An alternative view (optional)

Measuring homophily: Modularity

- The network is **assortative** if a significant fraction of the edges in the network run between nodes of the same type.
- **Problem:**
    - this fraction is 1 if all nodes belong to the same single type and 0 if there is a complete separation. We would like a measure that is large in non-trivial cases and small in trivial ones.
- A better measure quantifies the **level of non-randomness in the placement of edges in the network**.
- This is the difference between the **fraction of edges between nodes of the same type and the expected number of edges between all pairs of node of the same type**.

UNIVERSITÀ DI TORINO

Modularity

Group, class or type of node i	$g_i = 1 \dots N$ where N is the total number of groups, classes, or types
Fraction of edges between nodes of the same type	$\sum_{edges(i,j)} \delta_{g_i g_j} = \frac{1}{2} \sum_{ij} A_{ij} \delta_{g_i g_j}$ where $\delta_{g_i g_j} = \begin{cases} 1, & \text{if } g_i = g_j, \\ 0, & \text{if } g_i \neq g_j. \end{cases}$ Kronecker delta
Expected number of edges between all pairs of nodes of the same type	$\frac{1}{2} \sum_{ij} \frac{k_i k_j}{2m} \delta_{g_i g_j}$ where m is the total number of edges
Modularity	$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta_{g_i g_j}$

- **Modularity measures the extent to which similar nodes connect in a network.**
  - It is strictly less than 1 and takes positive values if there are more edges between nodes of the same type than we would expect by random chance.
  - It can also take negative values if there are fewer such edges than we would expect by chance.

Assortative mixing by ordered characteristics

We consider a **scalar characteristic** x_i for the node i, i.e., age or income.

Then consider the pairs of values (x_i, x_j) for the nodes i and j at the ends of each edge in the network and let us calculate their covariance over all edges, i.e. the joint variability of the values over the network.

$$cov(x_i, x_j) = \frac{\sum_{ij} A_{ij}(x_i - \mu)(x_y - \mu)}{\sum_{ij} A_{ij}}$$

$$\mu = \frac{1}{2m} \sum_i k_i x_i$$

- μ is the mean value of x_i at the end of an edge
  - note that it is an average over edges, and since a node with degree k_i lies at the ends of k_i edges, that node appears k_i times in the average

Assortativity coefficient

Then the covariance becomes:

$$cov(x_i, x_j) = \frac{1}{2}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2m})x_i x_y$$

The covariance will be positive if, on balance, values at either end of an edge tend to be both large or both small and negative if they tend to vary in opposite directions.

In other words, the **covariance will be positive when we have assortative mixing and negative for disassortative mixing**

Assortativity coefficient

If we want the covariance to take value 1 in a perfect assortative mixing, we can normalize it by the variance to get the **assortative coefficient** r:

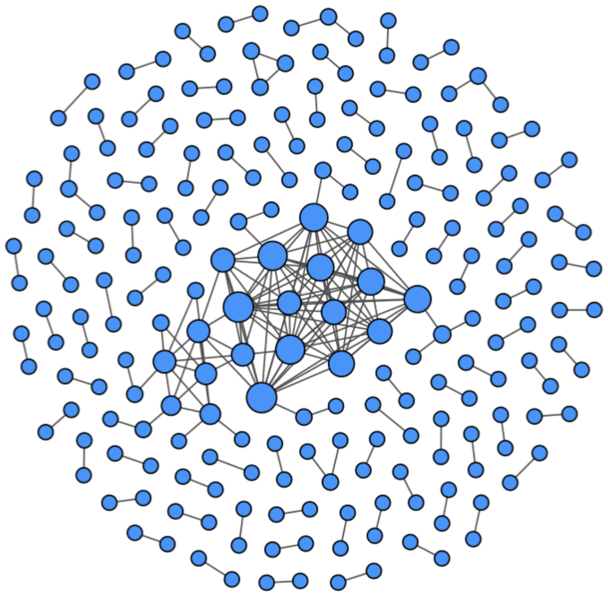
$$r = \frac{\sum_{ij}(a_{ij} - \frac{k_i k_j}{2m})x_i x_y}{\sum_{ij}(a_{ij}\delta_{ij} - \frac{k_i k_j}{2m})x_i x_y}$$

