Social Network Analysis

#12 Homophily

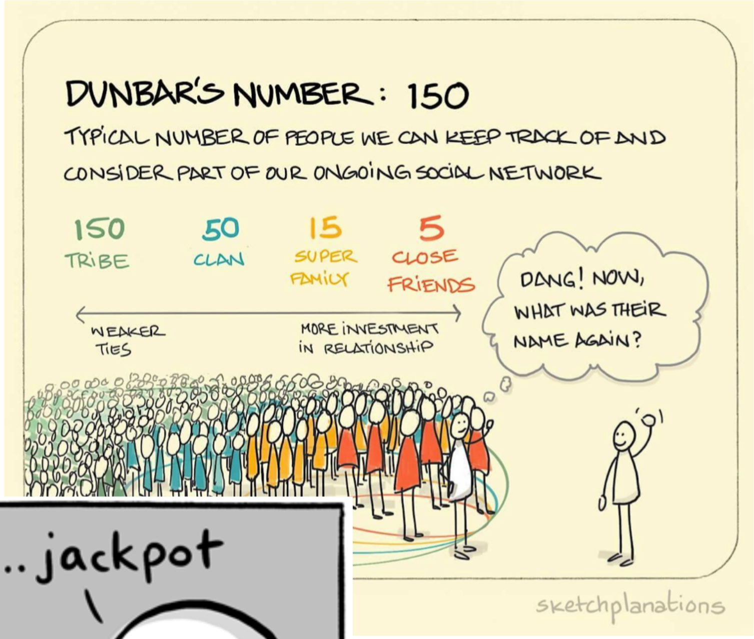
Homophily

Humans and social media

We have access to an unlimited amount of information, but we follow a limited number of sources

