Degree Correlation

Alberto Paccanaro
EMAp – FGV

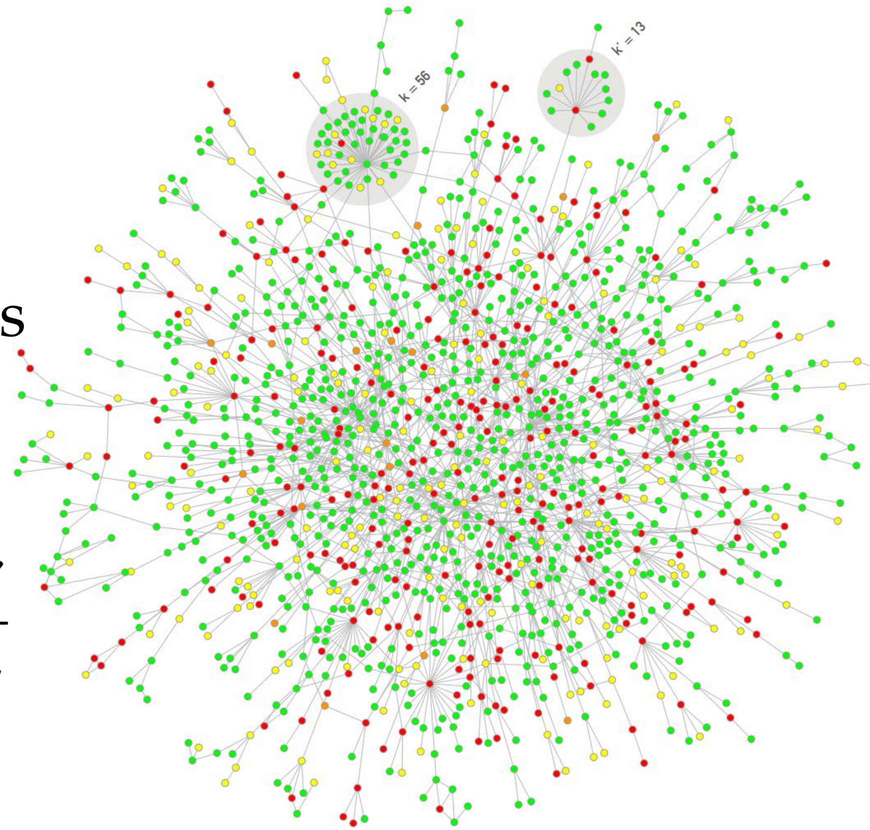
www.paccanarolab.org

Some material and images are from (or adapted from):
A. Barabási, and M. Pósfai. Network science, Cambridge University Press, 2016

Hubs link to hubs… or not ?

In social networks, hubs tend to have ties with hubs.

In protein interaction networks the opposite is true.



Prob. nodes with k and k′ degree connect to each other: $p_{k,k'} = \dfrac{kk'}{2L}$

$p_{1,2} = 0.0004$ but we see a lot of these

$p_{56,13} = 0.16$ but the hubs don't connect

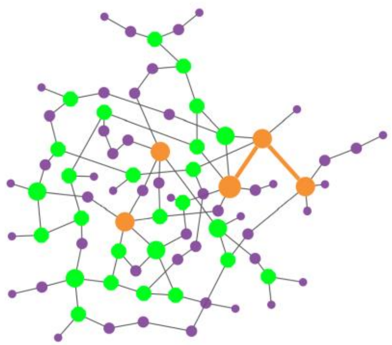
Real networks exhibit degree correlations

A.PACCANARO – F

Assortativity, disassortativity

3 networks with **identical degree sequence** but different topologies

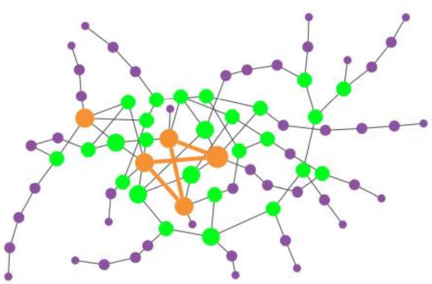
NEUTRAL



- The number of links between the hubs coincides with what we expect by chance

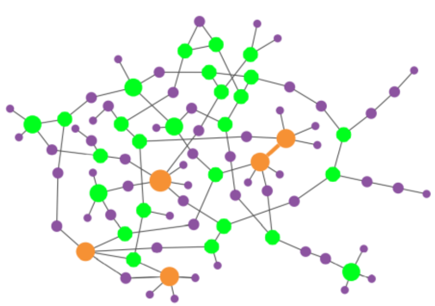
$$p_{k,k'} = \frac{kk'}{2L}$$

ASSORTATIVE



- Hubs tend to link to each other and avoid linking to small-degree nodes.
- Small-degree nodes tend to connect to other small-degree nodes

DISASSORTATIVE



- Hubs avoid each other, linking instead to small-degree nodes.
- The network displays a hub and-spoke character, making it disassortative.

Degree correlations

A network displays degree correlations if the number of links between the high and low-degree nodes is systematically different from what is expected by chance.

How can we quantify it?

Probability q_k that there is a degree-k node at the end of a randomly selected link

$$q_k = Ckp_k$$

k = number of stubs of the end node

p_k = prob. of node of degree k

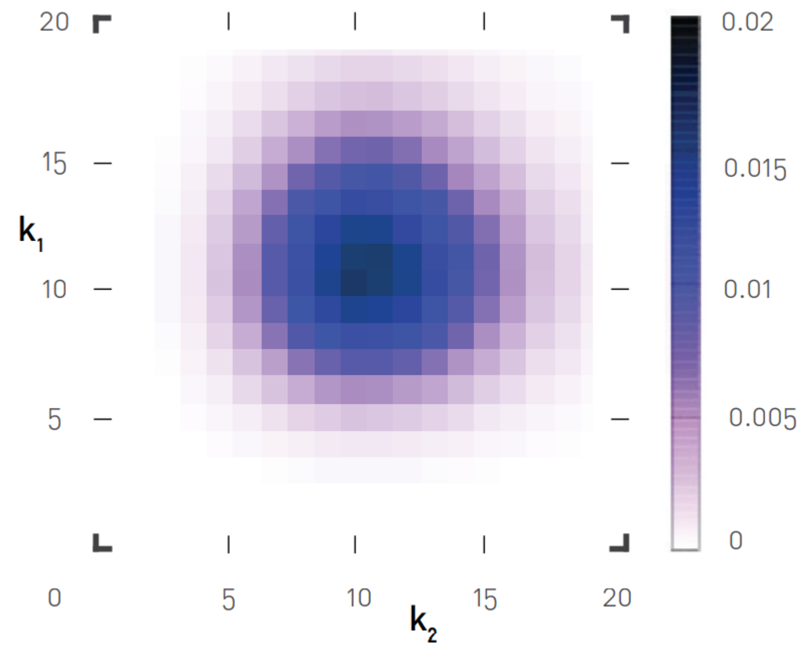
$$\sum_k q_k = 1 \quad \Rightarrow \quad C = \frac{1}{\sum_k kp_k} = \frac{1}{\langle k \rangle} \quad \Rightarrow \quad q_k = \frac{kp_k}{\langle k \rangle}$$