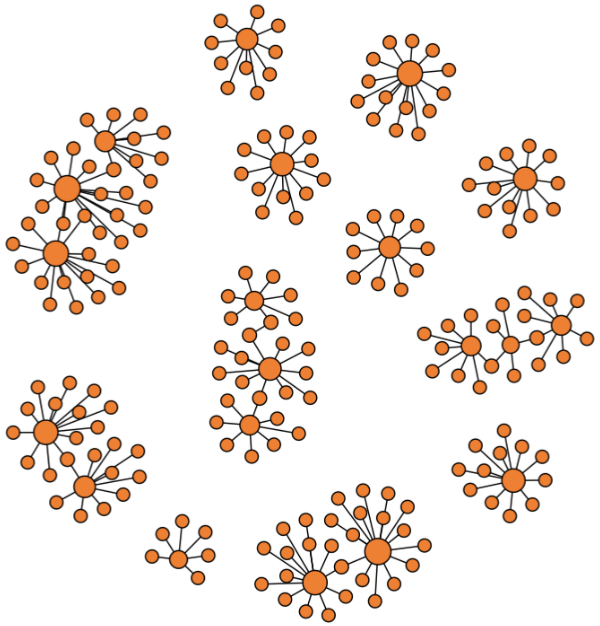
Note: this is an example of a correlation coefficient (Pearson)

The correlation coefficient varies in value between a maximum of 1 for a perfectly assortative network and a minimum of -1 for a perfectly disassortative one.

Degree assortativity

- A.k.a. degree correlation
- Assortative networks have a core-periphery structure with hubs in the core
    - Ex: social networks
- Disassortative networks have hub-and-spoke (or star) structure
    - Ex: Web, Internet, food webs, bio networks





Assortativity coefficient by degree

If we want to consider similar nodes with the same degree, then we can substitute x_i with k_i and obtain the **assortativity coefficient by degree**.

$$r = \frac{\sum_{ij}(a_{ij} - \frac{k_i k_j}{2m})k_i k_j}{\sum_{ij}(A_{ij}\delta_{ij} - \frac{k_i k_j}{2m})k_i k_j}$$

$$cov(k_i, k_j) = \frac{1}{2}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2m})k_i k_j$$

Measuring assortativity by neighbors

Another way to compute the degree assortativity is by measuring the **degree correlation function**:

- **the correlation between the degree and the average degree of the neighbors of nodes with that degree**.

Let's calculate the average of i's neighbors' degree, first:

$$k_{nn}(i) = \frac{1}{k_i}\sum_j A_{ij}k_j$$

With a little abuse of notation, let's define the degree correlation function as the average of all the $k_{nn}(k_i)$ for all the nodes whose degree is k.

$$k_{nn}(k) = \langle k_{nn}(i)\rangle_{i:k_i=k}$$

To visually check the network assortativity, we plot:

$$(k, k_{nn}(k))$$

It is possible to prove that, in a neutral network (i.e., where there is no correlation between a node's degree and its neighbors' average degree), plotting results in a horizontal line.

- Reference:
  - See Chapter 7 of [ns3] for more mathematical insights

UNIVERSITÀ DI TORINO

Example

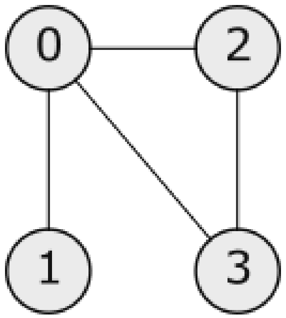
node i with k_i neighbors

$k_{nn}(i) = \frac{1}{k_i} \sum_j a_{ij} k_j$

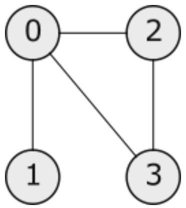
that is the average of i's neighbors' degrees

Let's compute $k_{nn}(i)$ for each node:

$$k_{nn}(0) = \frac{1}{3}(2+2+1) = \frac{5}{3}$$
$$k_{nn}(1) = 3$$
$$k_{nn}(2) = \frac{1}{2}(3+2) = \frac{5}{2}$$
$$k_{nn}(3) = \frac{1}{2}(3+2) = \frac{5}{2}$$