Because we are…

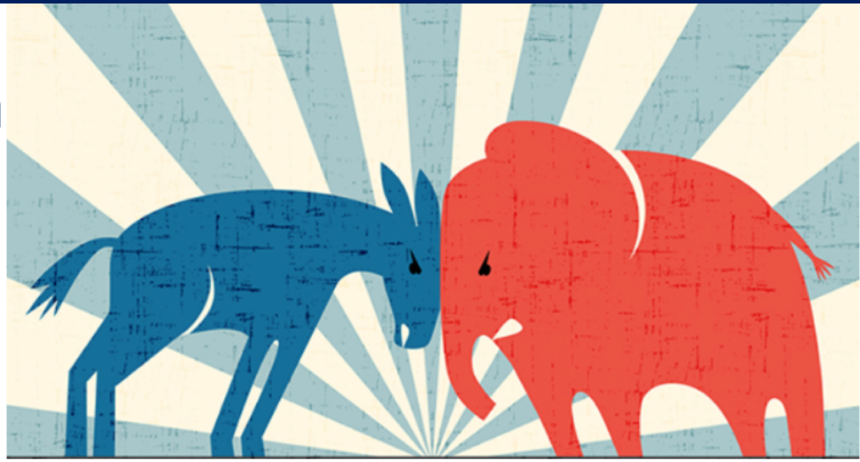
Bounded



Biased

Effects on online behaviour

Polarization



Homophily

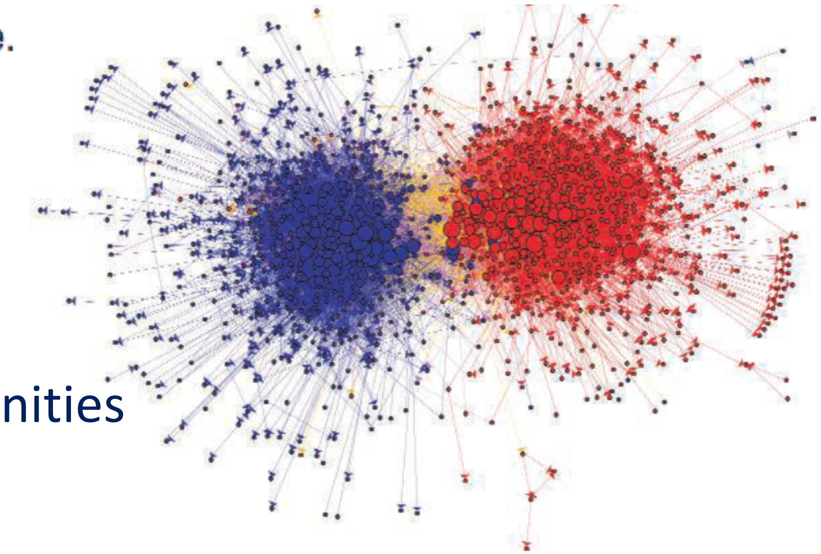
Selective exposure

Homophily

Homophily

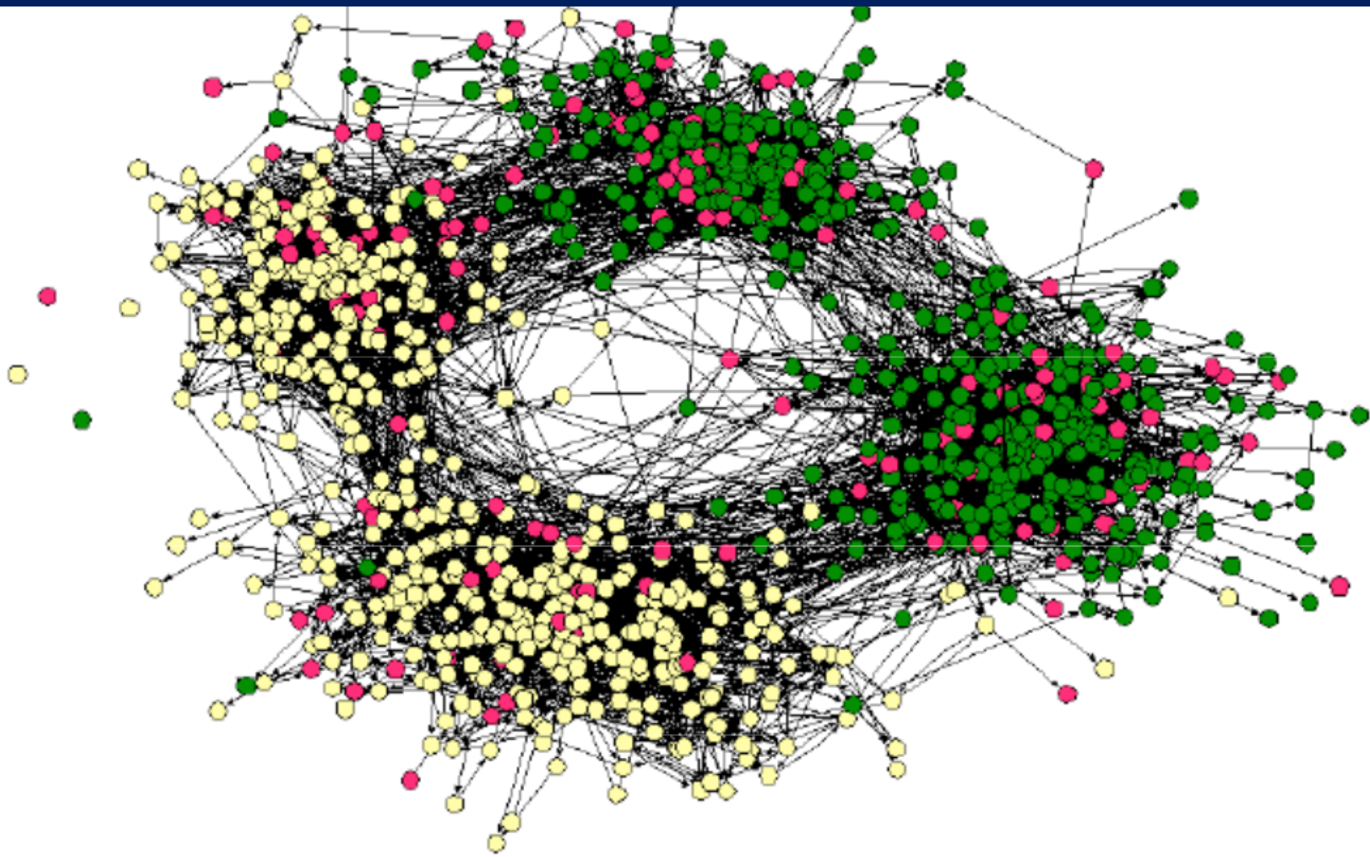
From Wikipedia, the free encyclopedia

Homophily (from Ancient Greek: *homoû*, 'together' + *philíē*, 'friendship, love') is the tendency of individuals to associate and bond with similar others, as in the proverb "birds of a feather flock together."[1] The presence of homophily has been discovered in a vast array of network studies: over 100 studies have observed homophily in some form or another, and they establish that similarity is associated with connection.[2] The categories on which homophily occurs include age, gender, class, and organizational role.



Political blog communities

Homophily at action: racial segregation



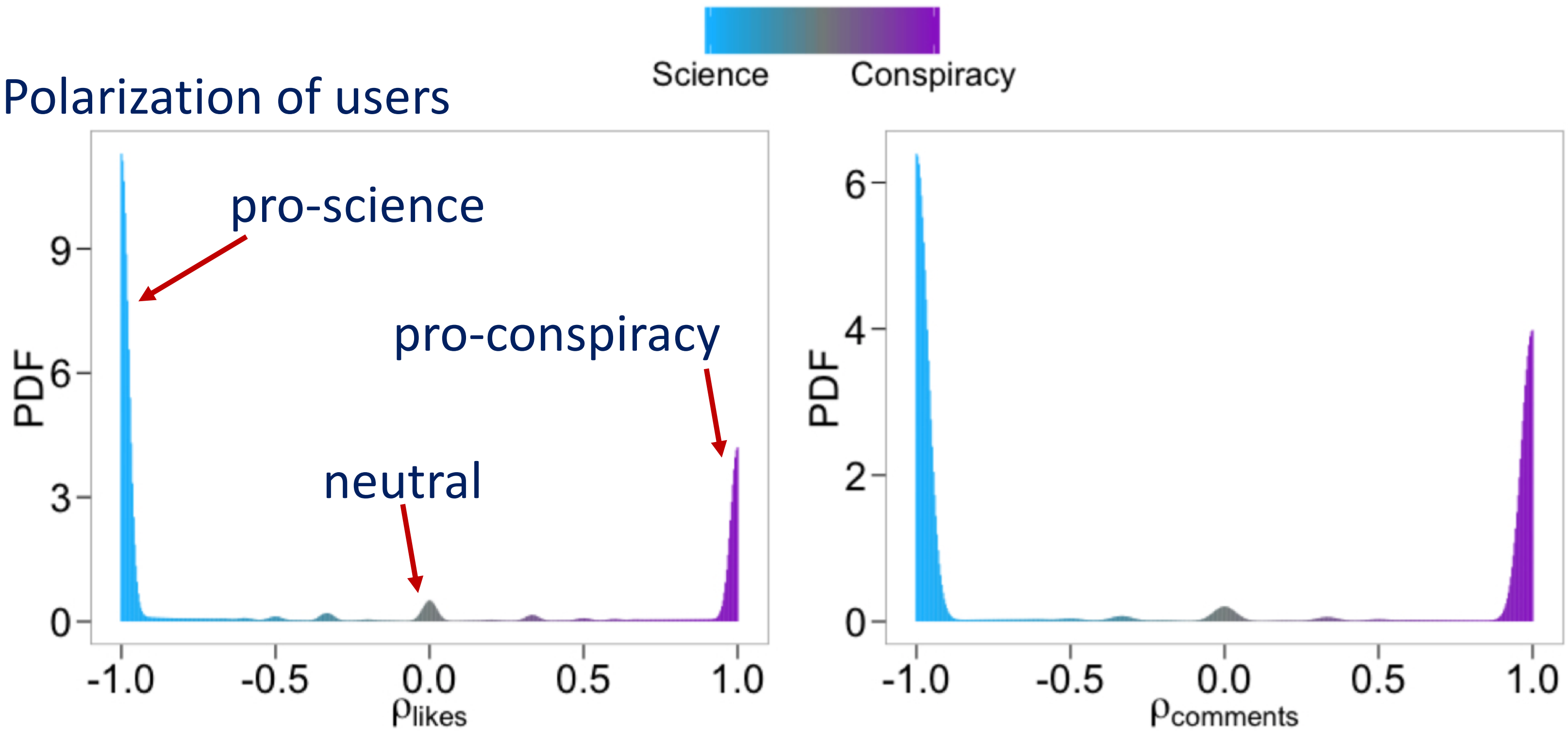
(Easley and Kleinberg, 2010)

Figure 4.1: Homophily can produce a division of a social network into densely-connected, homogeneous parts that are weakly connected to each other. In this social network from a town's middle school and high school, two such divisions in the network are apparent: one based on race (with students of different races drawn as differently colored circles), and the other based on friendships in the middle and high schools respectively [304].

Polarization

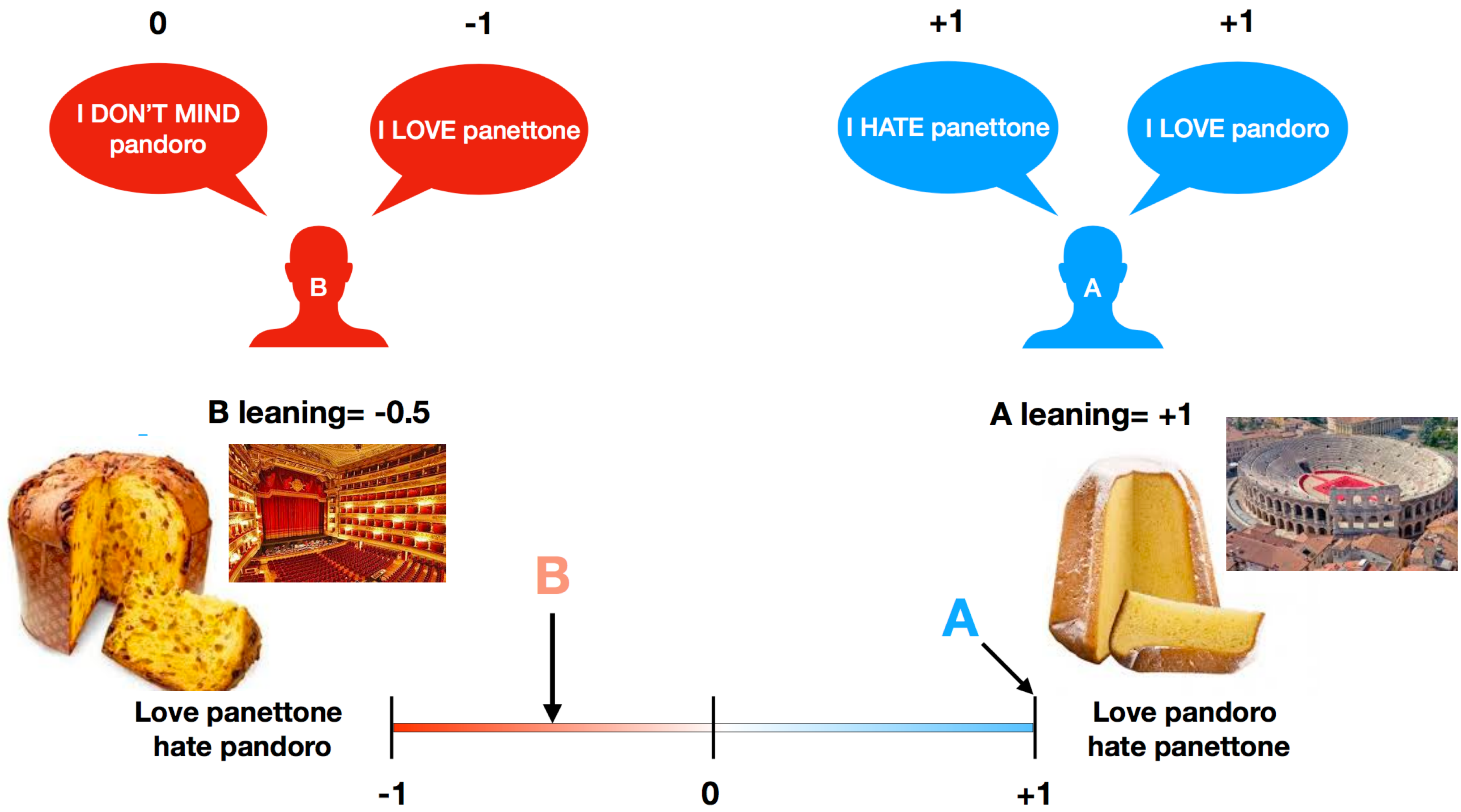
The extreme segregation of users into homogeneous communities based on their opinion on a controversial topic

