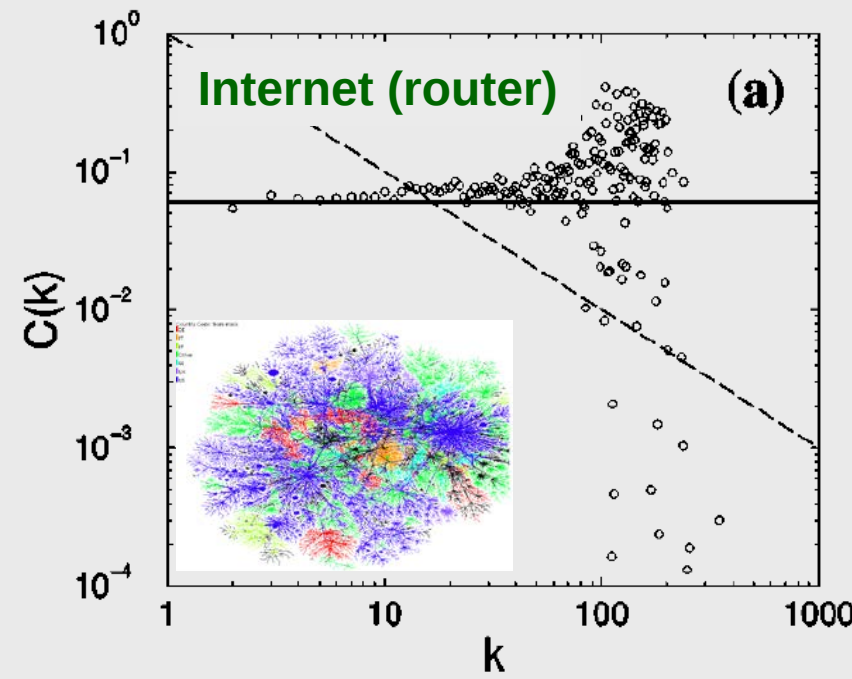


The metabolism forms a hierarchical network.



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 at stake, that needs to be considered— that of modularity.

All models predict

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

Is the exponent universal?

Or could we have for example:

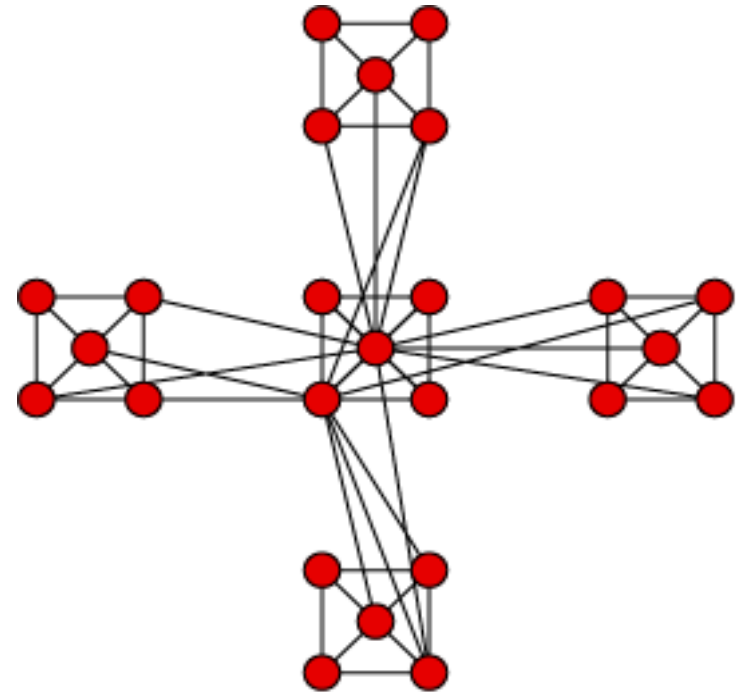
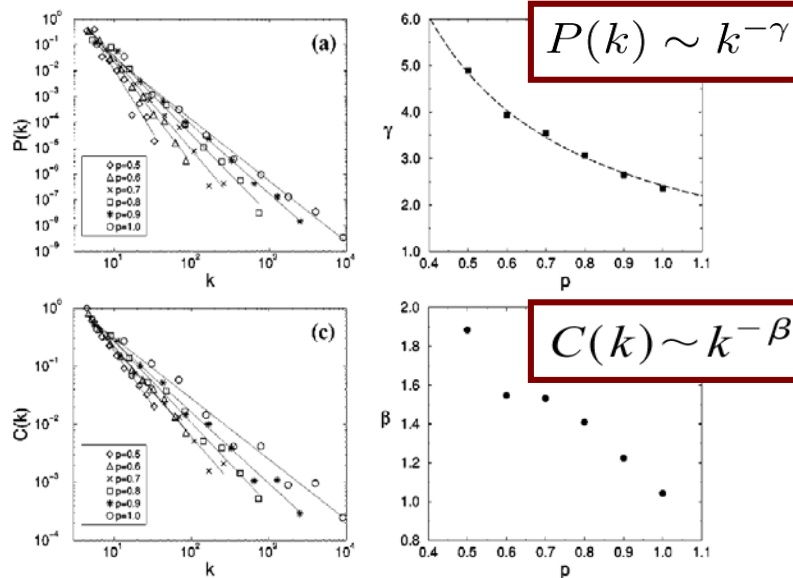
$$C(k) \sim k^{-\beta}$$

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.