Degree correlation matrix

e_{ij} is the probability of finding a node with degrees i and j at the two ends of a randomly selected link

We can collect all e_{ij} in a matrix where:

$$\sum_{ij} e_{ij} = 1 \qquad \sum_{j} e_{ij} = q_i$$

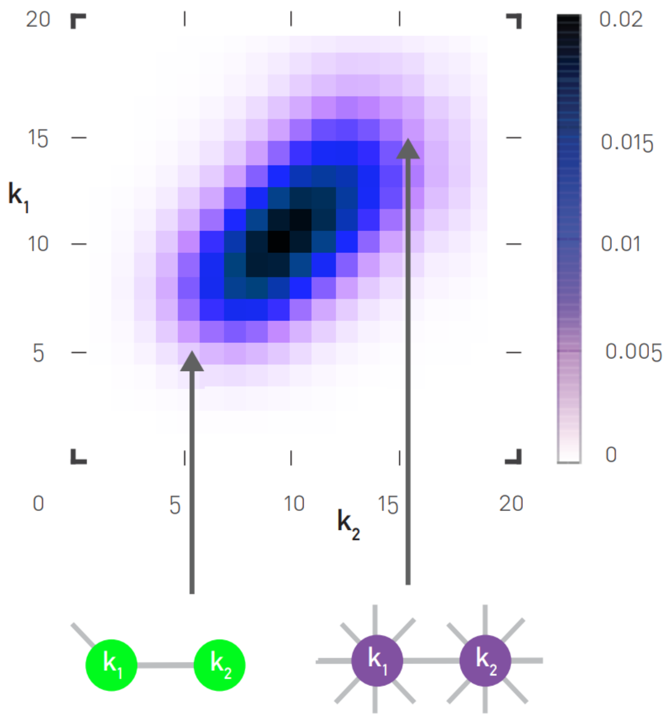


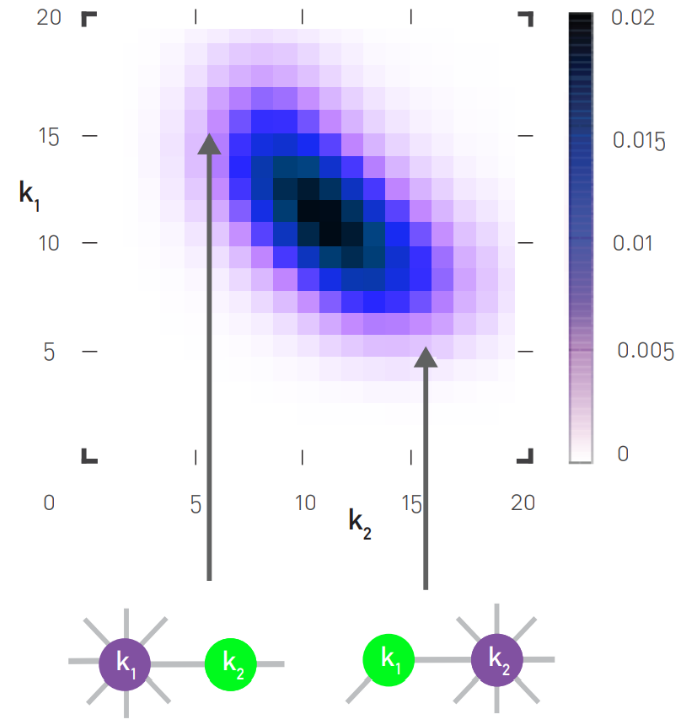
In a neutral network:

$$e_{ij} = q_i q_j$$

A network with degree correlations will deviate from this

A.PACCANAR





Assortative network **Disassortative network**

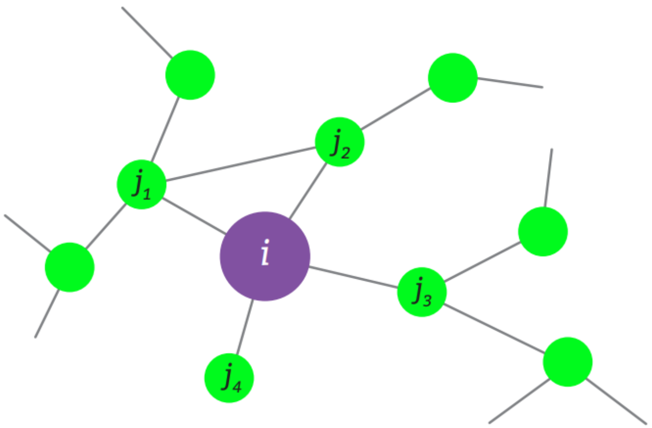
Information about degree correlations is carried by the degree correlation matrix

The degree correlation function

We want to capture the relationship between the degrees of nodes that link to each other.

Measure for each node *i* the average degree of its neighbours

$$k_{\mathrm{nn}}(k_i) = \frac{1}{k_i}\sum_{j=1}^{N} A_{ij}k_j$$



Degree correlation function: $\boxed{k_{\mathrm{nn}}(k) = \sum_{k'} k' P(k' \mid k)}$

$P(k' \mid k)$ conditional probability that following a link of a k-degree node we reach a degree-k' node.

$k_{nn}(k)$ is the average degree of the neighbours of all degree-k nodes

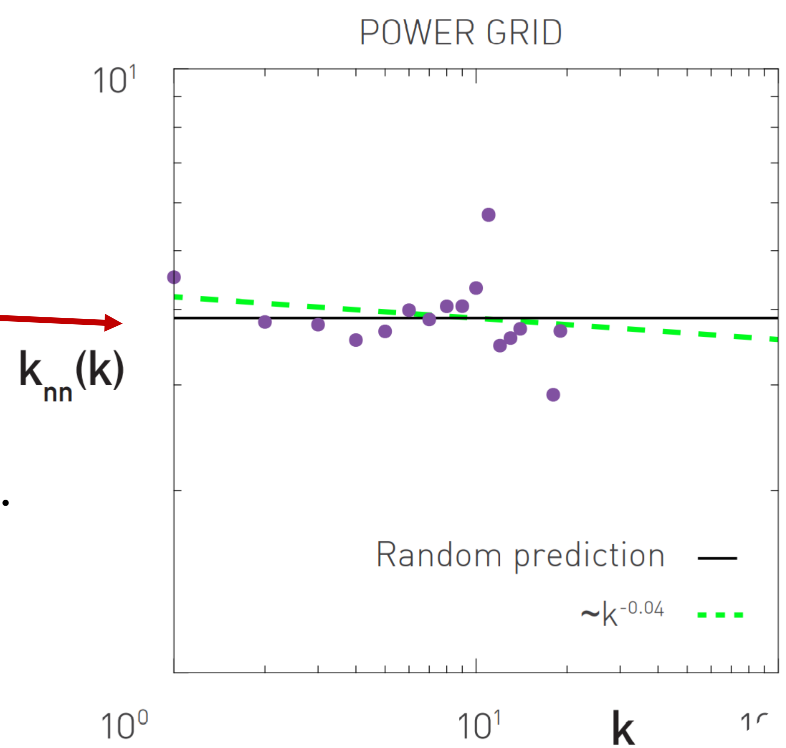
In a neutral network:

$$P(k' \mid k) = \frac{e_{kk'}}{\sum_{k'} e_{kk'}} = \frac{e_{kk'}}{q_k} = \frac{q_{k'} q_k}{q_k} = q_{k'}$$

$$q_k = \frac{k p_k}{\langle k \rangle}$$

$$k_{nn}(k) = \sum_{k'} k' q_{k'} = \sum_{k'} k' \frac{k' p(k')}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

The average degree of a node's neighbours is independent of the node's degree



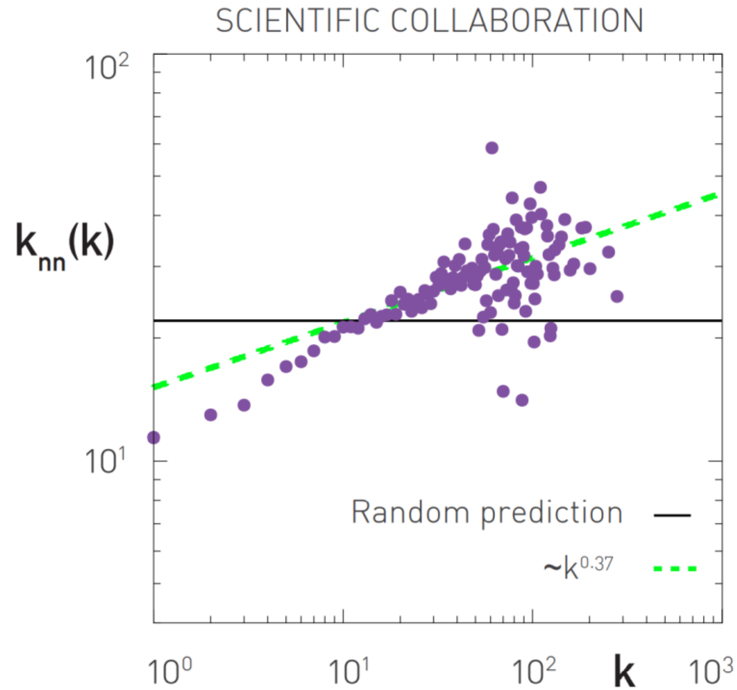
It depends only on the global network characteristics $\langle k \rangle$ and $\langle k^2 \rangle$.

Assortative networks:

Hubs tend to connect to other hubs

The higher is the degree k of a node, the higher is the average degree of its nearest neighbors.

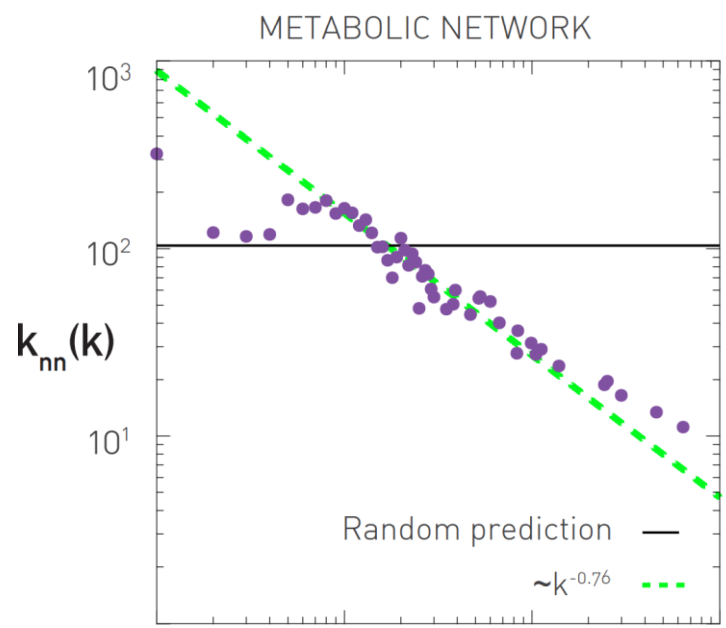
$k_{nn}(k)$* increases with *k



Disassortative network:

Hubs tend to link to low-degree nodes.

***$k_{nn}(k)$* decreases with k**



$$k_{\mathrm{nn}}(k) = ak^{\mu}$$

The nature of degree correlations is determined by the sign of the **correlation exponent μ**

- Assortative Networks: $\mu > 0$
- Neutral Networks: $\mu = 0$
- Disassortative Networks: $\mu < 0$

Structural cutoffs

There is a conflict between the scale-free property and degree correlations.

In a network with degree correlations $e_{kk'}$ the expected number of links between k and k' is:

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

Protein Interaction network: $E_{kk'} = \frac{k p_k k' p_{k'}}{\langle k \rangle} N = \frac{\frac{55}{300} \frac{46}{300}}{3} 300 = 2.8$

But we can have only one link !

For small k and k' we expect *less* than one link between the two nodes.

For nodes with degree above a structural cutoff k_s we expect multiple links:

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

Nodes with higher degree have $E_{kk'} > 1$

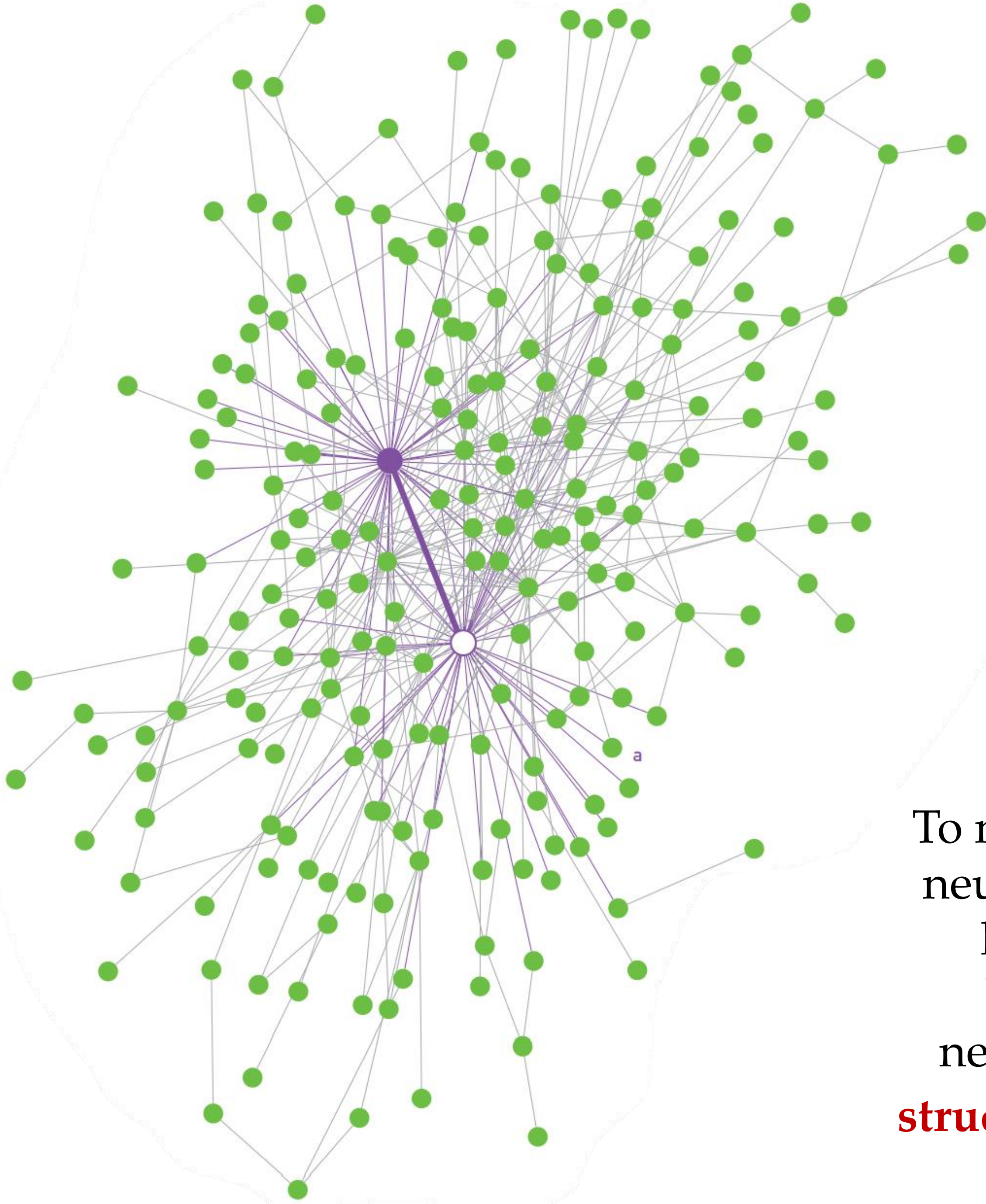
Which networks can have nodes with degree > k_s ?

Compare k_s with $kmax \sim N^{\frac{1}{\gamma-1}}$

random networks
scale-free networks with $\gamma \geq 3$ } $k_{max} < k_s$

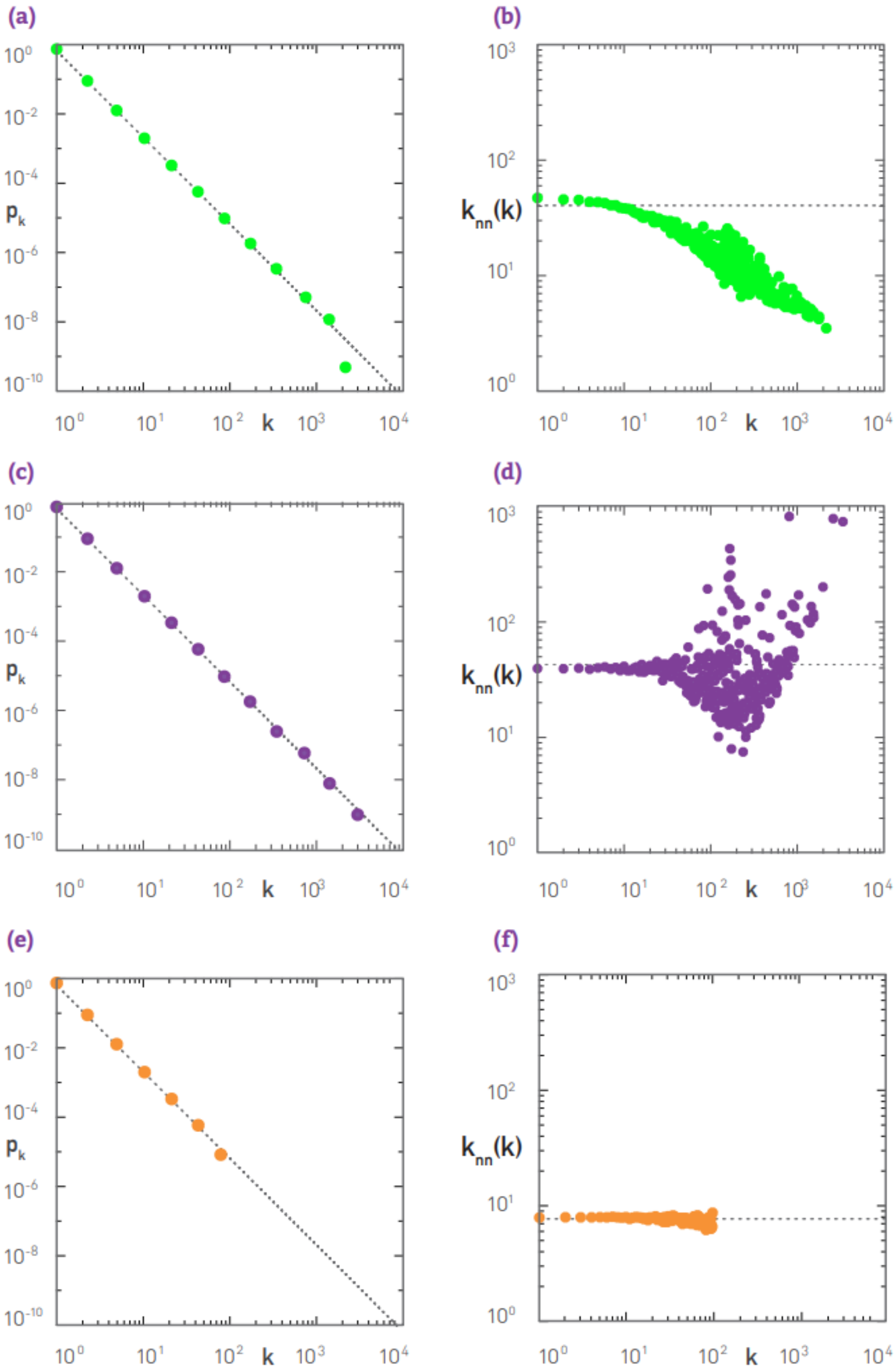
scale-fee networks with $\gamma < 3$ nodes with $k_s < k < k_{max}$ can violate $E_{kk'} > 1$

The network has fewer links between its hubs than expected.
Networks will become disassortative ➔ **structural disassortativity**.