Polarization of users



MIME

Users' leaning

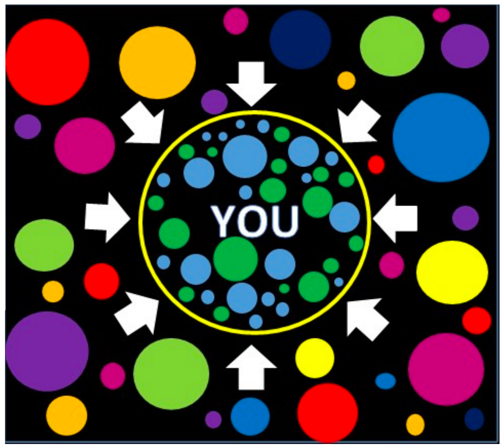


Eco chambers

Echo chamber (media)

From Wikipedia, the free encyclopedia

In news media, an **echo chamber** is a metaphorical description of a situation in which beliefs are amplified or reinforced by communication and repetition inside a closed system and insulates them from rebuttal.[1] By visiting an "echo chamber", people are able to seek out information that reinforces their existing views, potentially as an unconscious exercise of confirmation bias. This may increase social and political polarization and extremism.[2] The term is a metaphor based on the acoustic echo chamber, where sounds reverberate in a hollow enclosure. Another emerging term for this echoing and homogenizing effect on the Internet within social communities, such as Facebook, Instagram, Twitter, Reddit, etc; is cultural tribalism.[3]

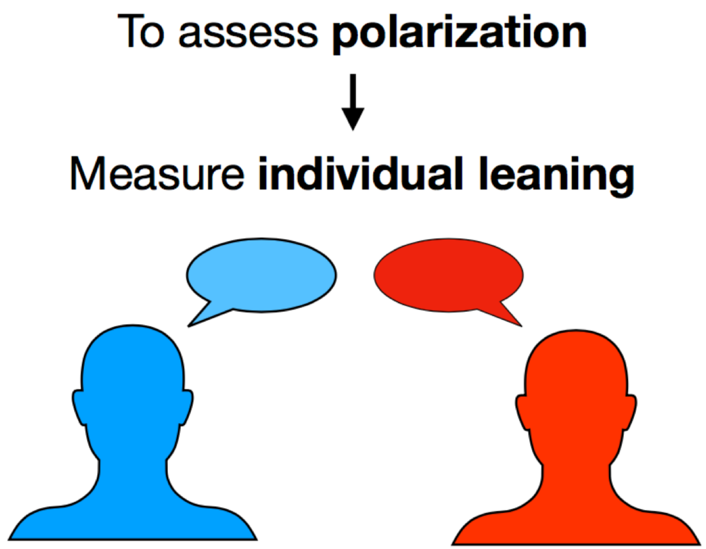


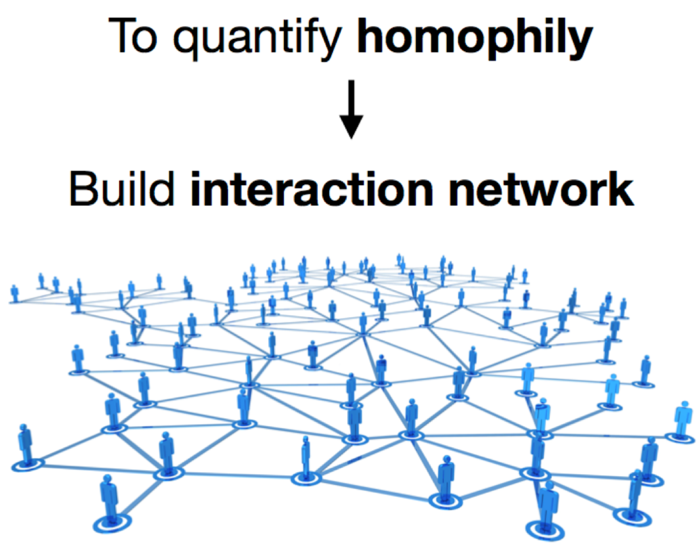
Definition of echo-chamber

Cinelli, Morales, Galeazzi, Quattrociocchi, Starnini (2020)
Echo chambers on social media: A comparative analysis
https://arxiv.org/pdf/2004.09603.pdf

