Social Data Management
Degree Correlations

Benoît Groz (slides and content are mostly from Silviu Maniu)

November 10th, 2023

M2 Data Science

Table of contents

Hubs Connecting to Hubs

Previously, we talked about **hubs**, i.e., high degree nodes in social networks.

We saw that new nodes tend to connect to high degree hubs.

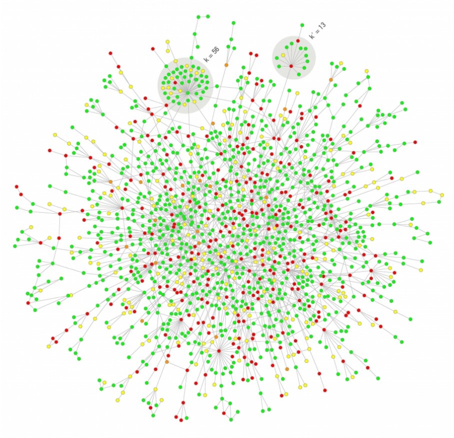
The question we wish to answer now is: **Is this behaviour general? Do nodes connect to other nodes of similar degree?**

Correlations in Social Networks

In social networks, hubs tend to connect to other hubs, e.g., celebrities dating other celebrities.

Correlations in Other Networks

In other networks, hubs tend to connect to very small degree nodes.



Protein interaction network (1870 proteins, 2277 links).

Degree Correlations

In random model, the probability that a given pair of nodes having degree k and k' are neighbours is :

$$p_{k,k'} \approx \frac{k \cdot k'}{2L} \tag{1}$$

- $p_{k,k'}$ predicts that high degree nodes *should* be more likely to connect to other high degree nodes
- in the protein-interaction network, this does not happen: the probability that a high degree node connects to a low degree node is higher than predicted

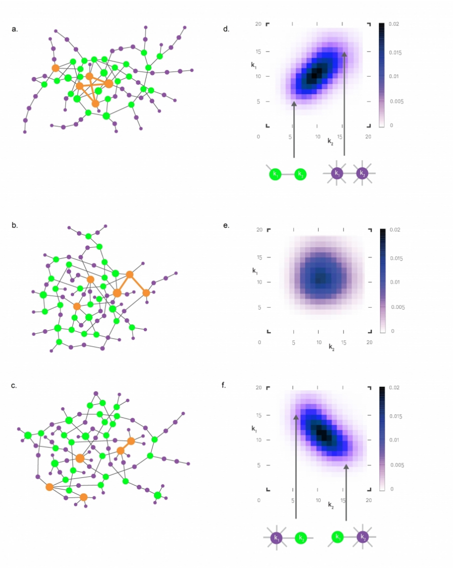
Table of contents

Assortativity and Disassortativity

Depending on how the hubs link to each other, we have 3 types of networks:

1. **Neutral networks**: the wiring is random, links between hubs correspond to the ones expected by chance as in Eq. 1
2. **Assortative networks**: in which nodes tend to connect to other nodes of similar degree
3. **Disassortative network**: networks in which hubs avoid other hubs

Assortativity and Disassortativity



Assortativity and Disassortativity

The information is captured in a degree correlation matrix e where e_{ij} encodes the probability that a node of degree i connects to a node of degree j.

Taking a *random network* the probability that there is a k degree node at the end of a randomly selected link:

$$q_k = \frac{kp_k}{\langle k \rangle},$$

giving

$$e_{ij} = q_i q_j.$$

A network exhibits degree correlations if it deviates from e_{ij}.

