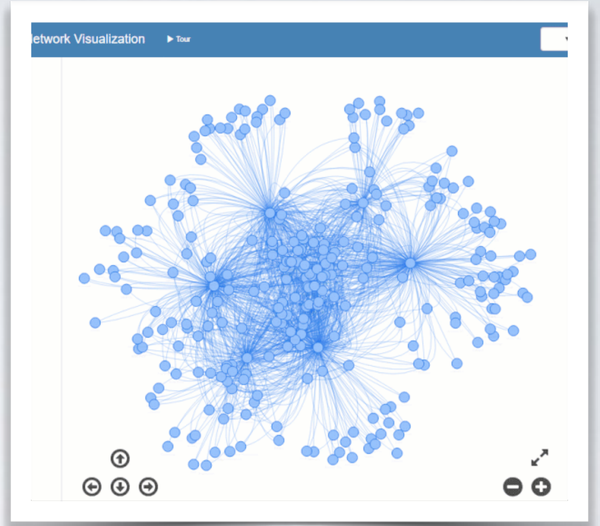
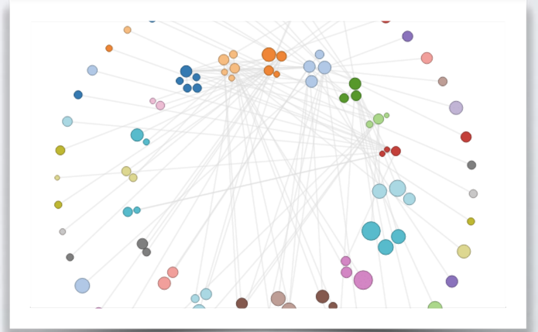
NETWORK VISUALISATION

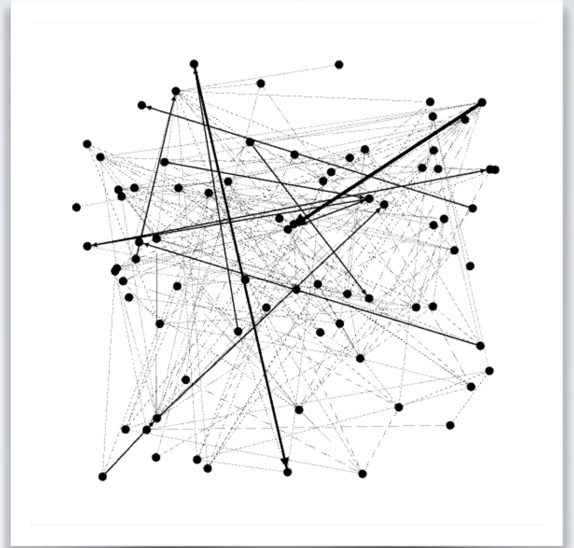
NETWORK VISUALIZATION



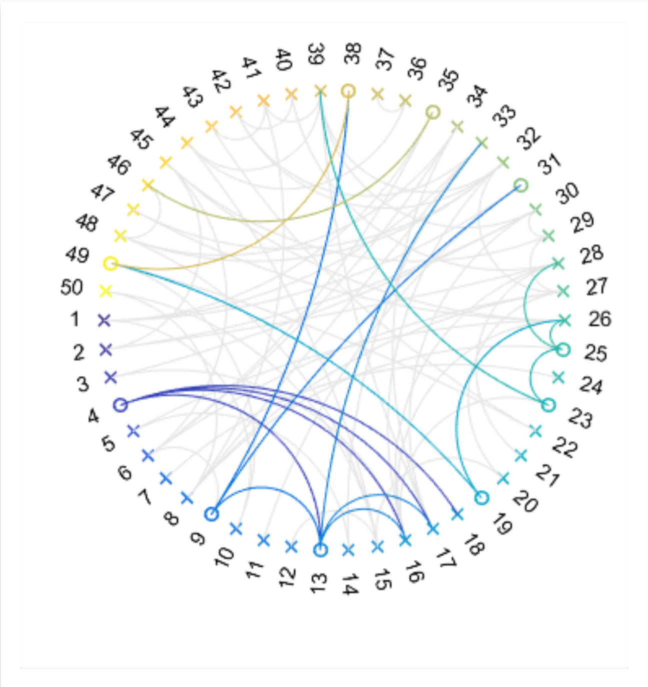
- How to interpret a network drawing?
- What does the position of nodes means?
- Can we draw conclusion from the drawing alone?



NETWORK VISUALIZATION

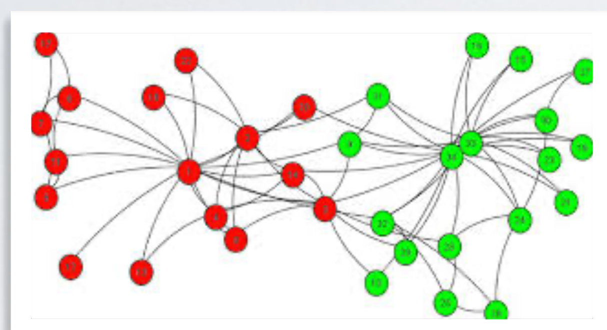


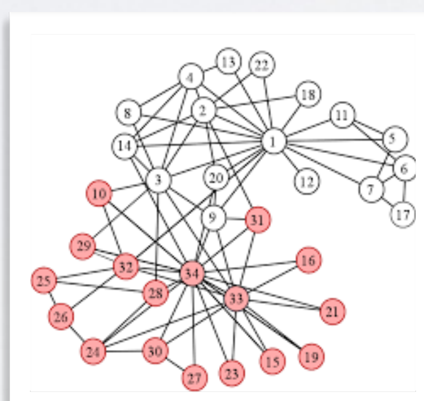
- Random layout
  - Assign random positions to nodes, draw edges
    - Useless for more than 5-6 nodes
- Geographical layout
  - The position of nodes is fixed a priori, often based on geographical location
  - Variant: position nodes on a circle based on a single, 1D property (age…)

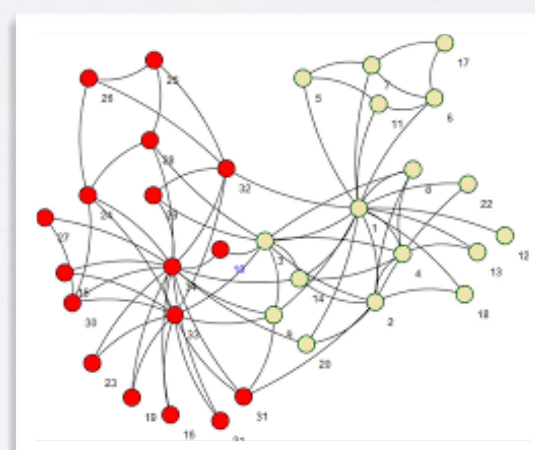


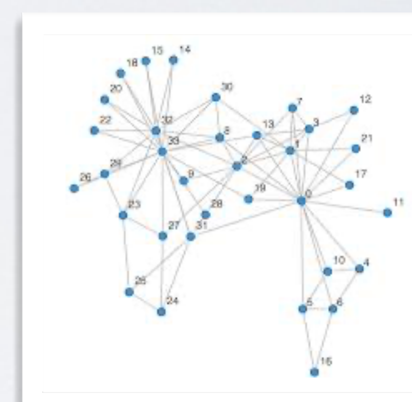
NETWORK VISUALIZATION

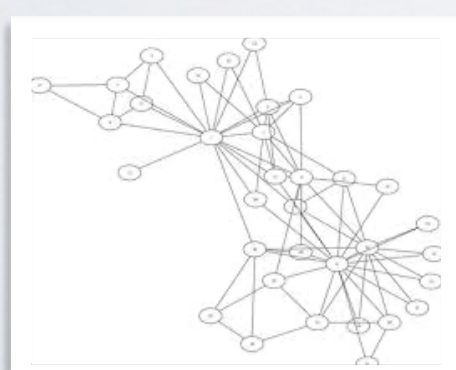
- Most commonly used: Automatic layout
  - Non deterministic
  - Tries to arrange nodes so that the network is easy to read and understand
    - Minimize edge crossings?
    - Most commonly, tries to put connected nodes close and unconnected nodes far

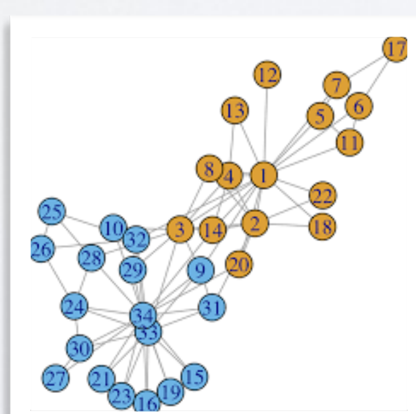












NETWORK VISUALIZATION

- Most common algorithms are variant of the **force directed** layout: physical system of bodies with forces acting on them.
  - Objective: minimize the energy of the system.
  - Fruchterman-Reingold: Spring+ repulsive forces
  - Kamada-Kawai: Springs+length proportional to graph distance
  - Gephi: Force Atlas (custom model)

- Principles of those models
  - Repulsive forces between nodes
  - Edges are attracting forces
  - Minimal (to avoid node overlap) and maximal (to avoid connected component drifting out of the figure) distances can be added.