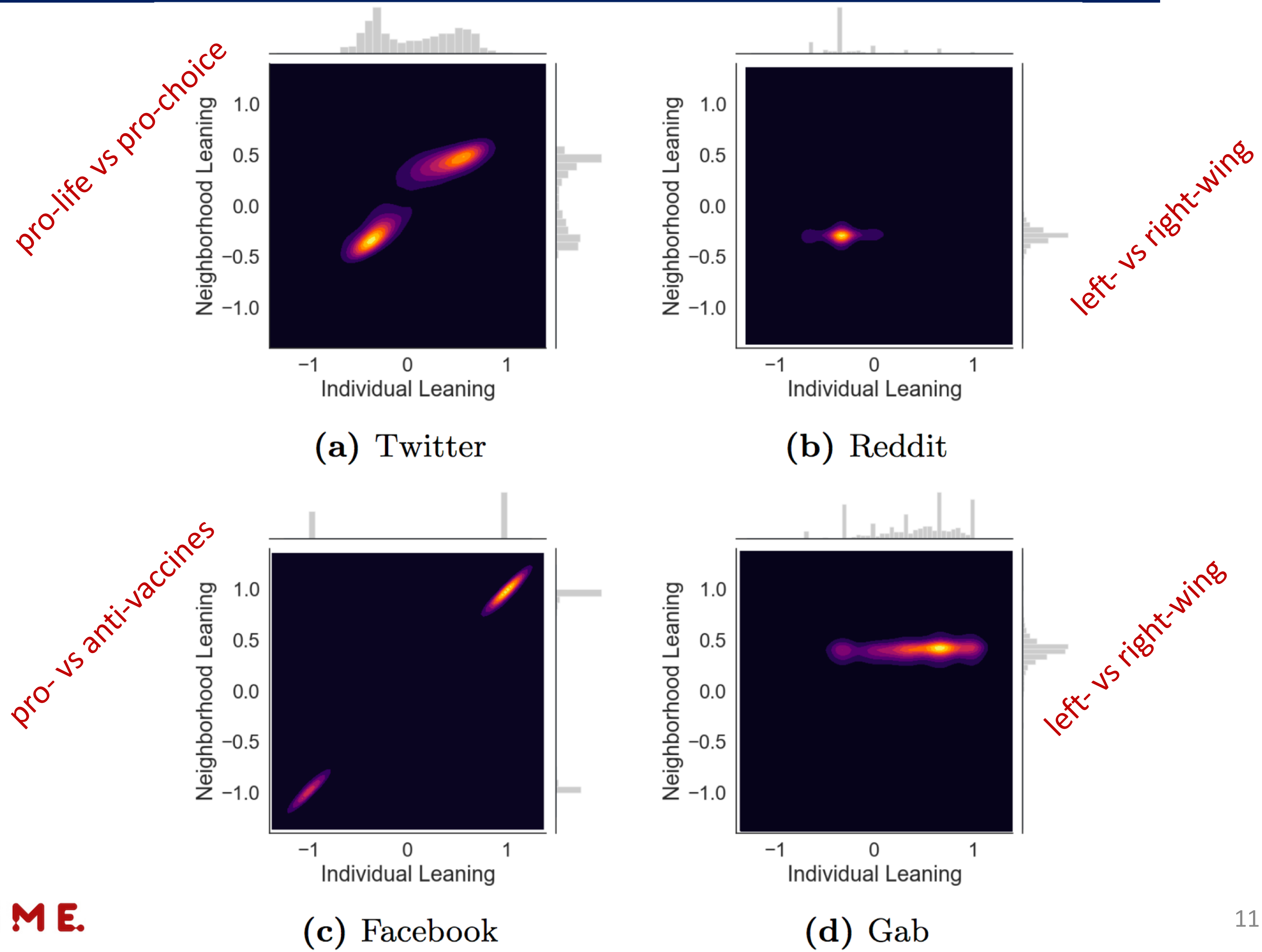
Coexistence of

- opinion polarization with respect to a controversial topic
- homophily in interactions





Echo-chamber effect in social networks



Filter bubbles

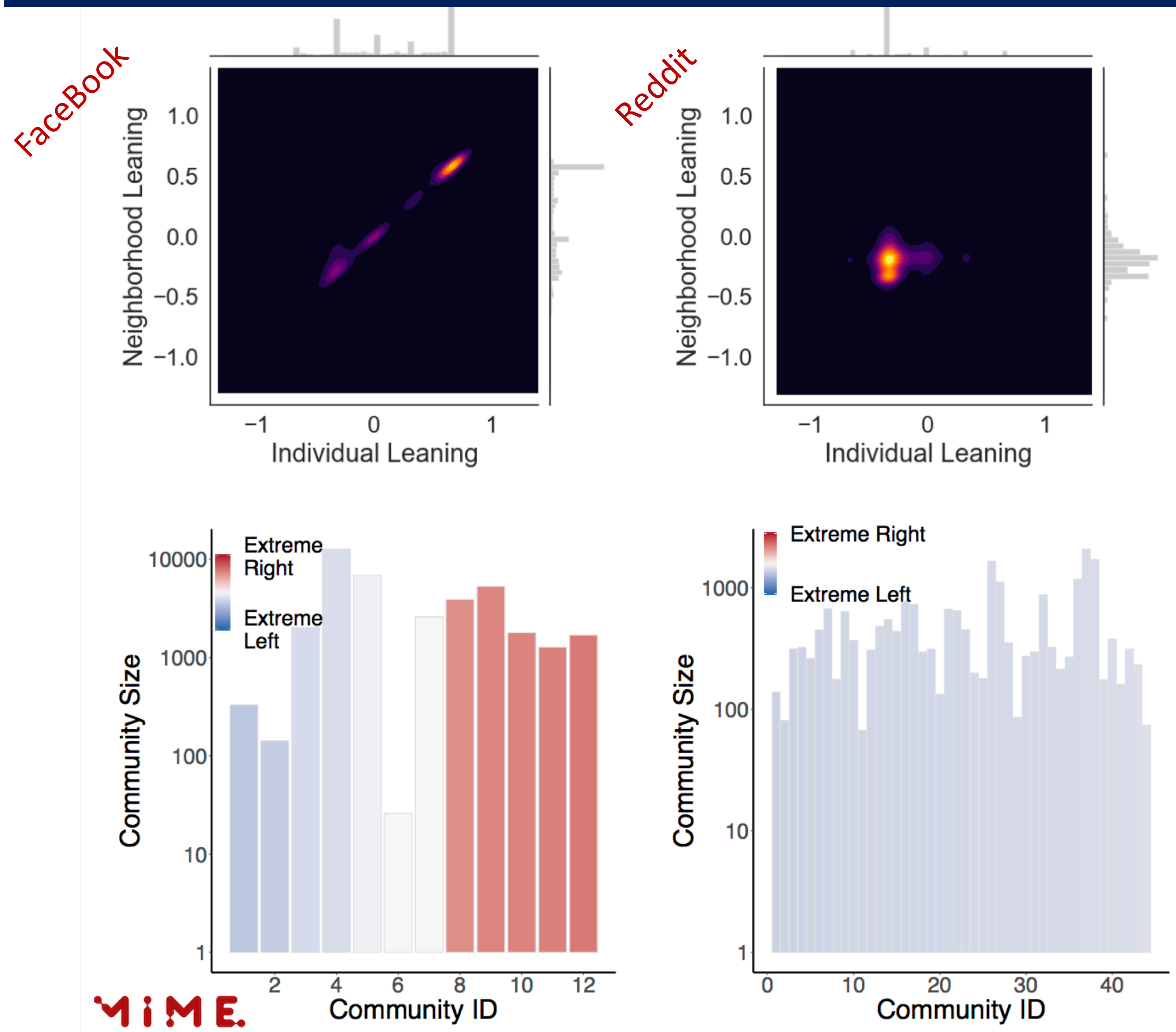
Filter bubble

From Wikipedia, the free encyclopedia

A **filter bubble** – a term coined by internet activist Eli Pariser – is a state of intellectual isolation[1] that allegedly can result from personalized searches when a website algorithm selectively guesses what information a user would like to see based on information about the user, such as location, past click-behavior and search history.[2][3][4] As a result, users become separated from information that disagrees with their viewpoints, effectively isolating them in their own cultural or ideological bubbles.[5] The choices made by these algorithms are not transparent.[6]

The term was coined by internet activist Eli Pariser circa 2010

Political leaning



- ❑ Same Topic: News
- ❑ Same leaning assigned to news sources
- ❑ Different platforms: Facebook has a strong social feeding algorithm, Reddit has not
- ❑ Different characteristics: Facebook shows segregation among groups with different leaning, Reddit has one group

Polarization in pro-life/pro-choice networks

Lejla Dzanko, Giulia Rizzoli, Sanja Milijanovic, Sara Shena, Lara Malin Schwarz
IP3 2019/20

Background

Abortion is one of the most controversial topics in social public, political and scientific debates in different disciplines