Perfect Assortativity

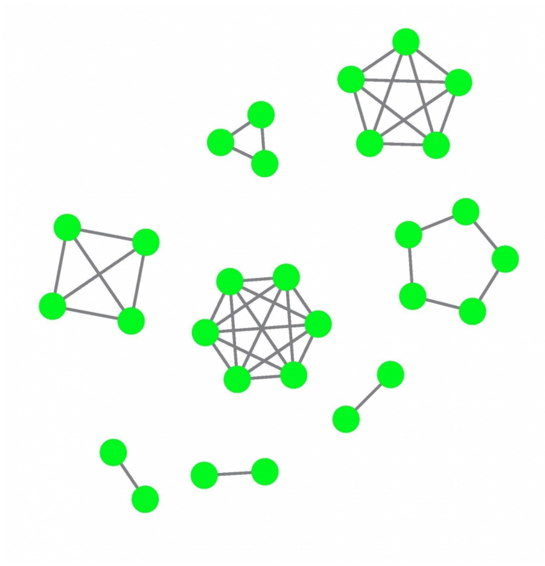


Table of contents

Correlation Function

Degree correlation function:

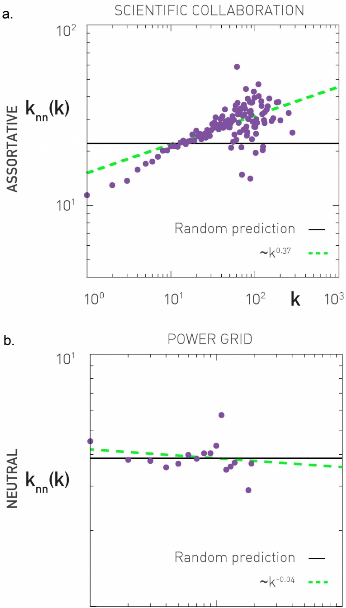
$$k_{nn}(k) = \sum_{k'} k' P(k'|k), \quad (2)$$

where $P(k'|k)$ is the conditional probability that following a link from degree k node we reach a node of degree k'

Depending on the network, we have:

- in neutral networks, $k_{nn}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle}$ – independent of the node's degree and only dependent on the global characteristics of the network;
- in assortative networks hubs tend to connect to other hubs – the higher the degree k is, the higher the avg. degree of the neighbours; and
- in disassortative networks, k_{nn} decreases with k.

Correlation Function



Approximating the Correlation Function

The above figures suggest a function of the form:

$$k_{nn}(k) = ak^{\mu}. \tag{3}$$

The sign of the **correlation exponent** μ characterizes the type of network:

- $\mu > 0$ assortative networks
- $\mu = 0$ neutral networks
- $\mu < 0$ disassortative networks

Degree Correlation Coefficient

We can also capture using a single value, the **degree correlation coefficient**:

$$r = \sum_{jk} \frac{jk(e_{jk} - q_i q_j)}{\sigma^2}, \tag{4}$$

where

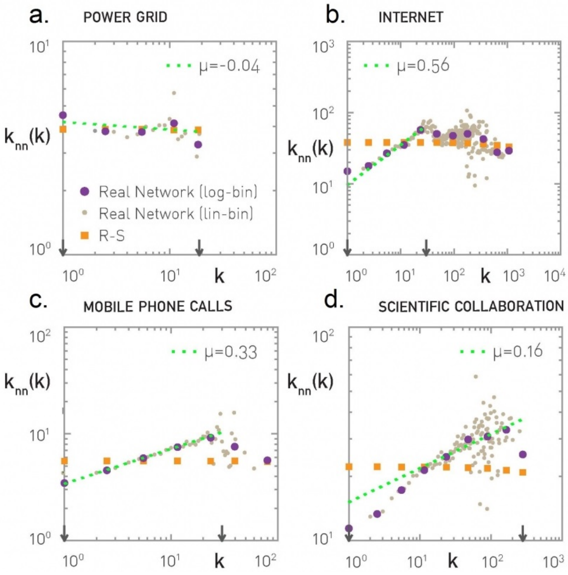
$$\sigma^2 = \sum_k k^2 q_k - \left(\sum_k k q_k \right)^2 .$$

This is equivalent to the **Pearson correlation coefficient** between the degrees of the nodes on each link.