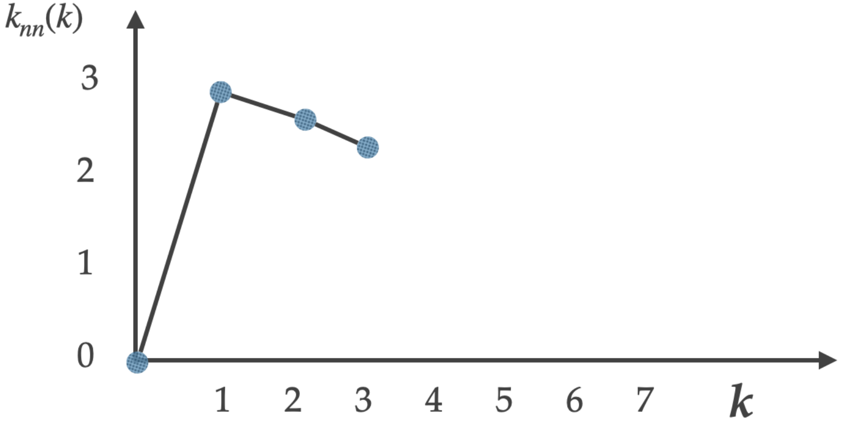


Example



$k_{nn}(0) = \frac{5}{3}, \quad k_{nn}(1) = 3, \quad k_{nn}(2) = \frac{5}{2}, \quad k_{nn}(3) = \frac{5}{2}$