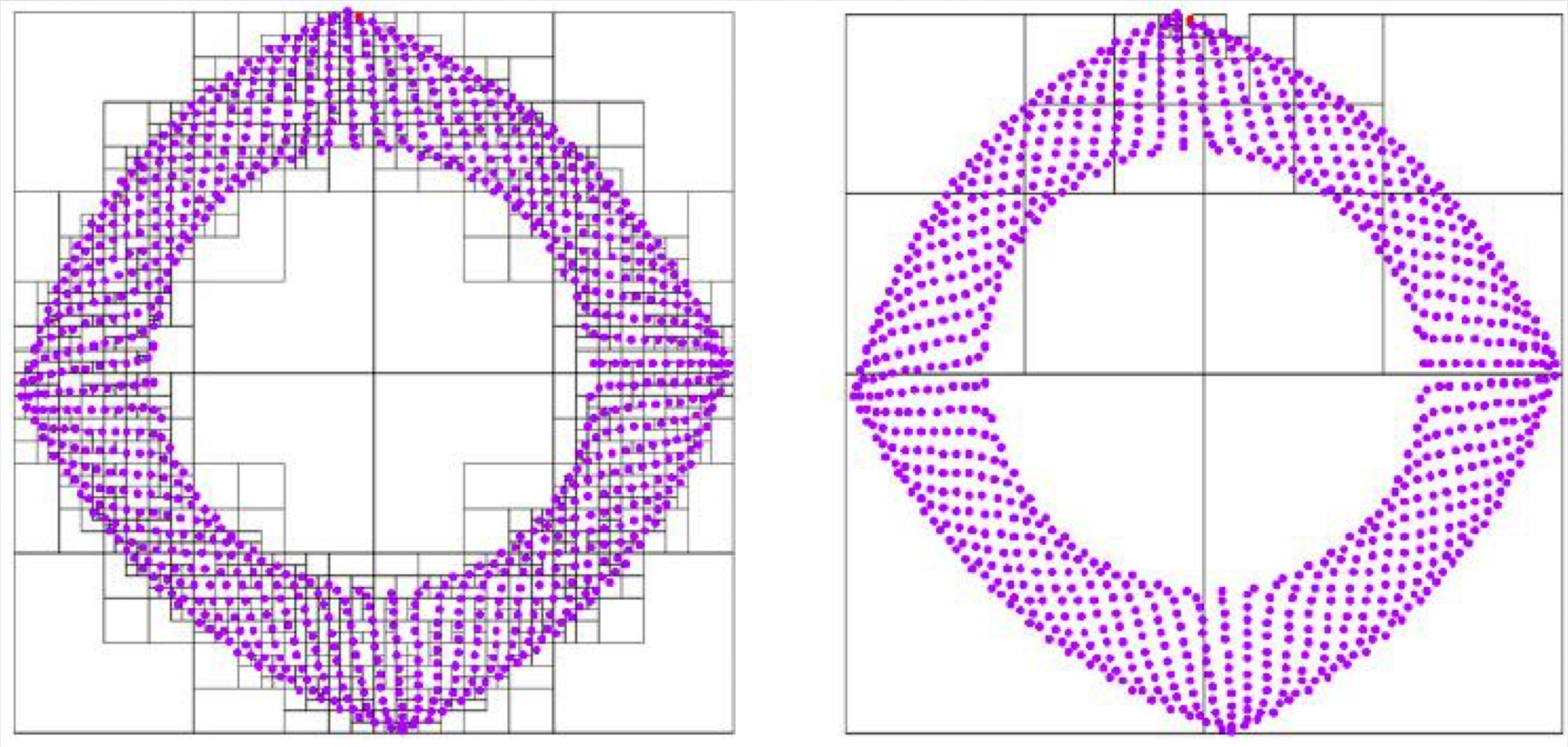
NETWORK VISUALIZATION

- Example: Kamada kawai
- $f(i, j) = ||x_i - x_j|| - d(i, j)$
  - x_i: position of node, $d(i, j)$: graph distance
- Energy of the system:
  - $\sum_{i \neq j} (||x_i - x_j|| - d(i, j))^2$

NETWORK VISUALIZATION

- Naive algorithm:
  - While (not converged)
    - For each node, compute forces on it and update position accordingly
- Problem: repulsive forces are among all pairs of nodes
  - Complexity $\mathcal{O}(n^2)$
  - Solution: multiscale computations…

NETWORK VISUALIZATION



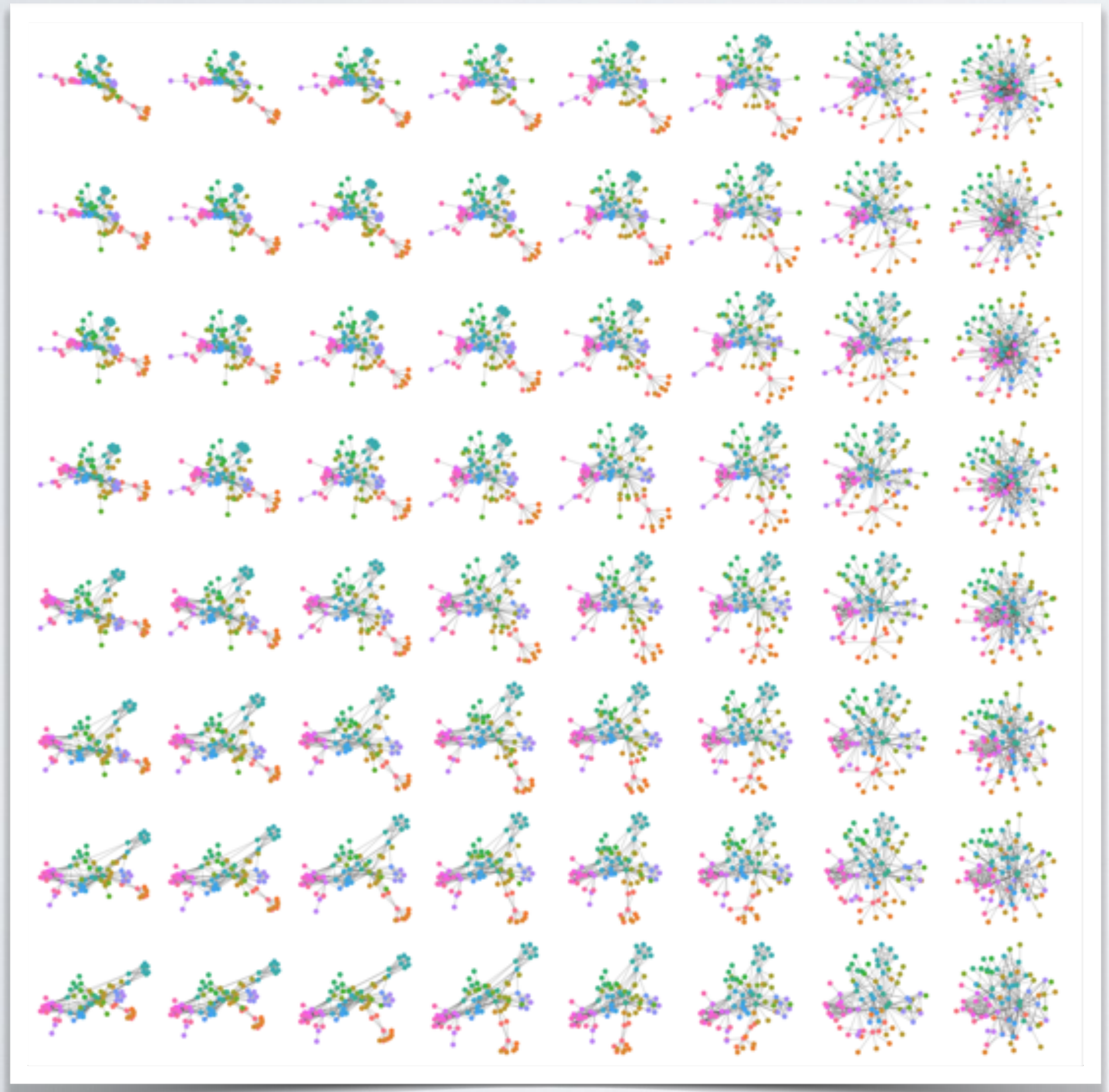
https://people.cs.clemson.edu/~isafro/na13/l22.pdf

NETWORK VISUALIZATION

- More recently, approaches using **graph embedding**:
- Maximize similarity between a notion of distance in the graph and the distance in the drawing
  - Graph distance can simply be number of hops, but also probability to reach by random walks, complex notion including communities, etc.