Often debates result in reforms of the law → USA 2019

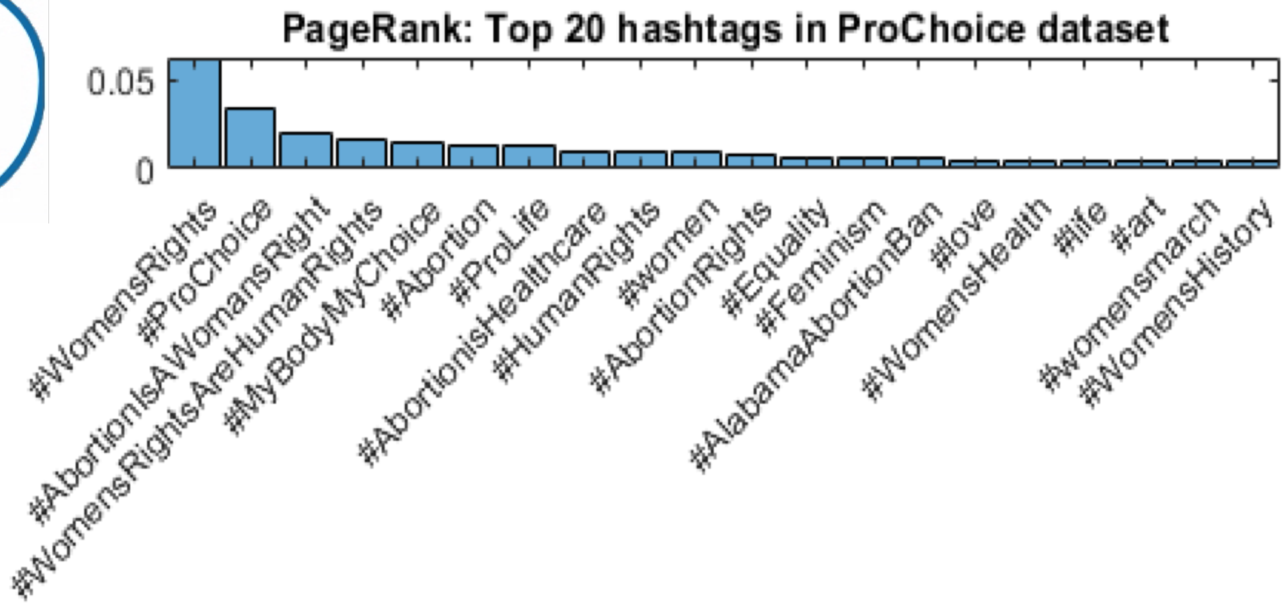
Two movements:

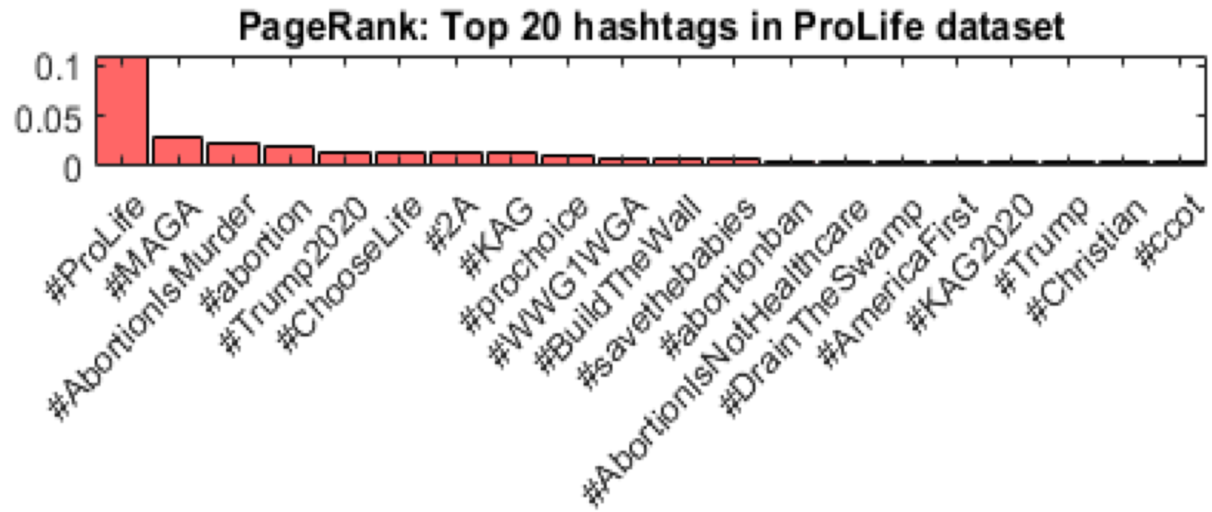
- Pro-Life: every human (embryo) has the right to live; abortion is murder → goal to ban it
- Pro-Choice: every woman should have the right to decide what to do with her body on her own → goal to keep abortion safe and legal

Data collection

Pro-Choice	Pro-Life
#prochoice	#prolife
#mybodymychoice	#savethebabies
#abortionishealthcare	#babiesarehuman
#abortionisawomansright	#chooselife
#abortionrights	#abortionban
#abortionismurder	#abortionismurder
#abortionsupportnetwork	#lovethemboth
#proabortion	#whywemarch