To maintain the network neutral we need about 3 links between these 2 hubs. This makes the network disassortative:

structural disassortativity



If we generate a scale-free network with the power-law degree distribution shown in (a), and we forbid self-loops and multi-links, the network displays structural disassortativity, as indicated by $k_{nn}(k)$ in (b). In this case, we lack a sufficient number of links between the high-degree nodes to maintain the neutral nature of the network, hence for high k the $k_{nn}(k)$ function must decay.

We can eliminate structural disassortativity by relaxing the simple network requirement, i.e. allowing multiple links between two nodes. As shown in (c,d), in this case we obtain a neutral scale-free network.

If we impose an upper cutoff by removing all nodes with $k \geq k_s \simeq 100$, as predicted by (7.15), the network becomes neutral, as seen in (f).

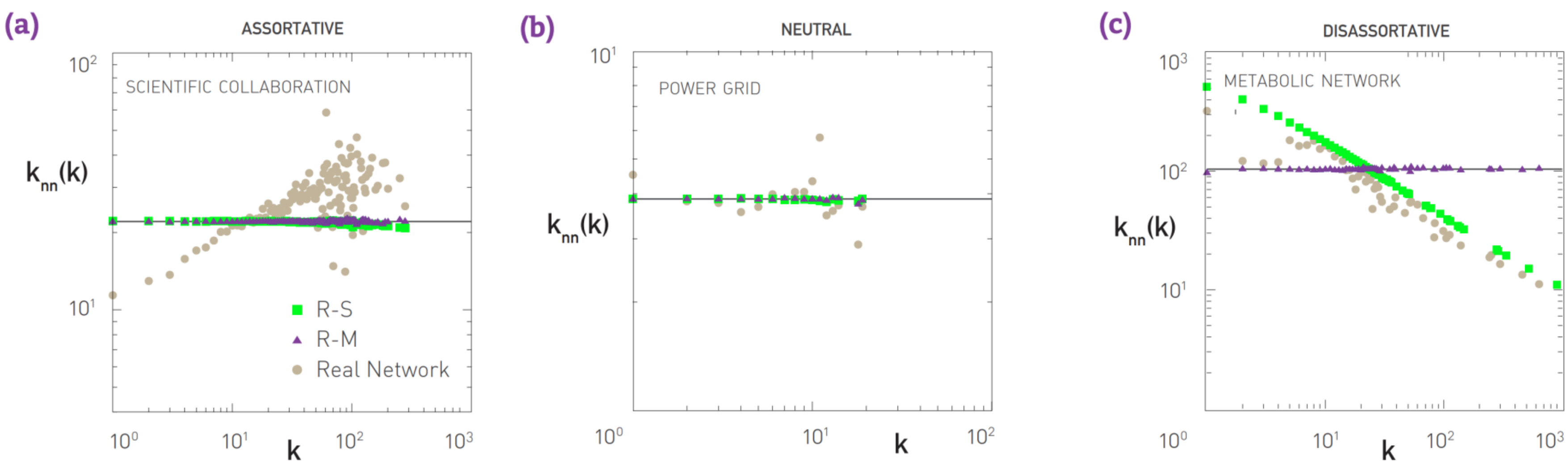
Scale-free network, N = 10,000 and $\gamma = 2.5$, configuration model

Degree Preserving Randomization with Simple Links (R-S)

Is the structural disassortativity that I measure "real"?

Degree-preserving randomization to the original network (no multi-links).

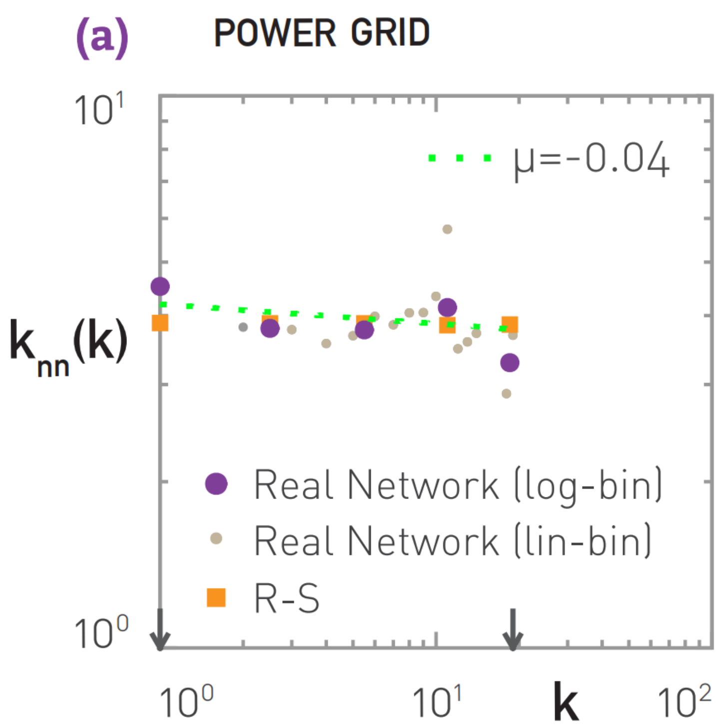
Compare the real $k_{nn}(k)$ and the randomized $k_{nn}^{R-S}(k)$