NETWORK VISUALIZATION

http://kwonoh.net/dgl/

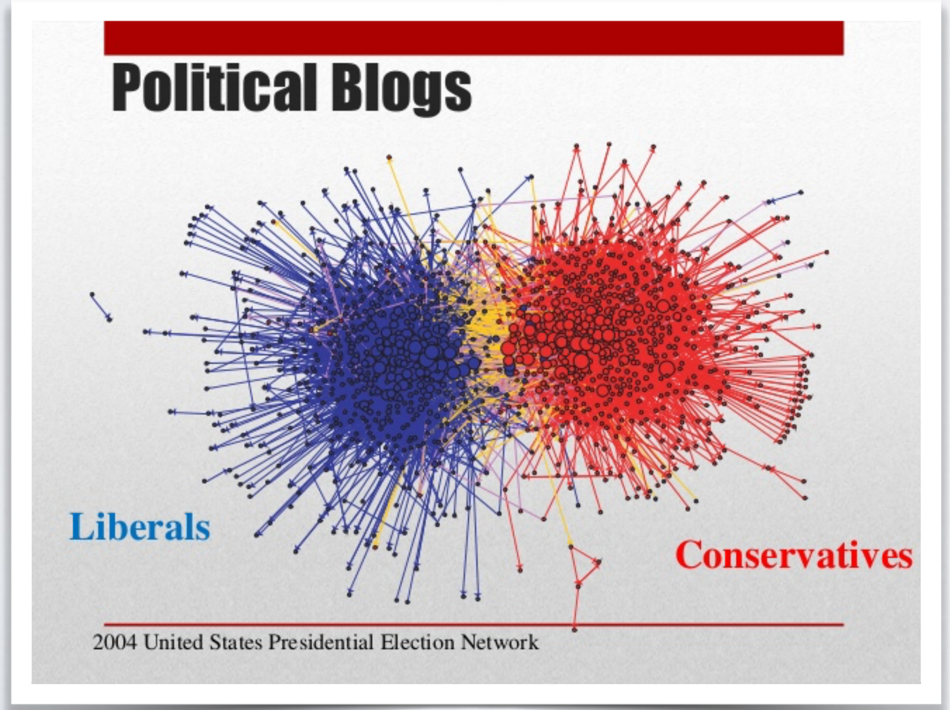


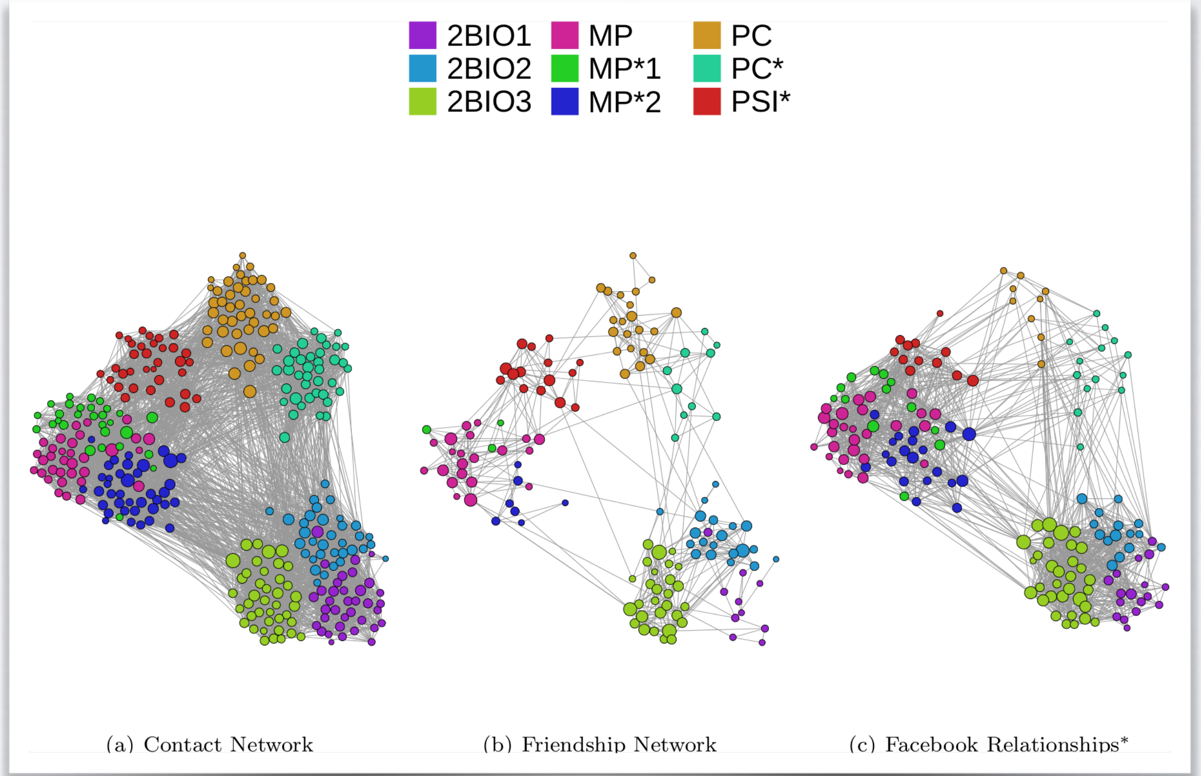
Deep Neural networks

GAN(Generative Adversarial Networks) approach
To "generalise" several existing layouts

NETWORK VISUALIZATION

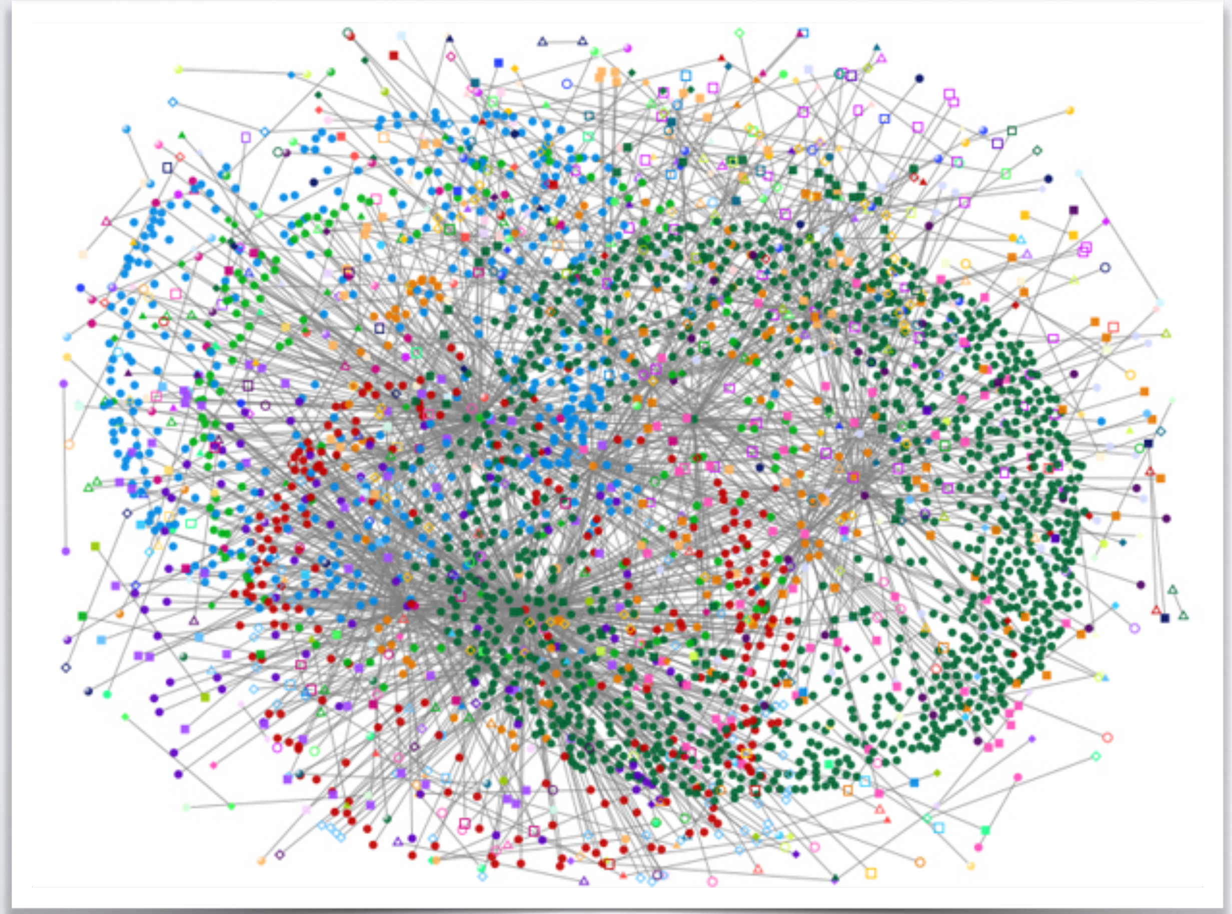
- Can we interpret a force layout?
  - Yes…