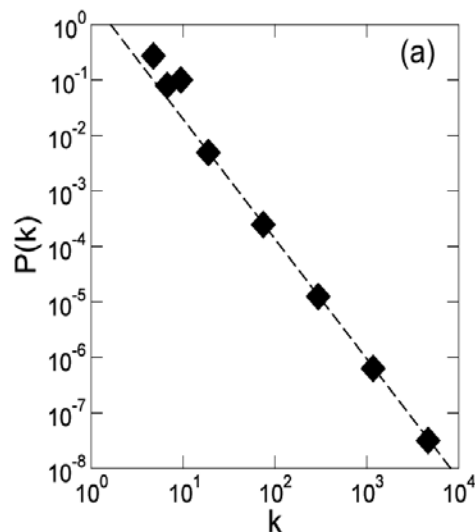
-at the next level p^2 fraction will link, back, then $p^3, \dots p^i$



SUMMARY

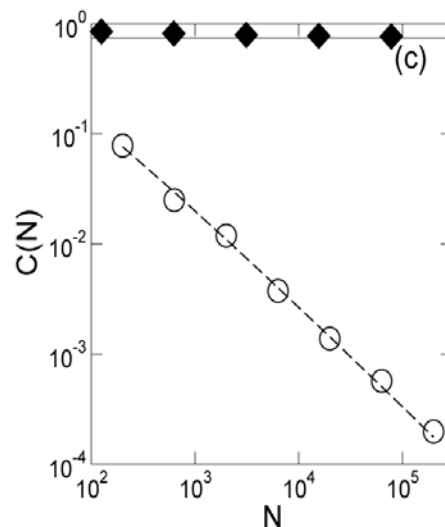
1. Scale-free

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



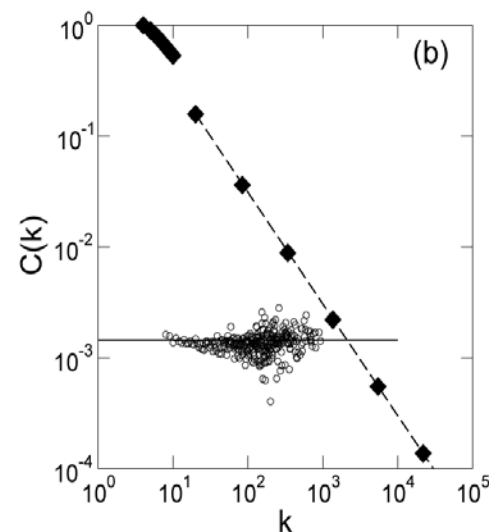
2. Clustering coefficient independent of N