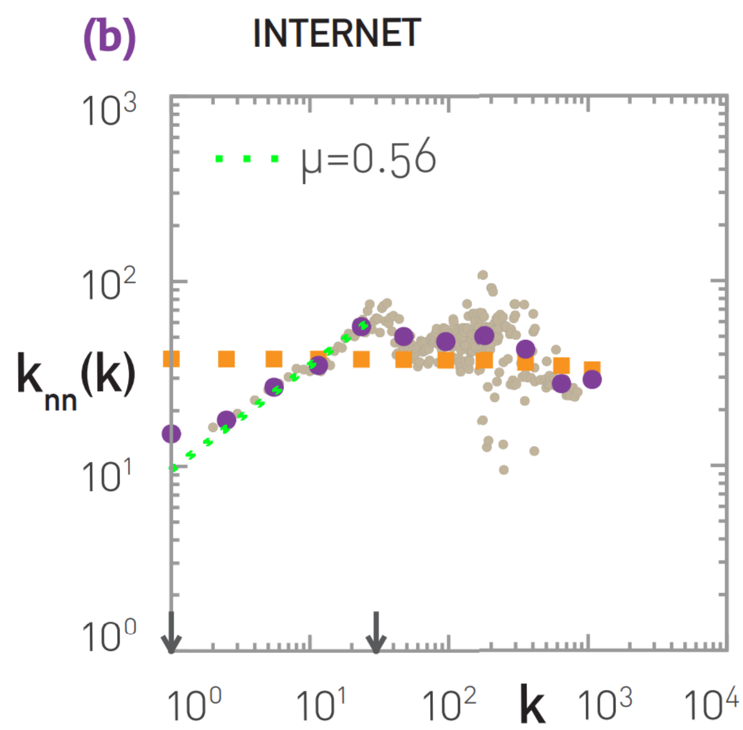


assortative correlations are *real*

disassortative correlations are <u>*not*</u> *real*

What happens in Real Networks





$k_{nn}(k)$ is flat and indistinguishable from its randomized version

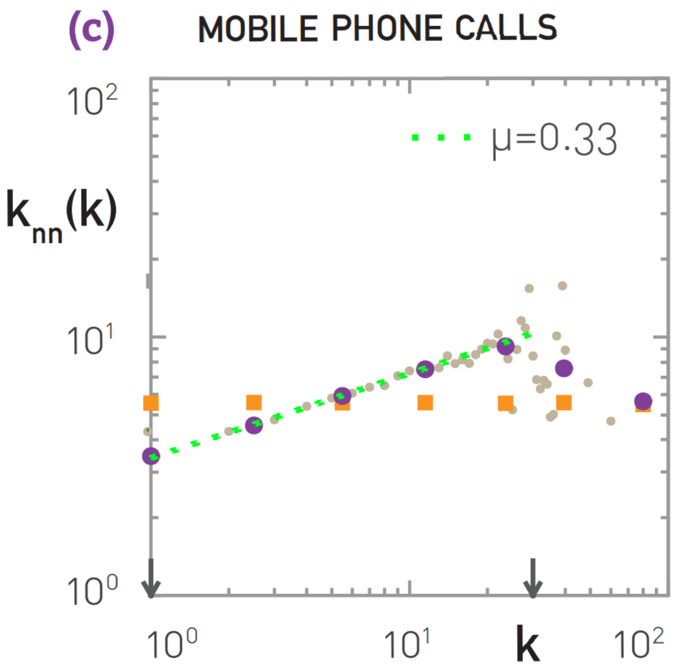
Lack of degree correlations

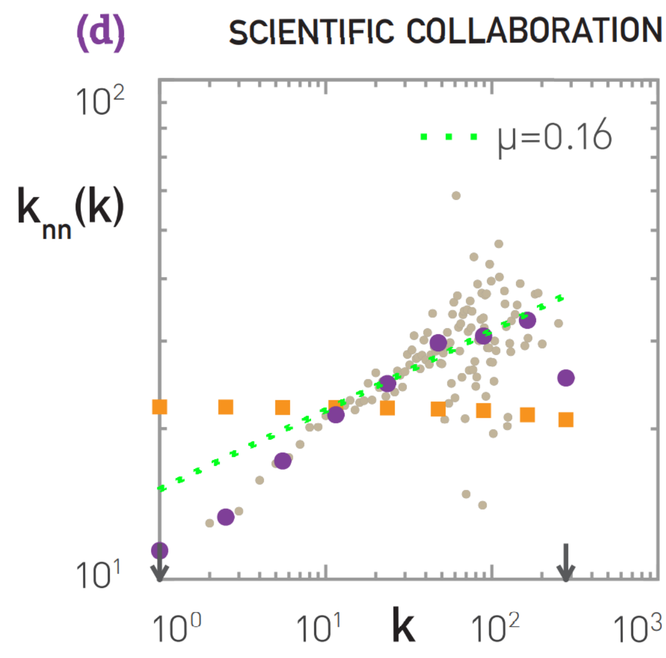
The power grid is neutral

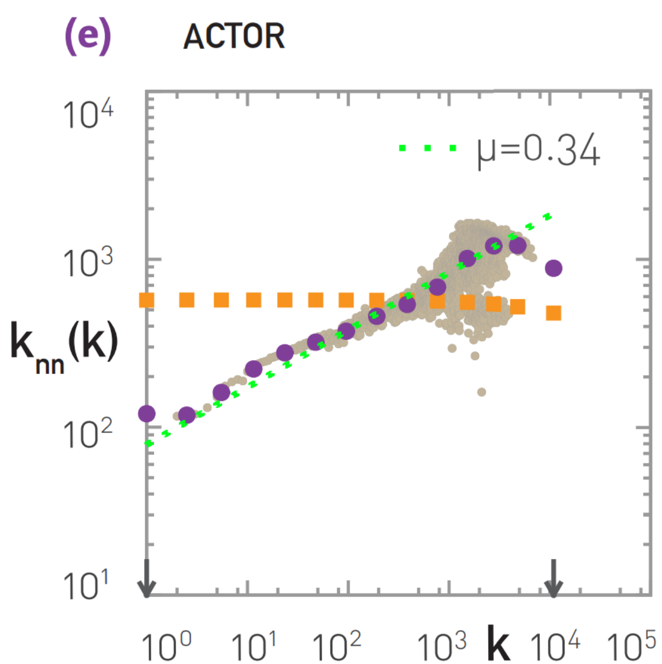
Clear assortative trend for small degrees, The effect levels off for high degrees.