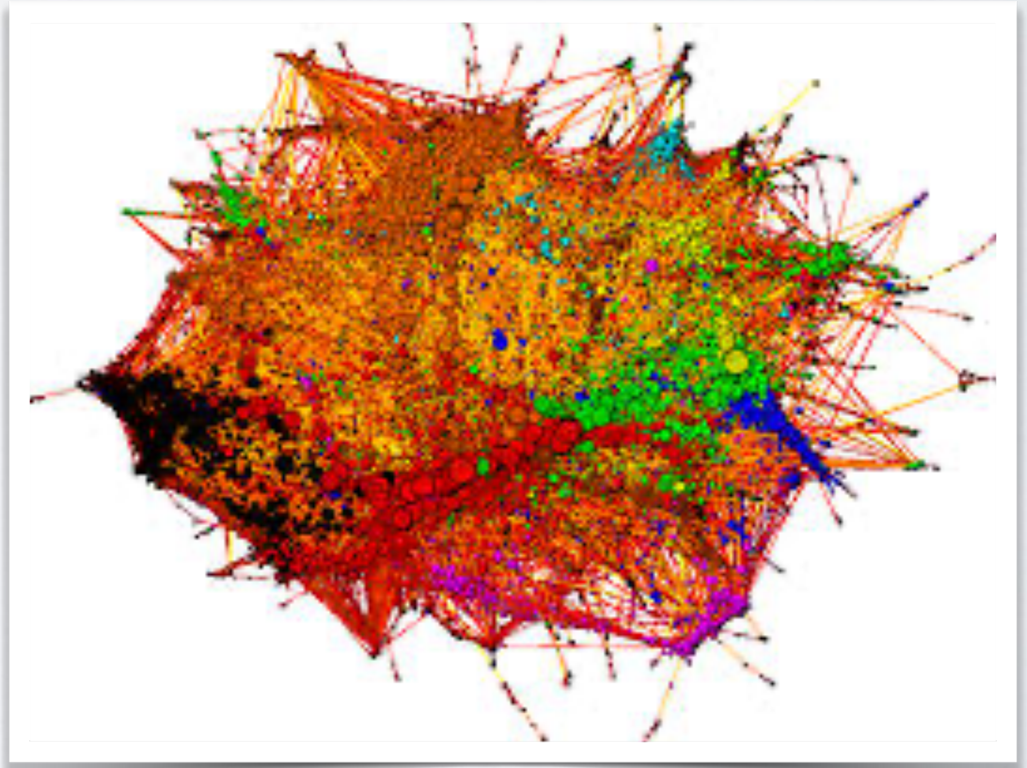




NETWORK VISUALIZATION

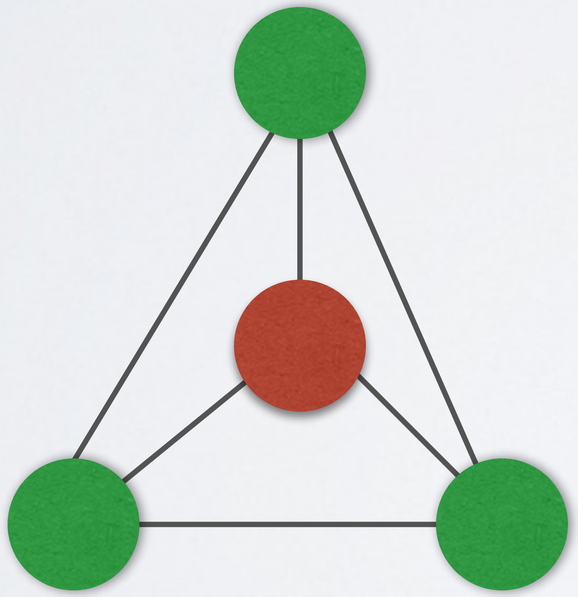
- Can we interpret a force layout?
  - Yes…
  - And no.

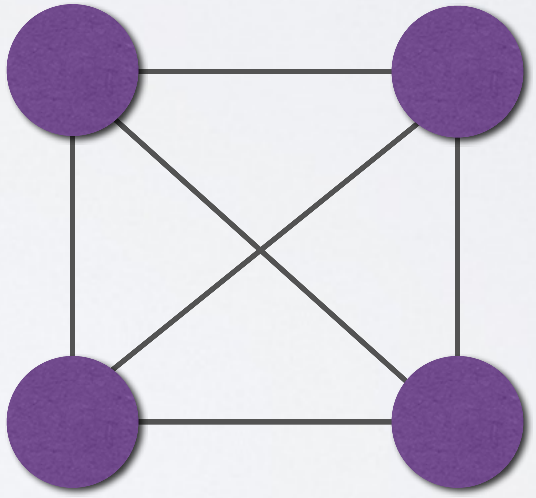




NETWORK VISUALIZATION

- Can we interpret a force layout?
  - Yes…
  - And no.





ASSORTATIVITY - HOMOPHILY

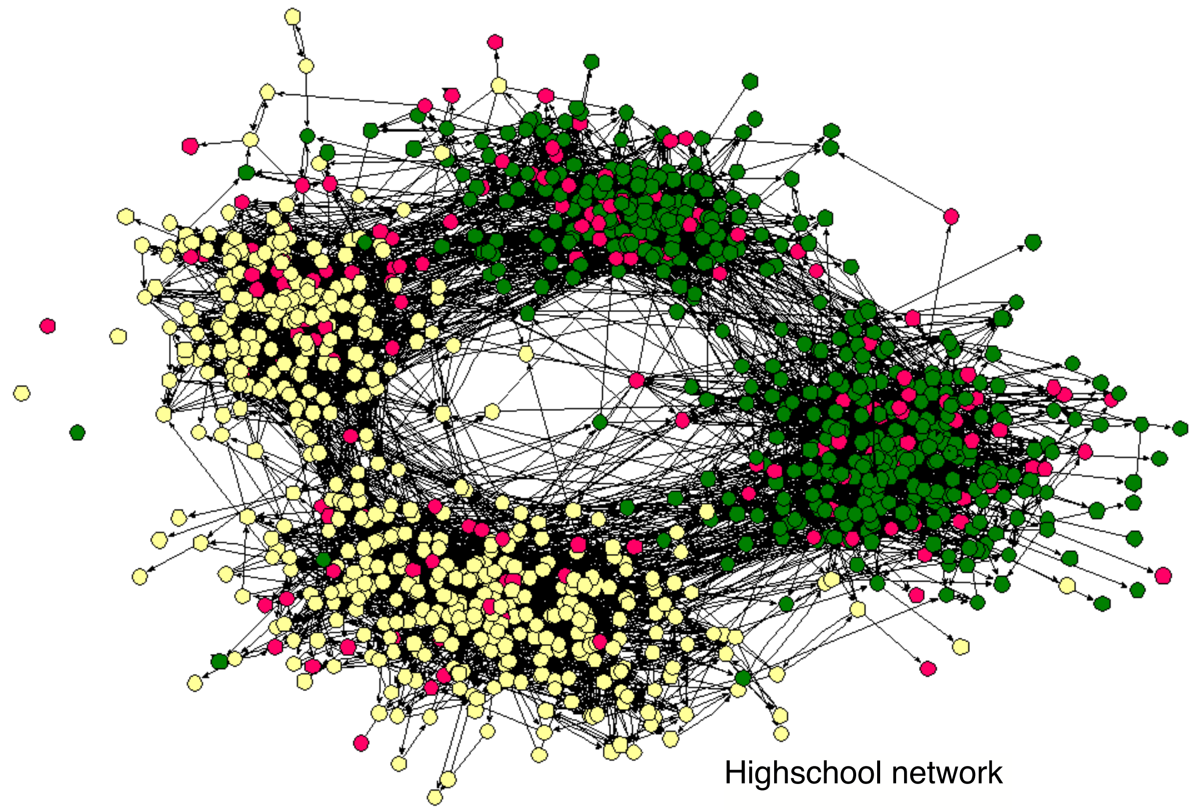
Homophily - Assortativity

"birds of a feather flock together"

- Property of (social) networks that nodes of the same attitude tends to be connected with a higher probability than expected
- It appears as correlation between vertex properties of $x(i)$ and $x(j)$ if $(i,j) \in E$

Vertex properties

- age
- gender
- nationality
- political beliefs
- socioeconomic status
- habitual place
- obesity
- …



Highschool network

Colored by ethnic groups (J Moody)

Homophily - Assortativity

Note on interpreting homophily

Homophily can be a link creation mechanism (nodes have a preference to connect with similar ones, so the network end up to be assortative), or a consequence of influence phenomenons (because nodes are connected, they tend to influence each other and thus become more similar).
Without access to the dynamic of the network and its properties, it is not possible to differentiate those effects.