$r \in [-1, 1]$ also characterizes the type of network:

- $r < 0$ assortative networks
- $r = 0$ neutral networks
- $r > 0$ disassortative networks

Correlations in Real Networks



Correlations in Real Networks

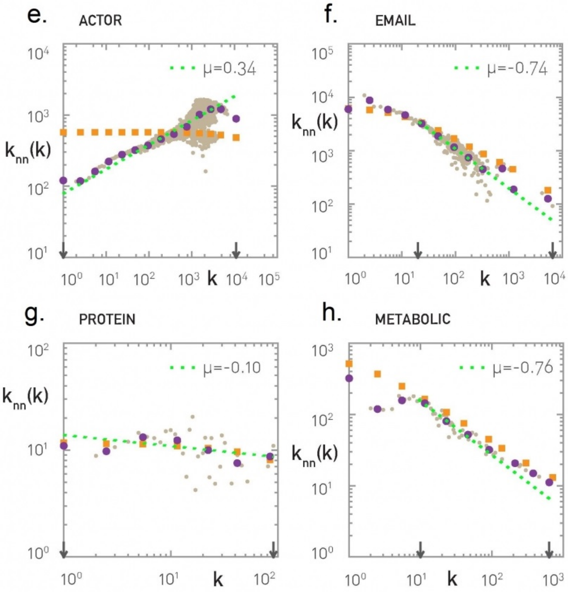
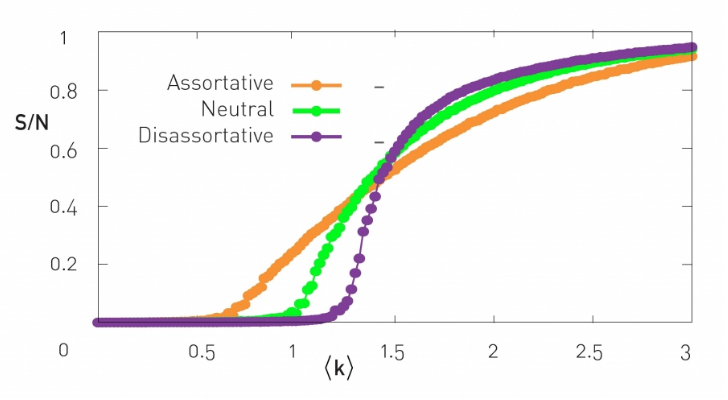


Table of contents

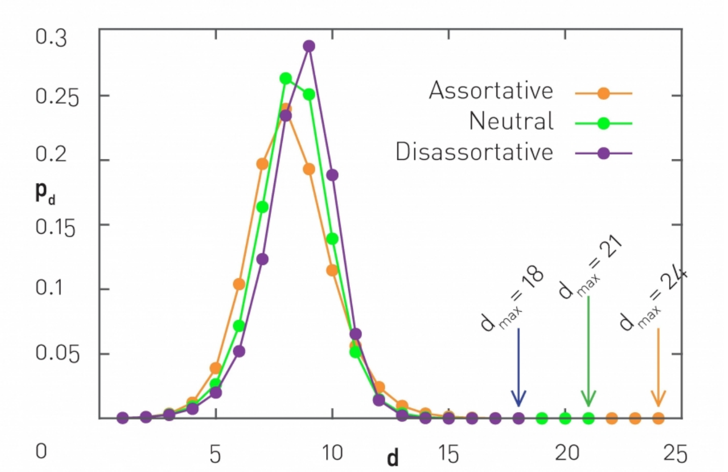
Giant Component

How does the assortativity influence the network? Depending on $\langle k \rangle$, the size of the giant component appears at different time steps – influence on **network robustness**



Other Consequences

- average path length is lower in assortative networks
- degree correlations influence stability (perturbations, stimuli)
- they influence greatly the cost of the vertex cover problem



Acknowledgments

Figures in slides 5, 9, 11, 14, 17, 18, 20, and 21 taken from the book "Network Science" by A.-L. Barabási. The contents is partly inspired by the flow of Chapter 7 of the same book.

`http://barabasi.com/networksciencebook/`

References i

Park, J. and Newman, M. E. J. (2003).
Origin of degree correlations in the internet and other networks.
Phys. Rev. E, 68.

Pastor-Satorras, R., Vázquez, A., and Vespignani, A. (2001).
Dynamical and correlation properties of the internet.
Phys. Rev. L., 87(25).