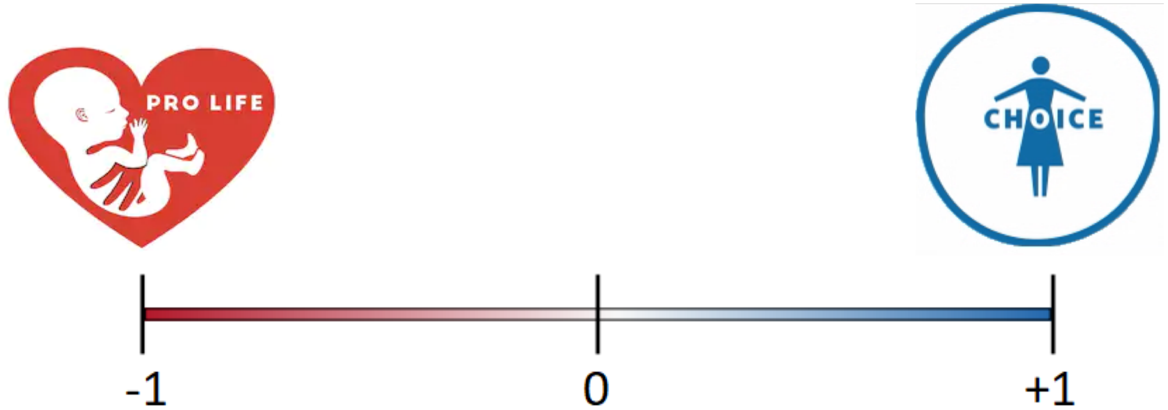
PageRank centrality





MIME

Hashtag polarization

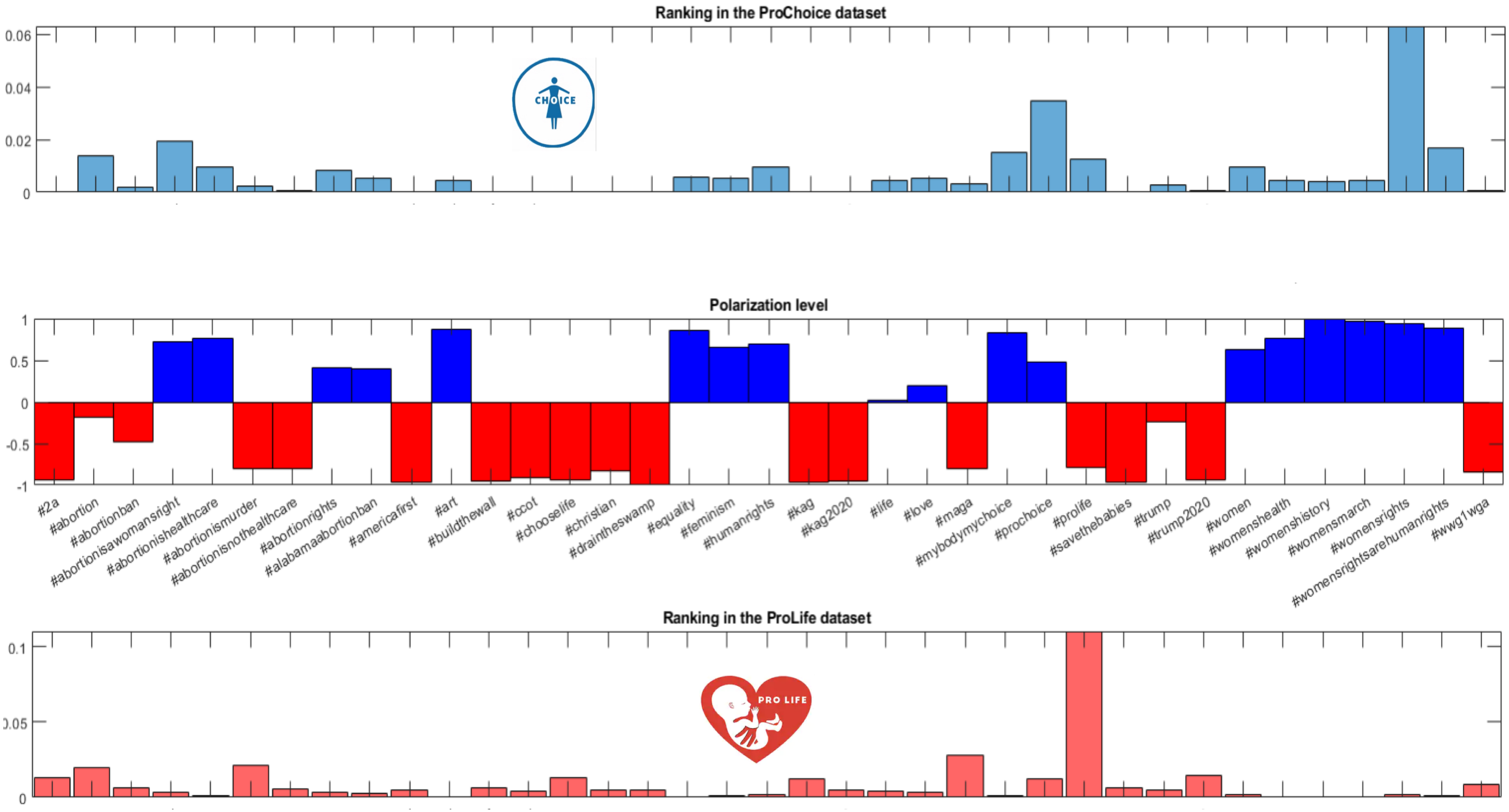


- Measure of hashtags centralities among the two dataset
- Extract which opinion an hashtag holds

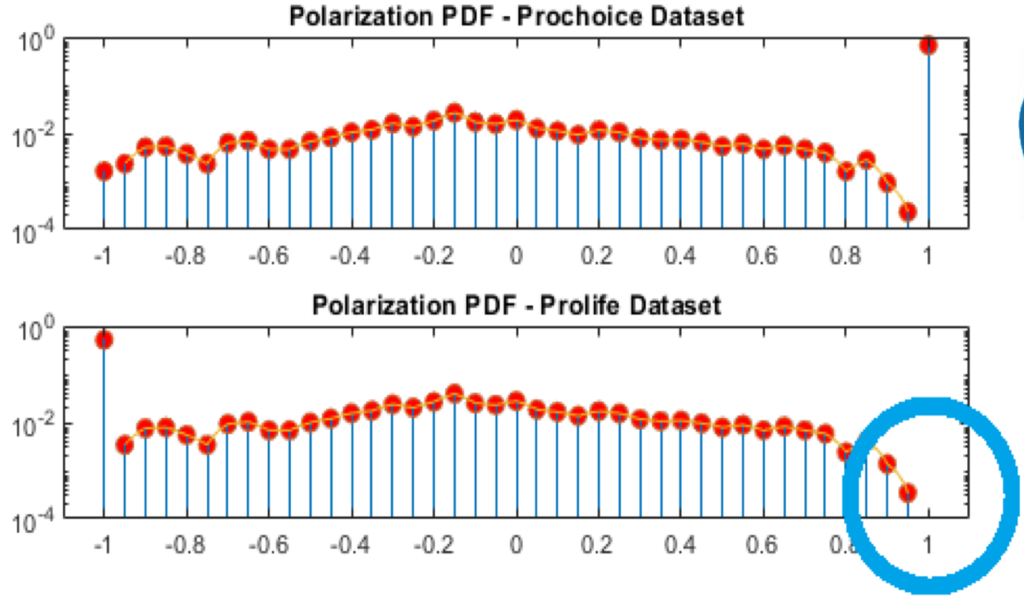
$$P_i = \frac{W_{pc_i} - W_{pl_i}}{W_{pc_i} + W_{pl_i}}$$

Prestige mapping

Hashtag polarization



Polarization effects



Absence of a debate?

Assortativity (degree homophily)

A.L. Barabási, Network science, http://barabasi.com/networksciencebook

Ch.7 “Degree correlation”

Correlation between hubs

- In some networks, hubs frequently connect with other hubs

  e.g., celebrity dating, actor networks

- In other cases hubs avoid connections with other hubs

  e.g., methabolic graphs, food webs (predators tend to differentiate their diet)

Assortativity

- Assortative network: high degree nodes connect with each other avoiding low degree nodes (tend to cliques)
- Disassortative network: opposite trend, hubs tend to avoid each other
- Neutral network: one with random wiring, i.e., aside from the (marginal) degree distribution of nodes, there is no correlation

MIME.

Assortativity

(dis)assortativity in sociology quantifies homophily in social networks, e.g., effects like:

- Rich people tend to be friends with each other
- People with the same education tend to hang out together

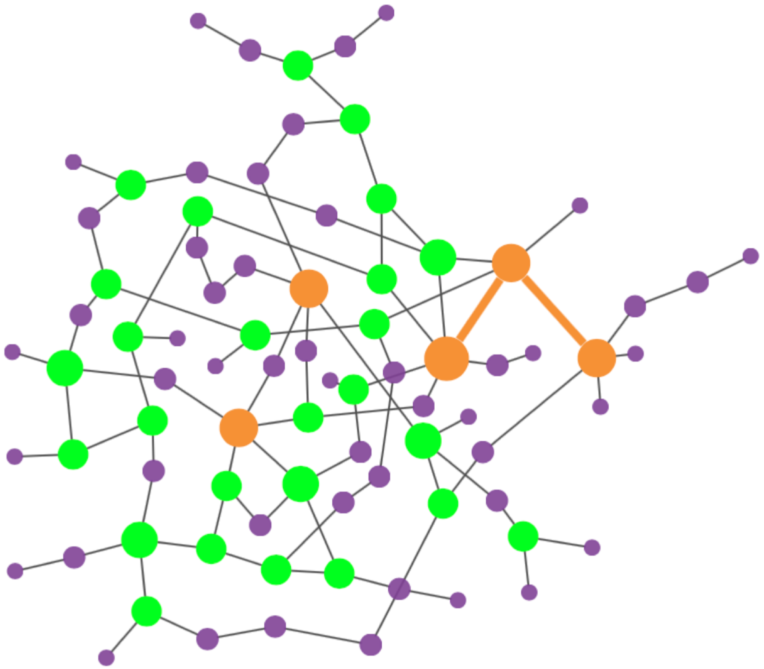
i.e., we expect social networks to be assortative