Homophily - Assortative mixing

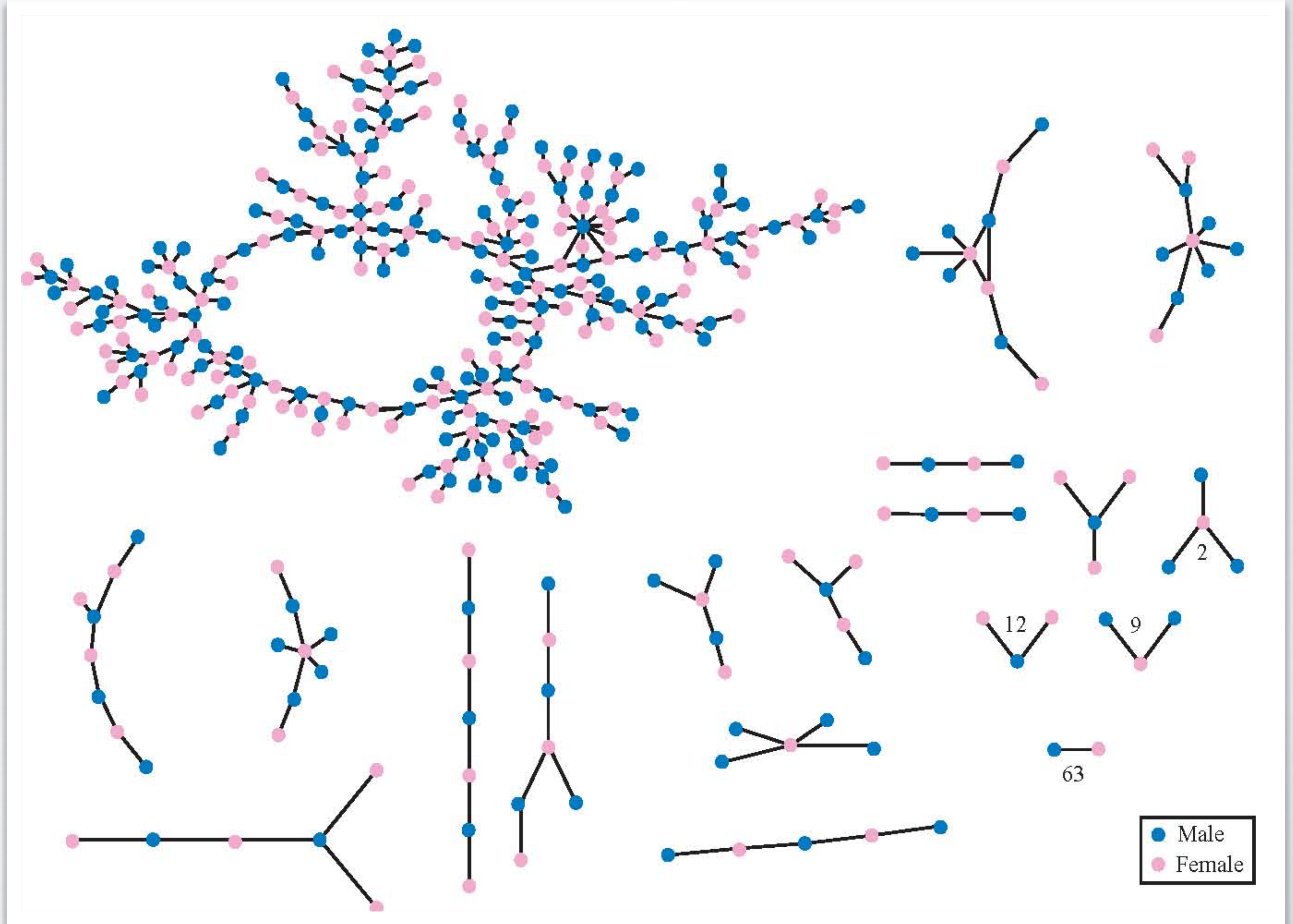
"Opposites attract"

Disassortativity - Heterophily

- Opposite of homophily: dissimilar nodes tend to be connected

Examples

- Sexual/Sentimental networks
- Predator - prey ecological networks



Homophily - Assortative mixing

To quantify homophily

- We can take into account…
  - **Categorical (Enumerative) attributes**: vertex features which are comparable but not quantifiable (e.g., gender, ethnicity, colour(of goods to sell..), shape, etc.)
  - **Scalar attributes**: vertex features which are comparable and sortable (age, weight, income, degree, …)

Homophily - Assortative mixing

Categorical attributes

$$r = \frac{\sum_i e_{ii} - \sum_i a_i^2}{1 - \sum_i a_i^2}$$

e_{ii}: fraction of edges between nodes with same attributes

a_i: fraction of all edges having at least an end with property i.

=>Sum of degrees of nodes with property i divided by L

No assortative mixing : r=0 ($e_{ij} = a_i^2$)

Perfectly assortative: r=1

Assortative: r>0

Homophily - Assortative mixing

Assortativity index - Example

Let's see a fictional example of how to compute the assortativity index. Nodes are individuals, edges represent for instance some social interaction. Columns/Rows correspond to blood types, and numbers are expressed in fraction of the total number of edges.

Blood Types	A	AB	B	O	a_i
A	0.30	0.05	0.1	0.05	*0.5*
AB	0.05	0.05	0	0	*0.1*
B	0.1	0	0.2	0	*0.3*
O	0.05	0	0	0.05	*0.1*
a_i	*0.5*	*0.1*	*0.3*	*0.1*	**1**

$$r = \frac{(0.3+0.05+0.2+0.05)-(0.5^2+0.1^2+0.3^2+0.1^2)}{1-(0.5^2+0.1^2+0.3^2+0.1^2)} = \frac{0.6+0.36}{1-0.36} = 0.375$$

Homophily - Assortative mixing

$$r = \frac{\sum_i e_{ii} - \sum_i a_i^2}{1 - \sum_i a_i^2}$$

Assortativity and Modularity

Assortativity is related to the Modularity, a measure of the quality of *communities*, by the following relation:

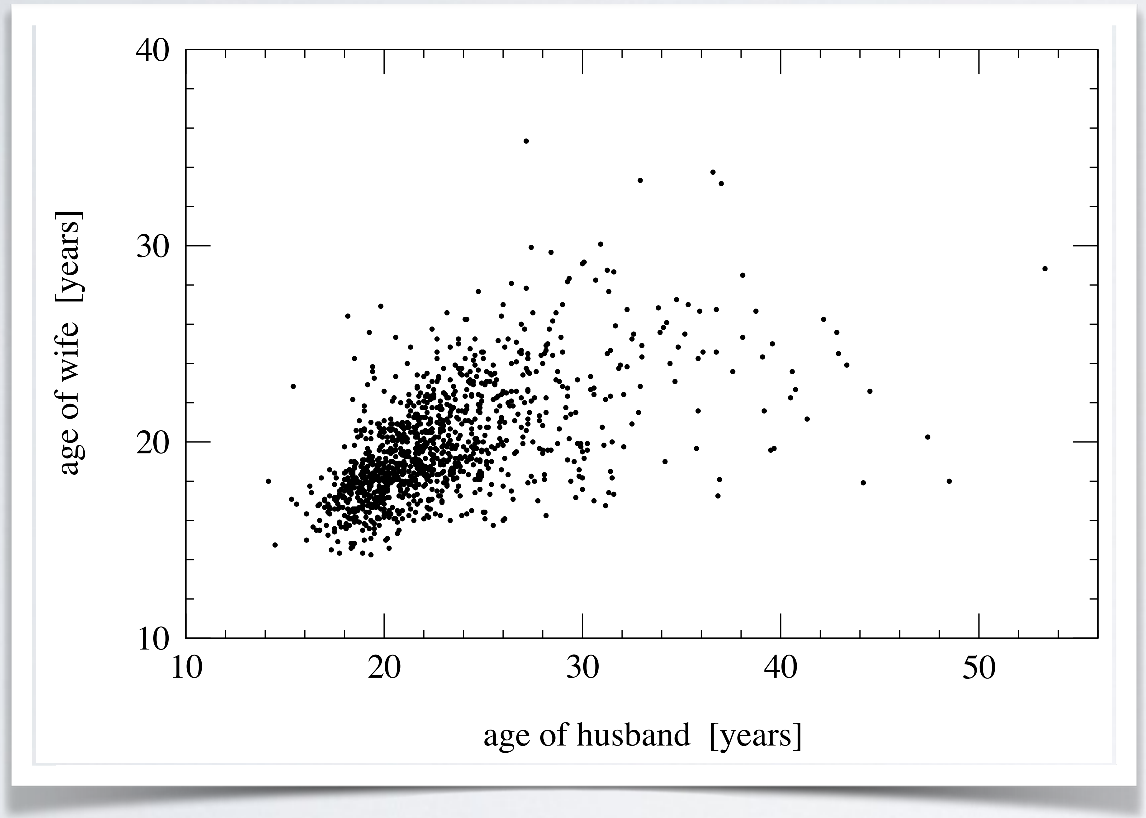
$$r = \frac{Q}{Q_{max}}$$

Indeed, $\sum_i e_{ii} - \sum_i a_i^2$ corresponds to the definition of the Modularity, while $1 - \sum_i a_i^2$ corresponds to the maximal value that the Modularity could reach if all nodes were in the same communities.