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



3. Clustering spectrum

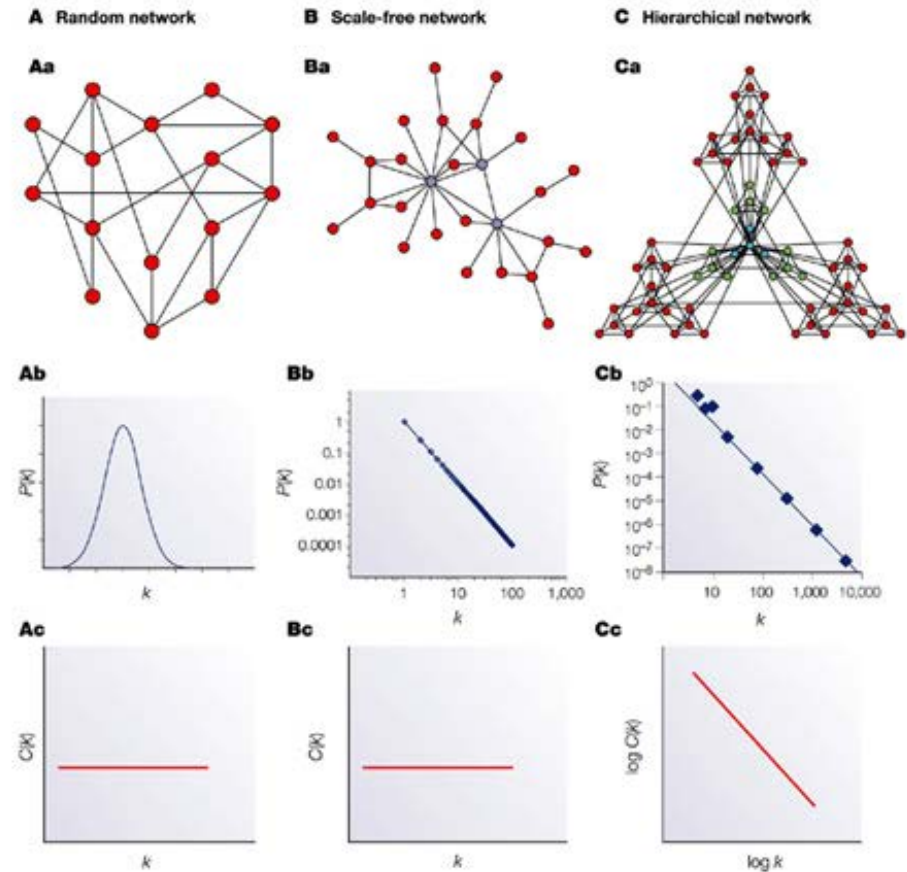
$$C(k) \sim k^{-1}$$



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.



What does happen in real systems? Is a prediction that all systems with $\gamma < 3$ should be automatically disassortative, or have a cutoff – is this the case?

Let's see: www, $\gamma=2.1$, no cutoff, disassortative NICE

Actor network, no cutoff, but it is ASSORTATIVE (how is this possible?).

Internet: $\gamma=2.5$, disassortative, cutoff, NICE

Networks with $\gamma < 3$ 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.

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

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

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.

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

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

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.

$$\langle k_m \mid i \leftrightarrow j \rangle = \sum_{k_m} k_m P(k_m \mid i \leftrightarrow j) = 2$$

$$P(k_m \mid i \leftrightarrow j) = \frac{P(k_m, i \leftrightarrow j)}{P(i \leftrightarrow j)} = \frac{P(i \leftrightarrow j \mid k_m) P(k_m)}{P(i \leftrightarrow j)} \quad \text{Bayes' theorem}$$

$P(k_m \mid i \leftrightarrow j)$: 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 $P(k)$:

$$P(i \leftrightarrow j) = \frac{2L}{N(N-1)} = \frac{\langle k \rangle}{N-1}$$

$$P(i \leftrightarrow j \mid k_m) = \frac{k_m}{N-1}$$

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

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

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} = 2$$

$\kappa > 2$: a giant cluster exists

$\kappa < 2$: many disconnected clusters

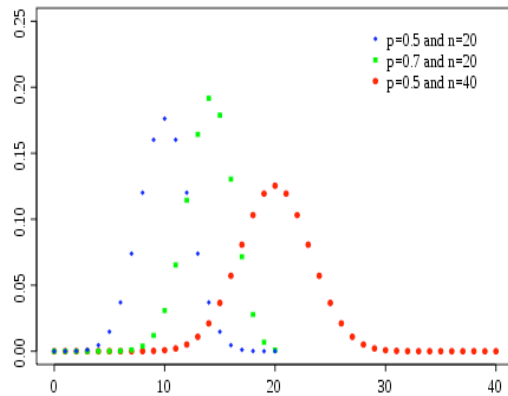


Apply the Malloy-Reed Criteria to an Erdos-Renyi Network

Discrete Formulation

-binomial distribution-

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



$$\langle k \rangle = (N-1)p$$

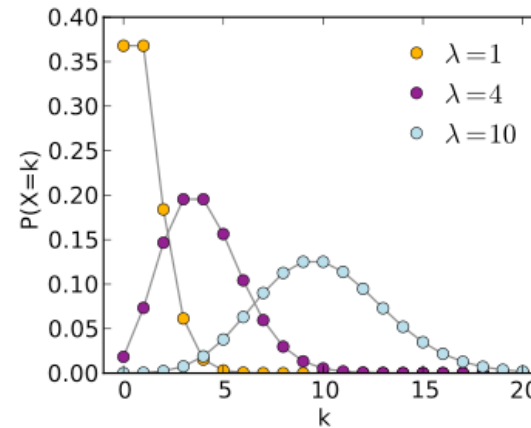
$$\langle k^2 \rangle = p(1-p)(N-1) + p^2(N-1)^2$$

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

Continuum Formulation

-Poisson distribution-

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$



$$\langle k \rangle = \langle k \rangle$$

$$\langle k^2 \rangle = \langle k \rangle (1 + \langle k \rangle)$$

$$\sigma_k = (\langle k^2 \rangle - \langle k \rangle^2)^{1/2} = \langle k \rangle^{1/2}$$

Probability Distribution Function
(PDF)

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.

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} = 2$$

$\kappa > 2$: a giant cluster exists;

$\kappa < 2$: many disconnected clusters;

$$\langle k \rangle = \langle k \rangle$$

$$\langle k^2 \rangle = \langle k \rangle (1 + \langle k \rangle)$$

$$\sigma_k = (\langle k^2 \rangle - \langle k \rangle^2)^{1/2} = \langle k \rangle^{1/2}$$

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\langle k \rangle (1 + \langle k \rangle)}{\langle k \rangle} = 1 + \langle k \rangle = 2$$

$$\langle k \rangle = 1$$

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