The degree correlations vanish in the randomized version of the Internet

Assortative, but structural cutoffs eliminate the effect for high k

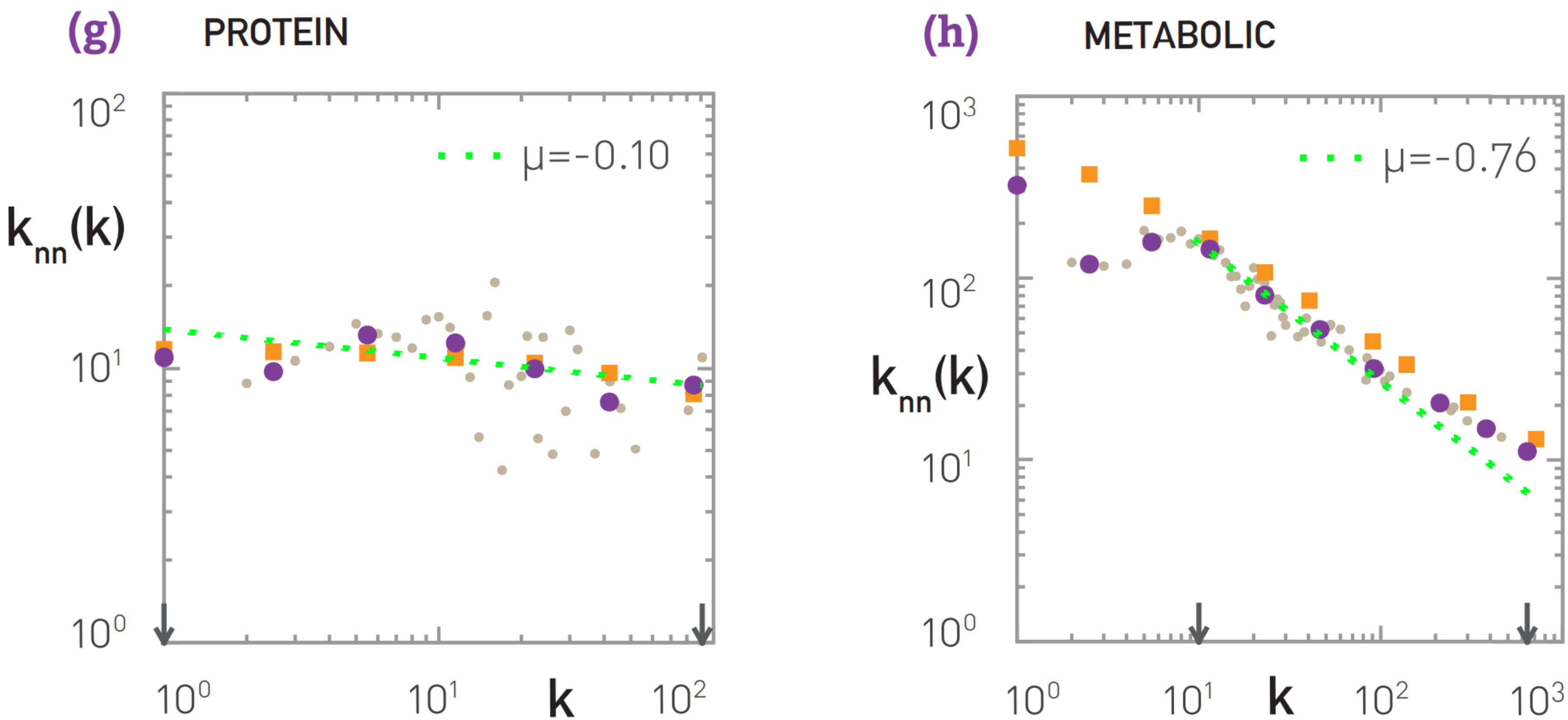






Assortative

Hubs tend to link to other hubs and low-degree nodes tend to link to low-degree nodes.



A negative μ, suggesting disassortativity.
BUT, the observed structural **disassortativity is rooted in the scale-free nature of these networks**

Conclusions:

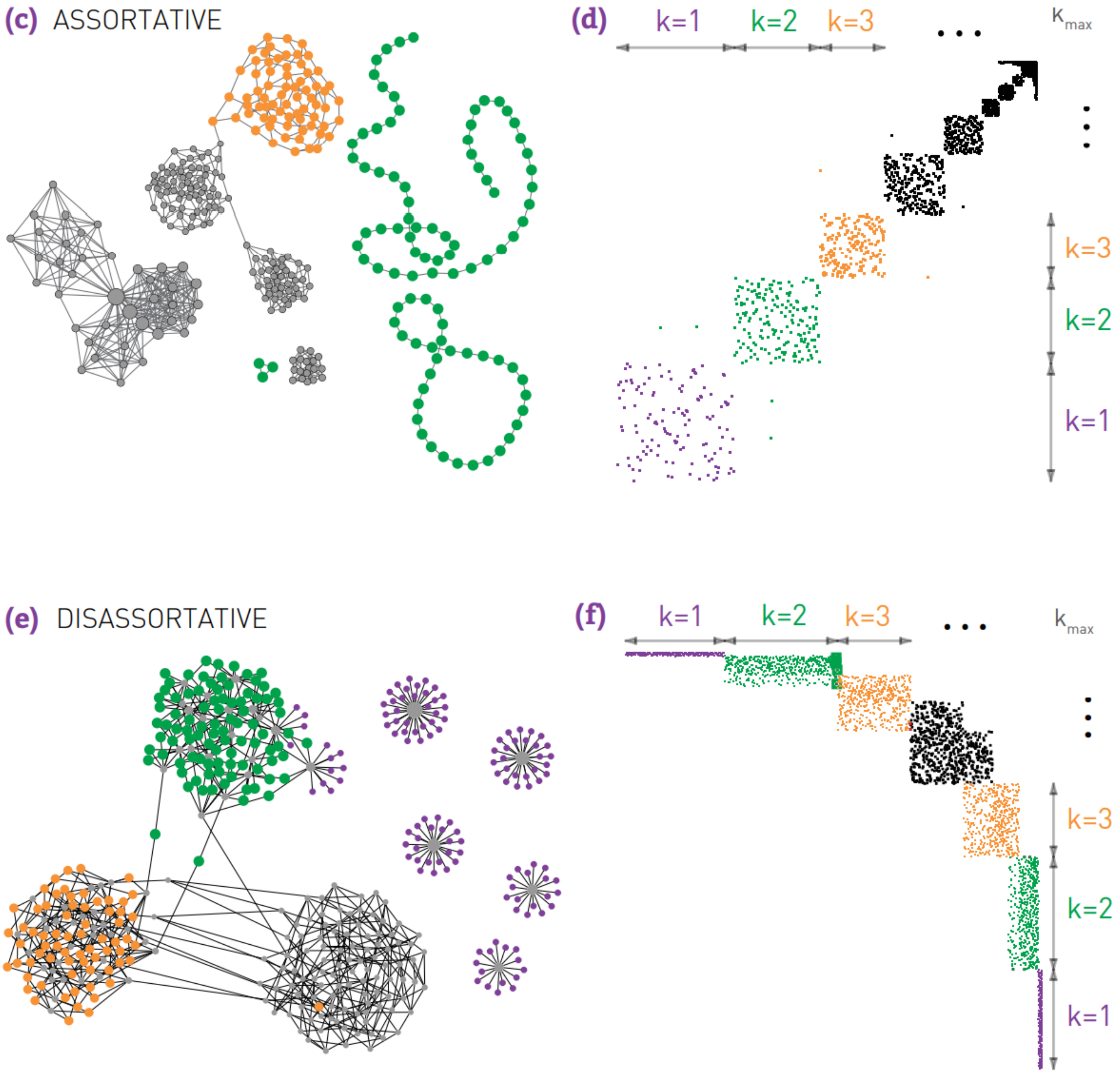
- most real networks display degree correlations -- of the 10 networks, the power grid is the only truly neutral network.
- All networks that display disassortative tendencies (email, protein, metabolic) do so thanks to their scale-free property. Hence, these are all structurally disassortative.
- Only the WWW shows disassortative correlations that are only partially explained by its degree distribution.
- The degree correlations characterizing assortative networks are not explained by their degree distribution.
- Most social networks are assortative and so is the Internet and the citation network.