Homophily - Assortative mixing

Numeric attributes

Pearson correlation coefficient of properties
at both extremities of edges



e_{xy}: fraction of edges joining nodes with values x and y

$$\sum_{xy} e_{xy} = 1, \qquad \sum_{y} e_{xy} = a_x, \qquad \sum_{x} e_{xy} = b_y$$

$$r = \frac{\sum_{xy} xy(e_{xy} - a_x b_y)}{\sigma_a \sigma_b},$$

with σ_a standard deviation of a_x

(Here, discrete version)

Limit of assortativity coefficient

Limits of Assortativity

A limit of assortativity coefficients as we have defined them is that they summarize the whole network as a single value. However, different parts of the network might have different types of assortativity.

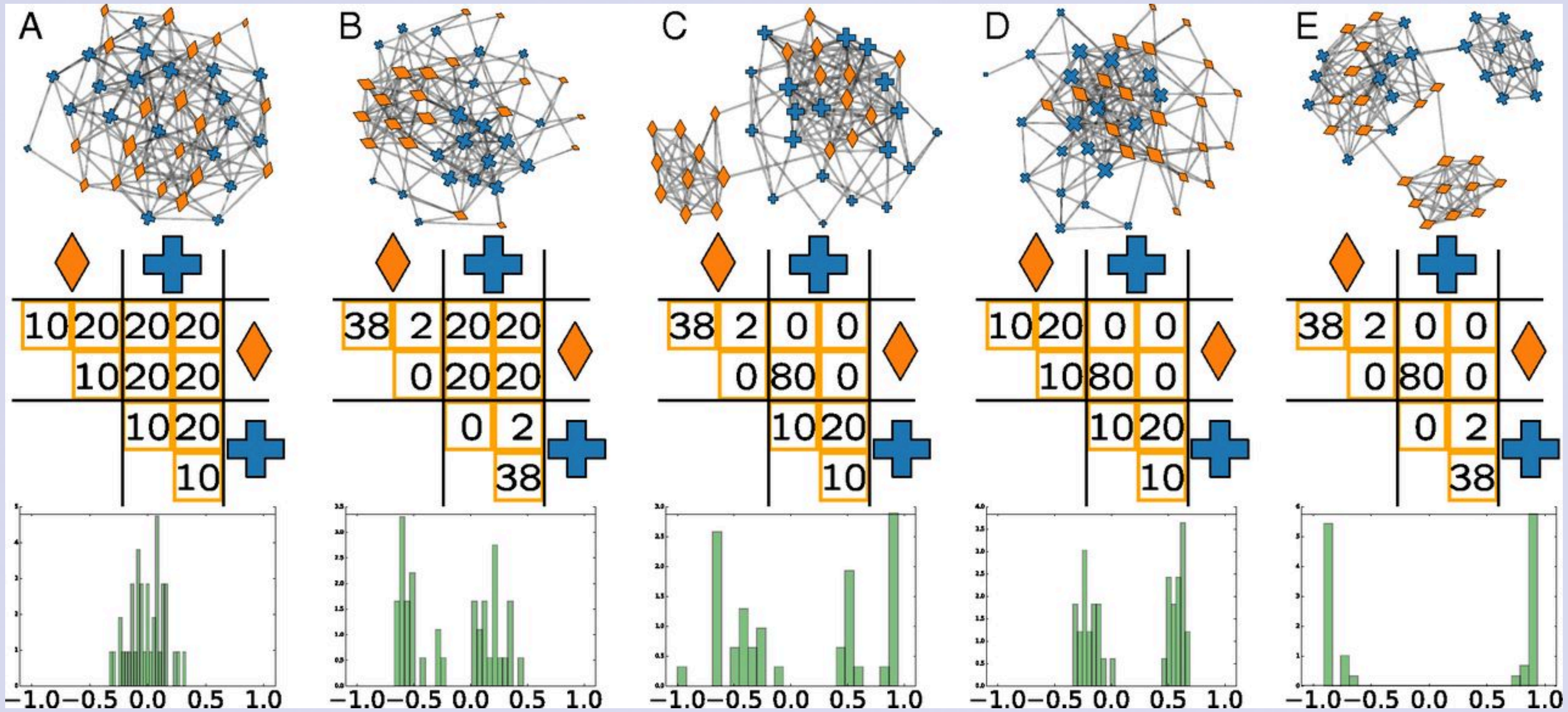
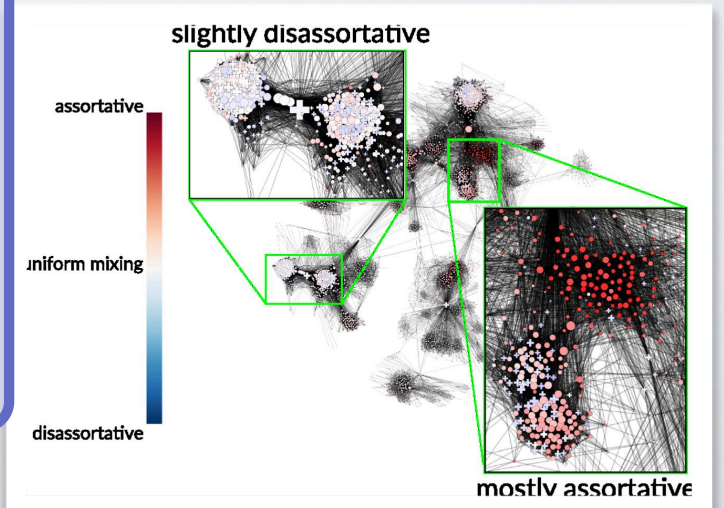


Illustration of different local assortativity behaviors leading to the same global assortativity value (bottom: distribution of local assortatvity). Figure from[a], in which the authors propose a measure of **multiscale assortativity**.

[a] Peel, Delvenne, and Lambiotte 2018.

Gender in Social networks

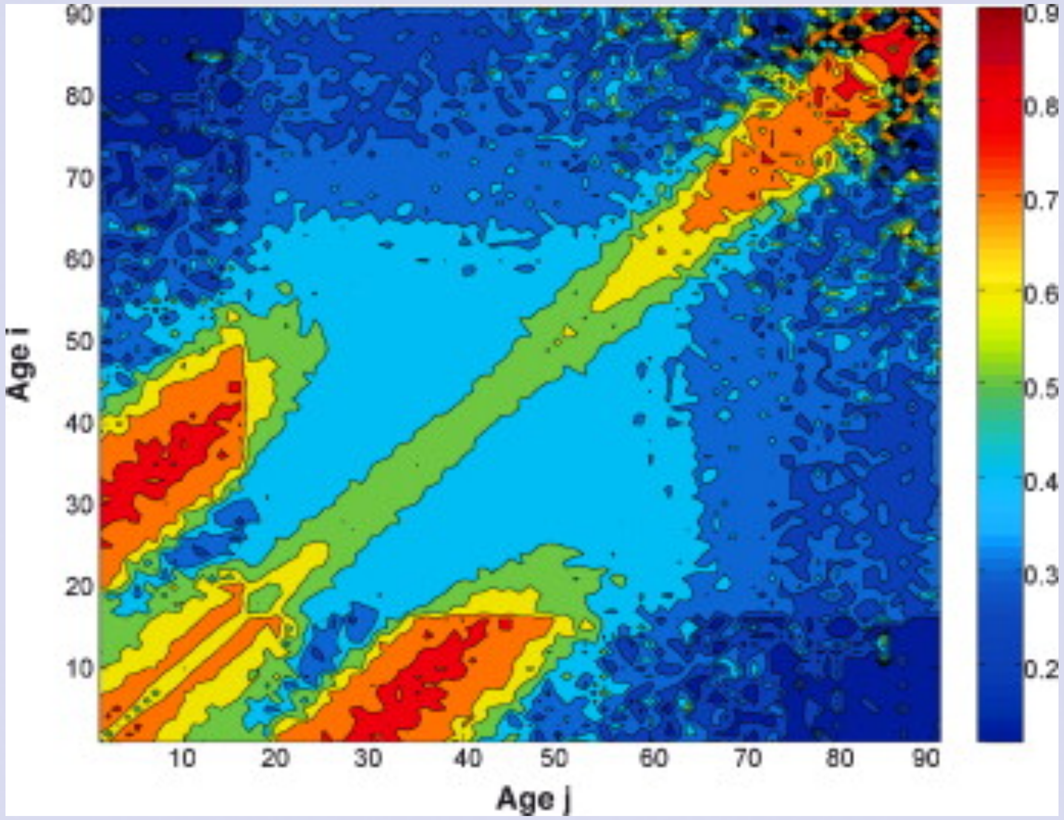


Mixing patterns

Beyond assortative and disassortative, we can study more generally **Mixing patterns**,
=>preference of nodes with attribute **a** to connect with nodes with attribute **b** (where a,b can be identical or different)