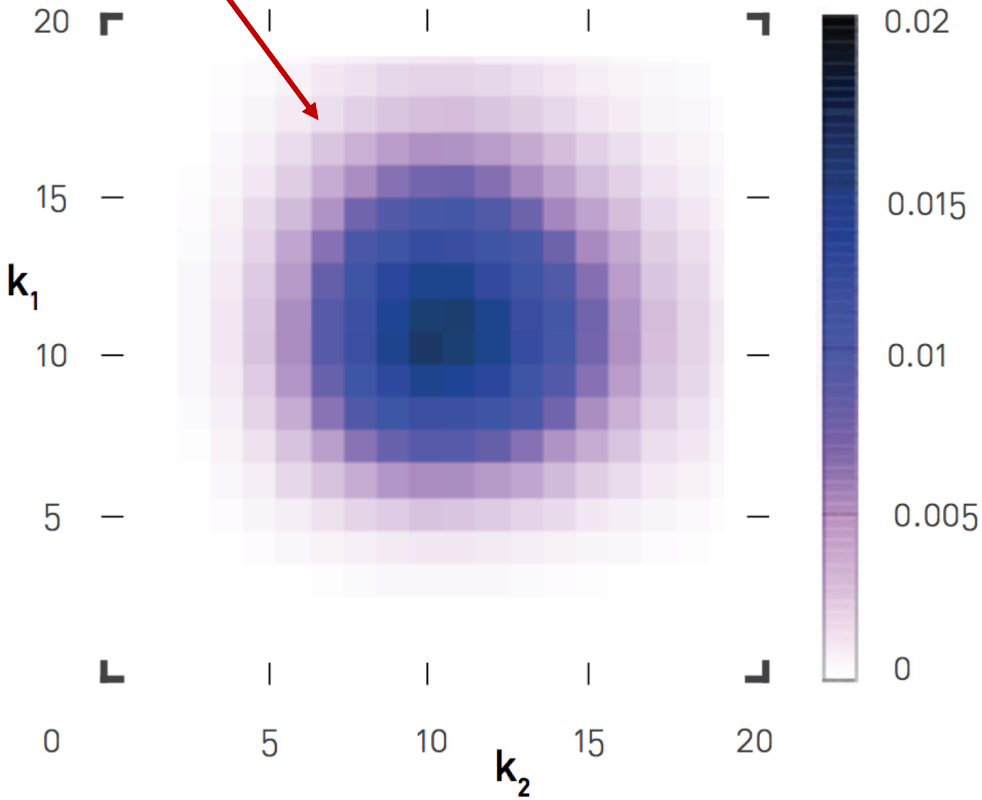
Neutral networks

The degree correlation matrix $E_{k1,k2}$ is visually centred around the average degree