Let us plot $(k, k_{nn}(k))$

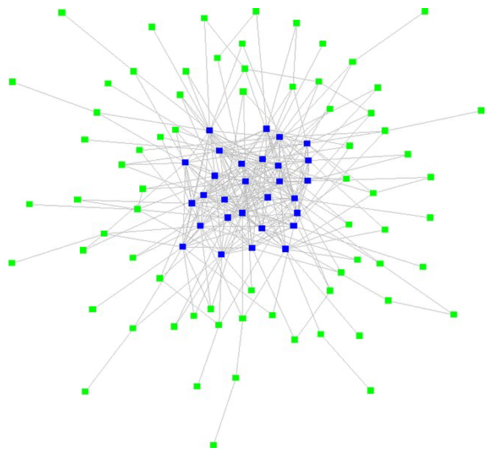


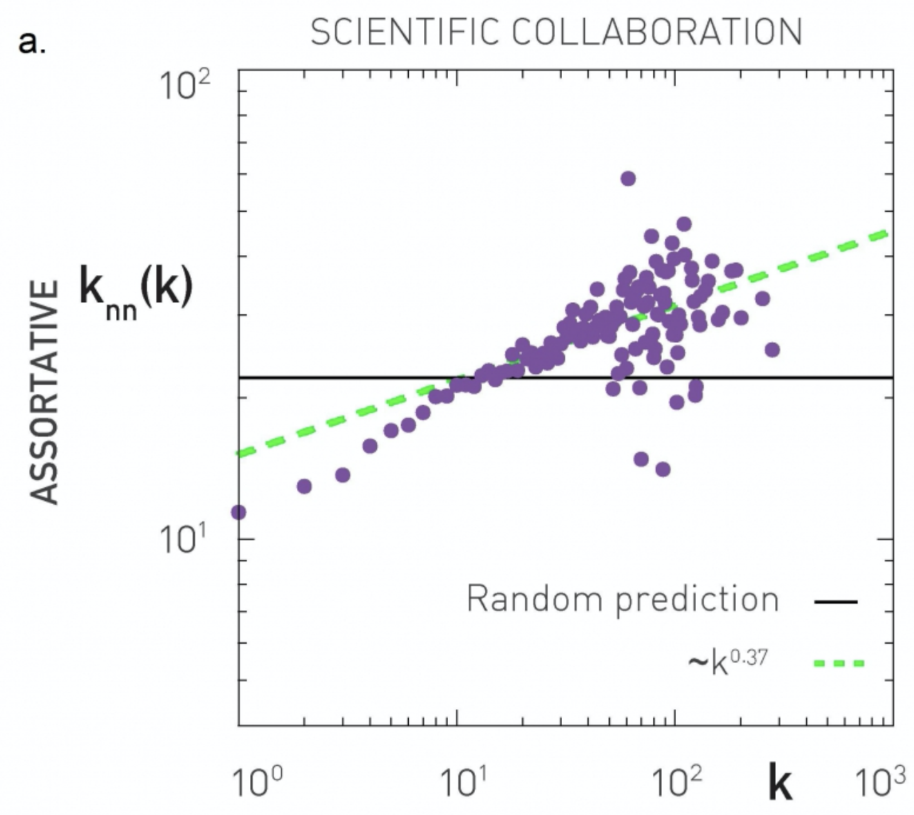
Positive assortativity

- **Scientific collaboration network**

The increase with k indicates that the network is assortative.

- ex. core-periphery

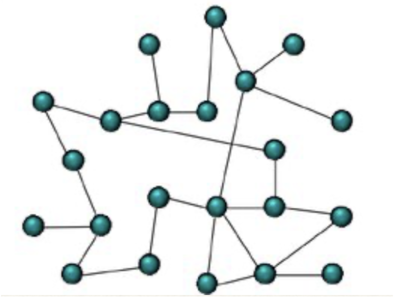


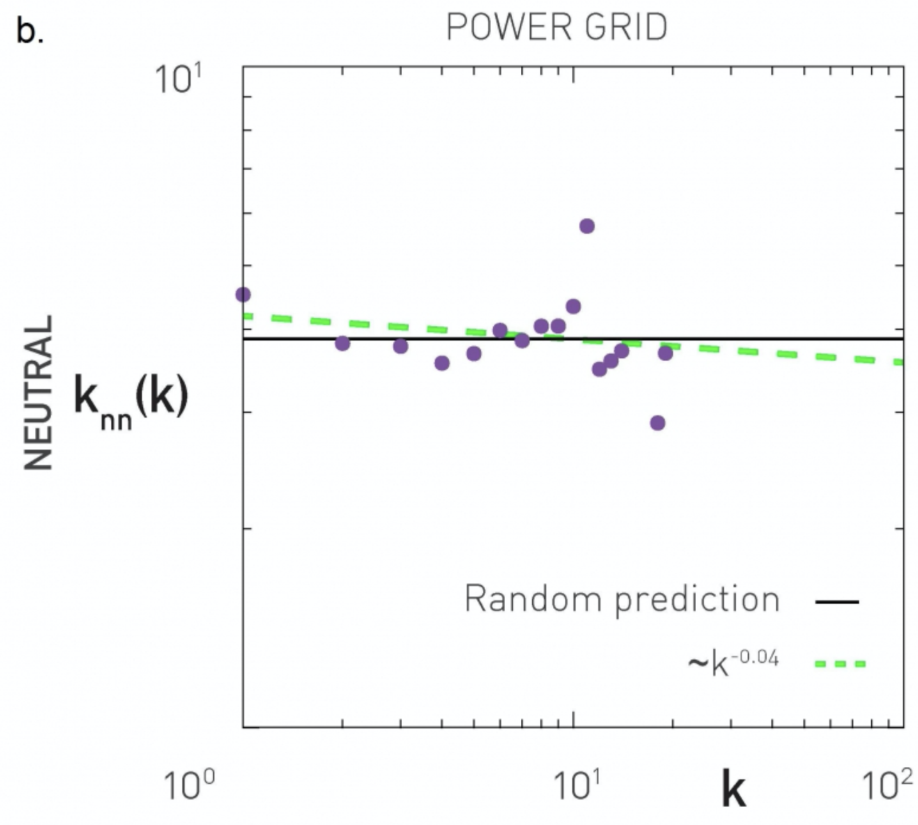


Neutral assortativity

- **Power grid**

The horizontal indicates the lack of degree correlations, in line with our expectations for neutral networks.





Negative assortativity

- **Metabolic Network**

The decreasing documents the network's disassortative nature.