Mixing Patterns - example

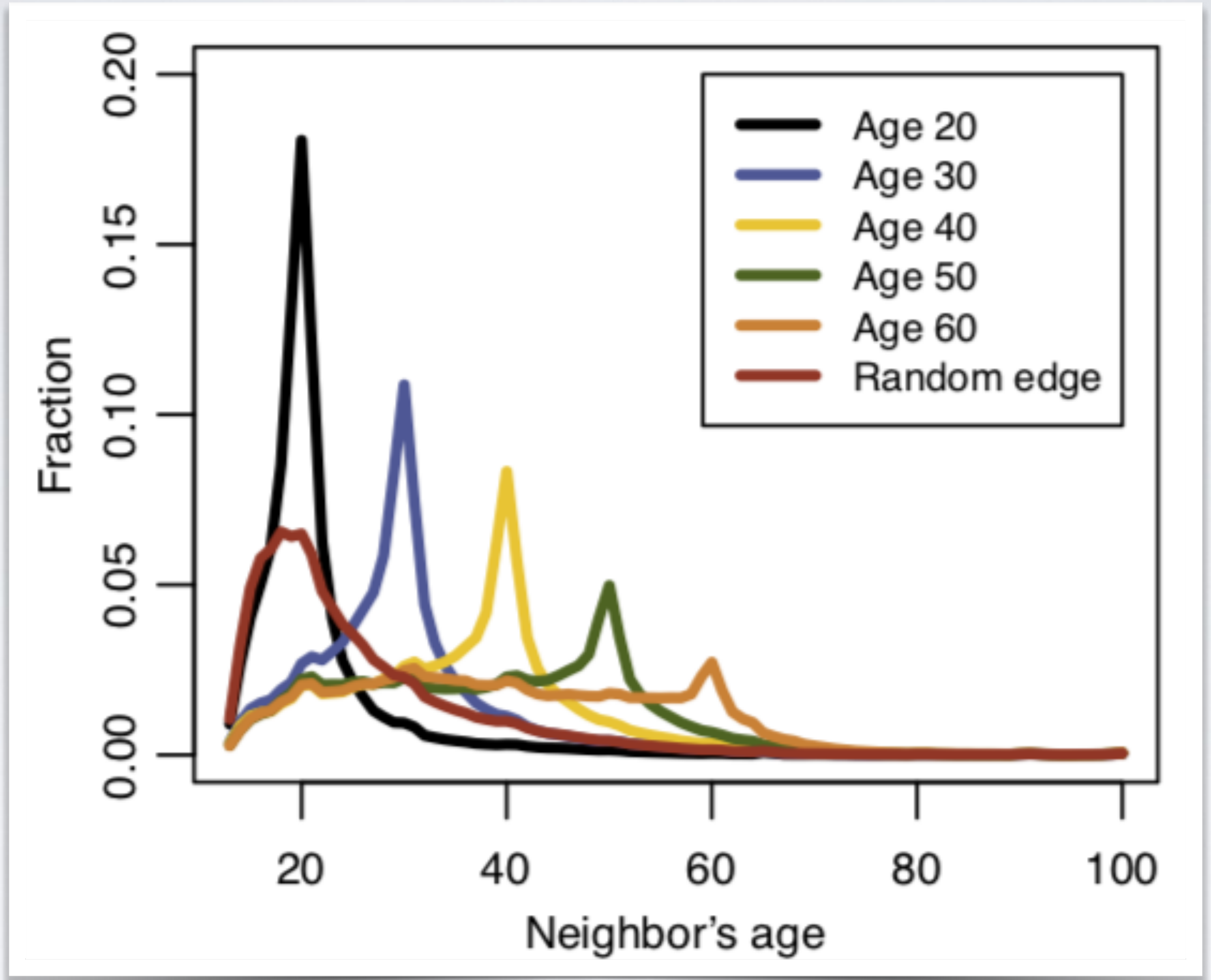
Example of mixing patterns of age in a network of interaction between individuals, reproduced from[a].



We can see that there is some level of assortativity (hig hvalues on the diagonal), but that there are also some more complex mixing patterns, for instance between age 10 and 40, approximately, here interpreted as child-parents relationships.

[a] Del Valle et al. 2007.

Mixing patterns



- [The Anatomy of the Facebook Social Graph, Ugander et al. 2011]

Degree-degree correlation

- Assortativity often used for **degree assortativity**
- An application of assortativity to the case of degrees used as node properties:
  - Are *important nodes* connected to other important nodes with a higher probability than expected?
  - The degree can be used as any other scalar property

	network	type	size n	assortativity r	error σ_r
social	physics coauthorship	undirected	52 909	0.363	0.002
	biology coauthorship	undirected	1 520 251	0.127	0.0004
	mathematics coauthorship	undirected	253 339	0.120	0.002
	film actor collaborations	undirected	449 913	0.208	0.0002
	company directors	undirected	7 673	0.276	0.004
	student relationships	undirected	573	−0.029	0.037
	email address books	directed	16 881	0.092	0.004
technological	power grid	undirected	4 941	−0.003	0.013
	Internet	undirected	10 697	−0.189	0.002
	World-Wide Web	directed	269 504	−0.067	0.0002
	software dependencies	directed	3 162	−0.016	0.020
biological	protein interactions	undirected	2 115	−0.156	0.010
	metabolic network	undirected	765	−0.240	0.007
	neural network	directed	307	−0.226	0.016
	marine food web	directed	134	−0.263	0.037
	freshwater food web	directed	92	−0.326	0.031

M. E. J. Newman, Phys. Rev. Lett. **89,** 208701 (2002)

Average nearest-neighbour degree

- More detailed characterisation of degree-degree correlations
- k_{annd}: **a**verage **n**earest **n**eighbours **d**egree
- k_{annd} can be written as: $$k_{annd}(k) = \sum_{k'} k' P(k' \mid k) = \frac{\sum_{k'} k' e_{kk'}}{\sum_{k'} e_{kk'}}$$
    - where $P(k'|k)$ is the conditional probability that an edge of a node with degree k points to a node with degree k'
- If there are no degree correlations:

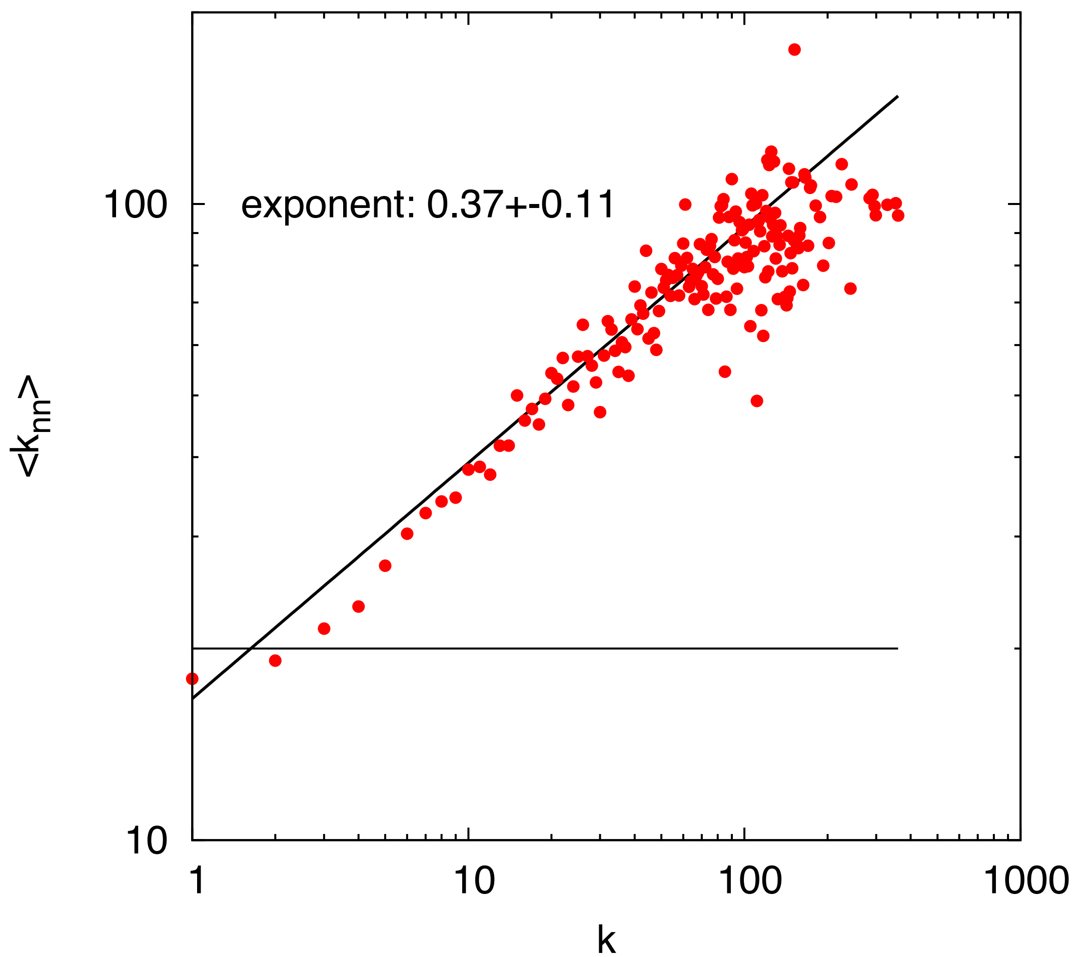
$$k_{annd}(k) = \ldots = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