Generating correlated networks

Our models are neutral if we allow multi-links, and develop structural disassortativity if we force simple networks.

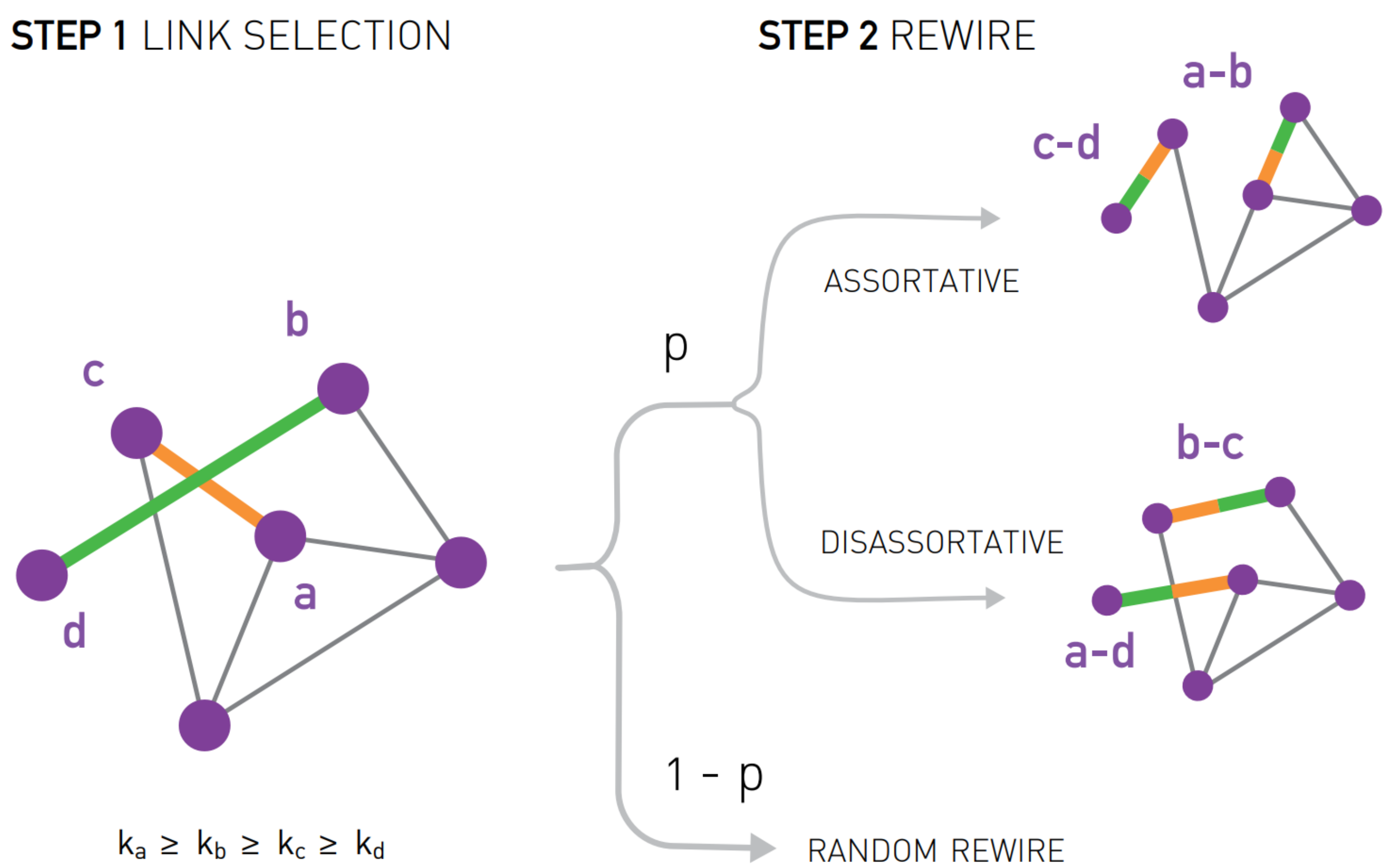
1. Choose at random two links.
2. Label the four nodes at the end of these two links with *a, b, c, d* such that their degrees are ordered as: $k_a \geq k_b \geq k_c \geq k_d$
3. Break the selected links and rewire them. Depending on the desired degree correlations:
   a) Assortative: pair the two highest degree nodes (a with b) and the two lowest degree nodes (c with d)
   b) Disassortative: pair the highest and the lowest degree nodes (a with d and b with c)

[Xalvi-Brunet and Sokolov algorithm]



Rewiring of a Scale-free network
N = 1,000, L = 2,500,
$\gamma = 3.0$.

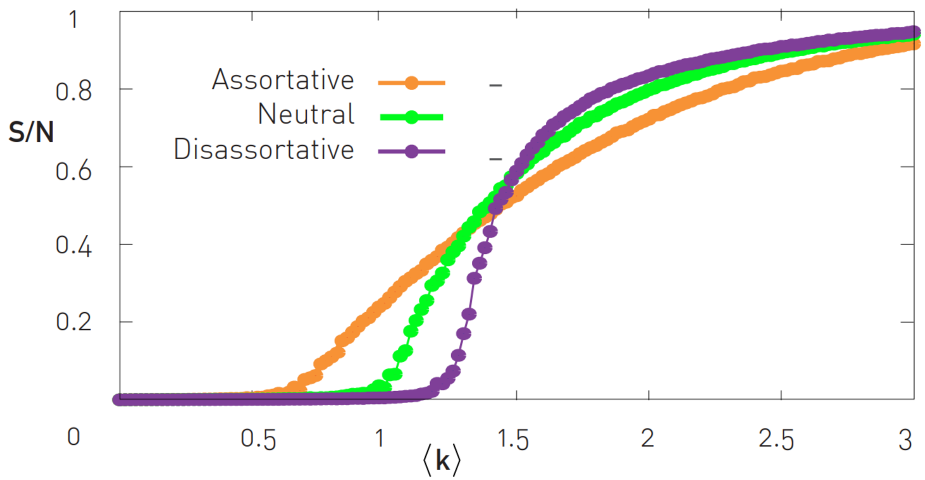
Tuning Degree Correlations



Execute the deterministic rewiring step with probability p, and with probability 1 − p randomly pair the a, b, c, d nodes with each other.

Impact of degree correlations

Relative size of the giant component for networks with different degree correlations



Size and the structure of the giant component have implications on:

- diffusion processes (e.g. spread of diseases)
- robustness
- network distances (assortative: shorter path length)