  - k_{annd} is independent of k (nodes of any degrees should have the same nearest neighbors degree)

- If the network is assortative $k_{nn}(k)$ is a positive function
- If the network is disassortative $k_{nn}(k)$ is a negative function

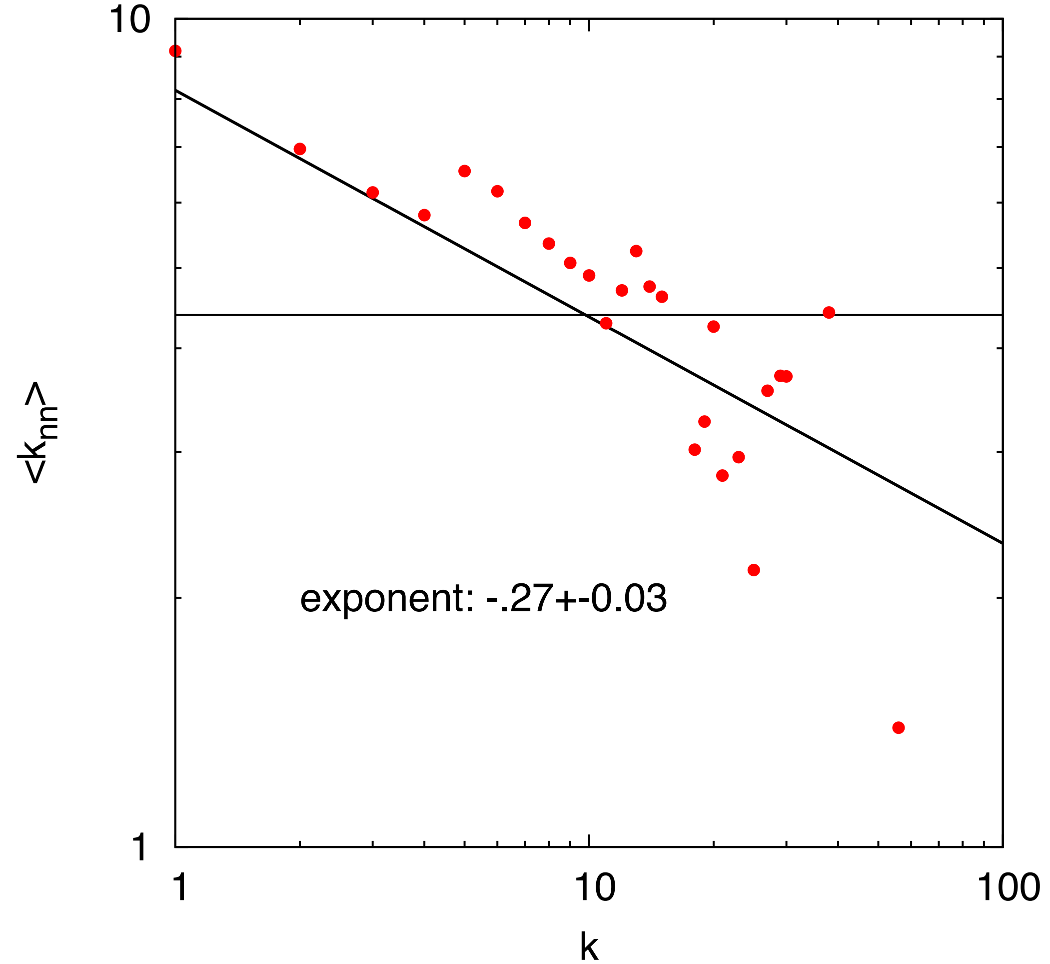
R. Pastor-Satorras, A. Vázquez, A. Vespignani, Phys. Rev. E 65, 066130 (2001)

Nearest neighbour degree



Astrophysics co-authorship network

Assortative

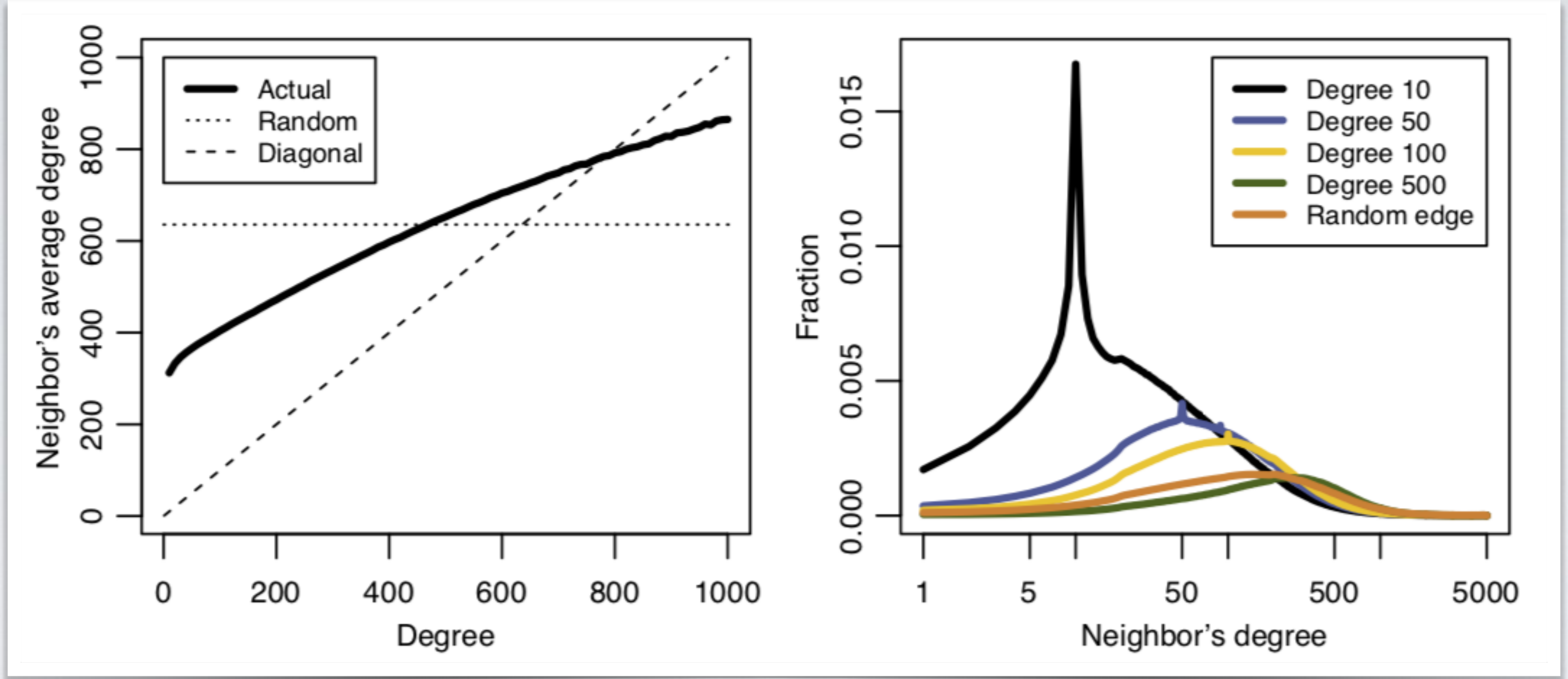


Yeast PPI

Disassortative

Nearest neighbour degree

On Facebook



- [The Anatomy of the Facebook Social Graph, Ugander et al. 2011]

Rich-club coefficient

- How well connected are the well connected among themselves
- It is calculated on a list of node degree sorted in ascendant order as

$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k}-1)}$$

- $N_{>k}$ denotes the number of nodes with degree k or larger than k
- $E_{>k}$ measures the number of links between them
- Results are usually compared to random references
  - configuration model of equivalent synthetic network
  - configuration model of the empirical network

Algorithm

- rank nodes by degree
- remove nodes in an ascendant degree order
- measure the density of the remaining network