In the neutral case we expect $E_{k1,k2} = q_{k1}\, q_{k2}$, i.e., independence

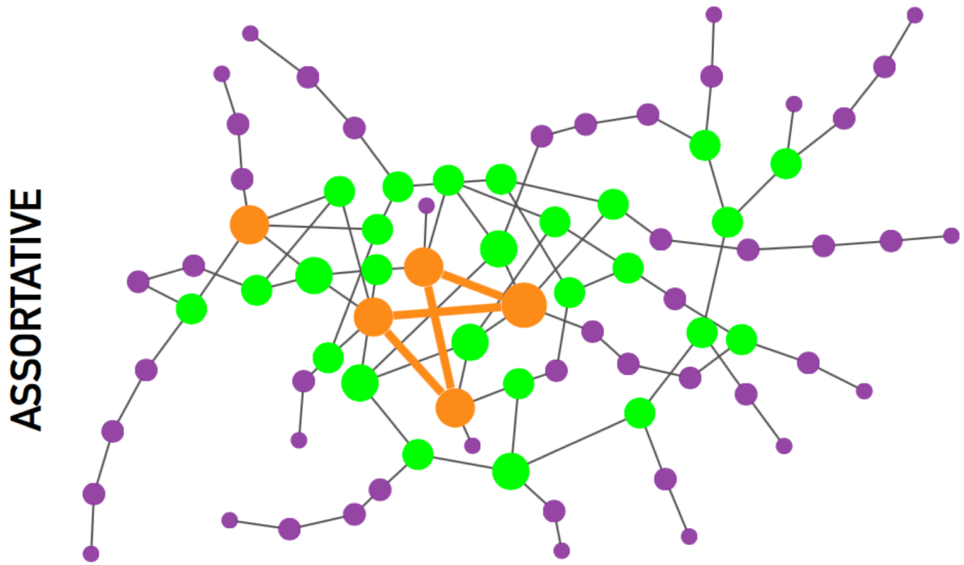
NEUTRAL

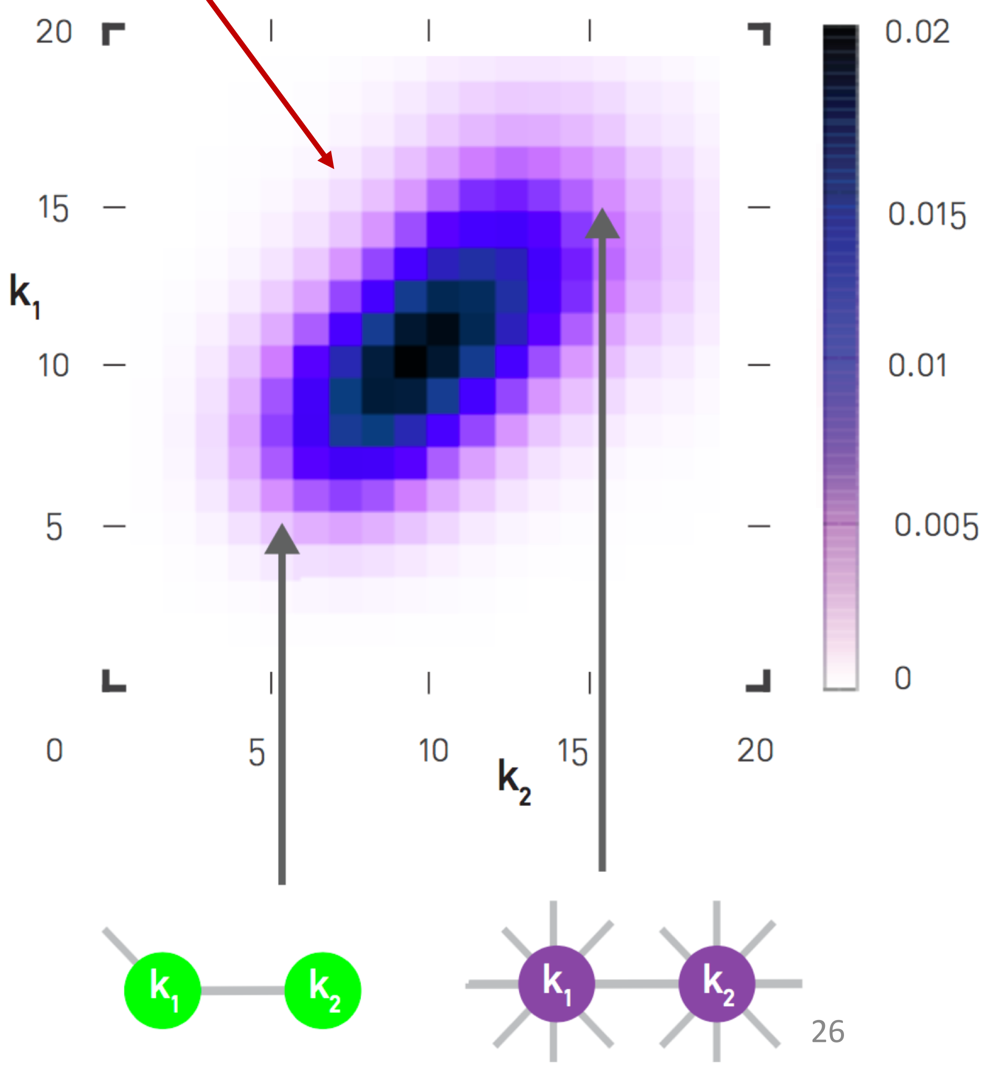




Assortative networks

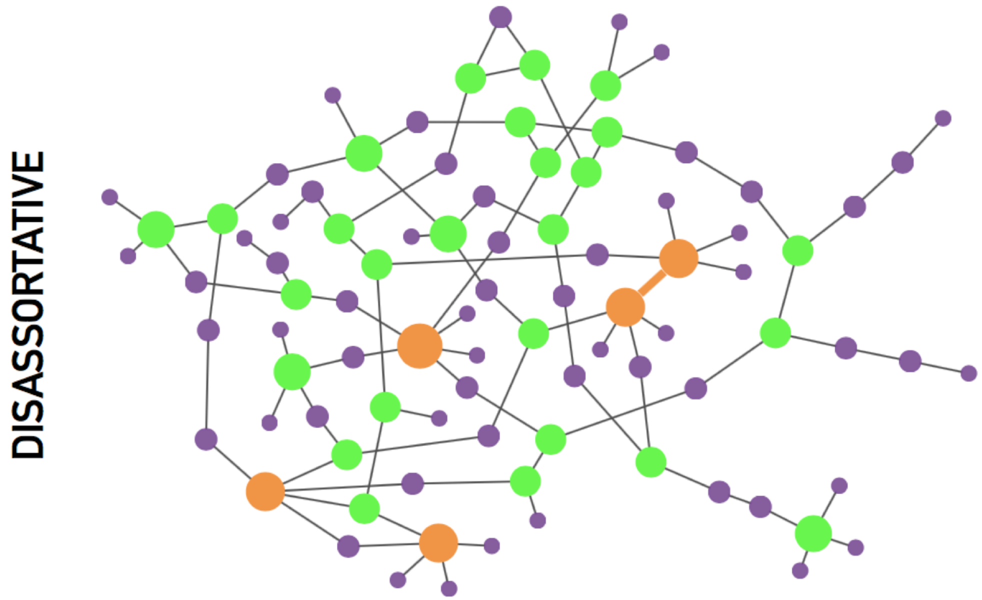
The degree correlation matrix $E_{k1,k2}$ is turning to the right

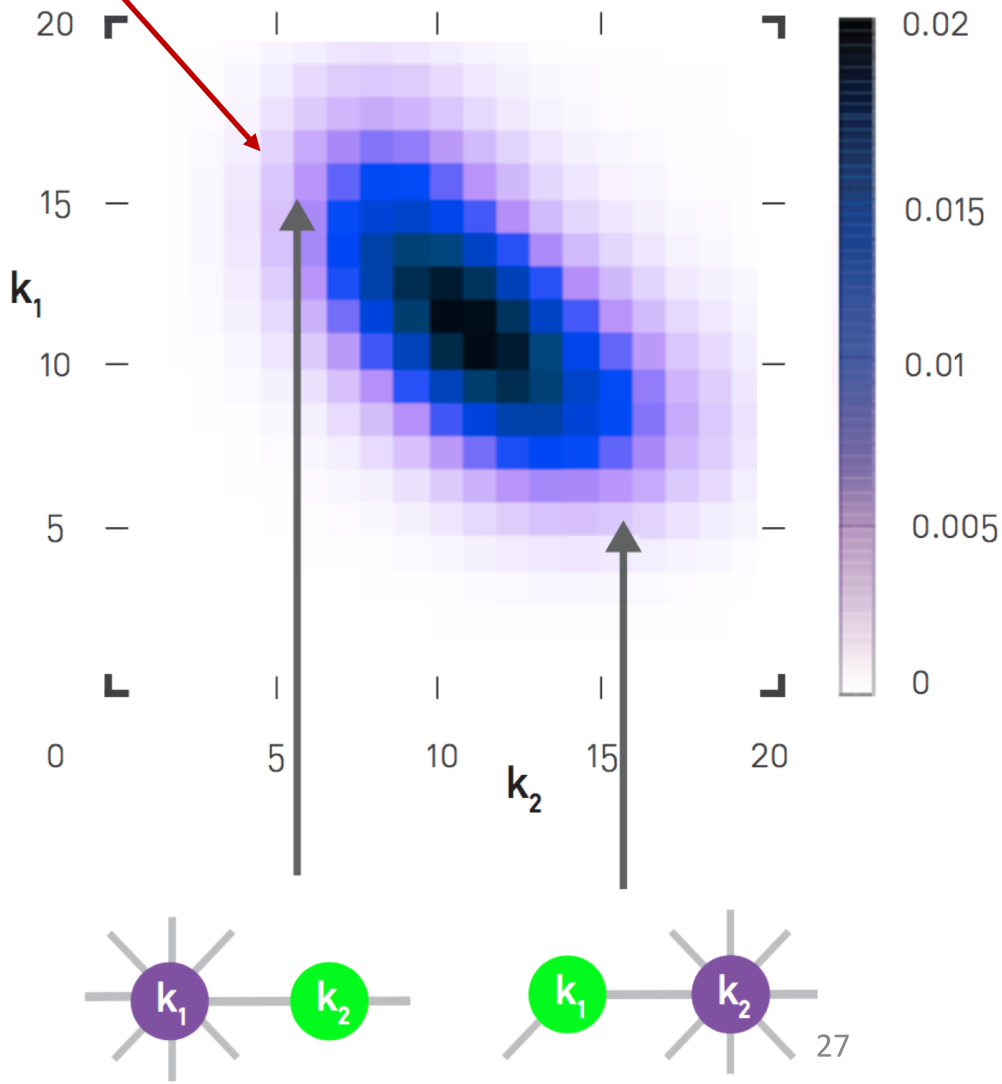




Disassortative networks

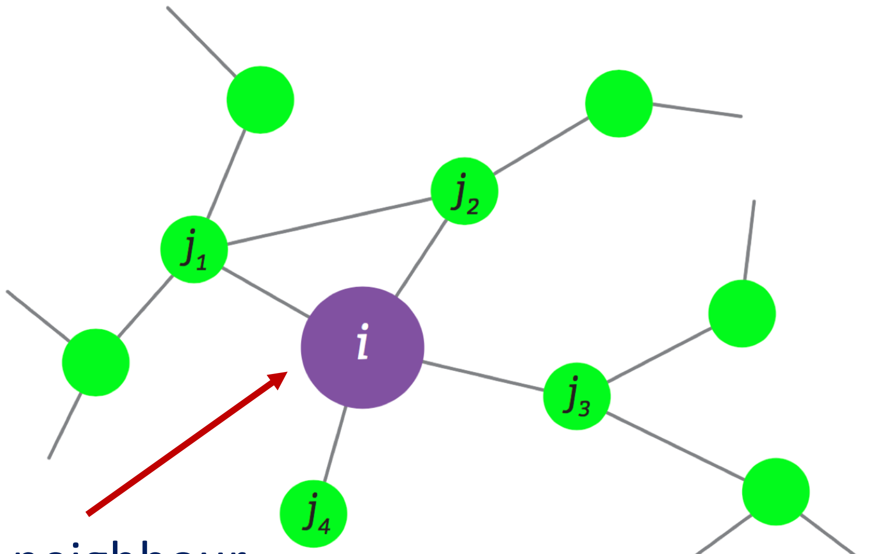
The degree correlation matrix
$E_{k1,k2}$ is turning to the left





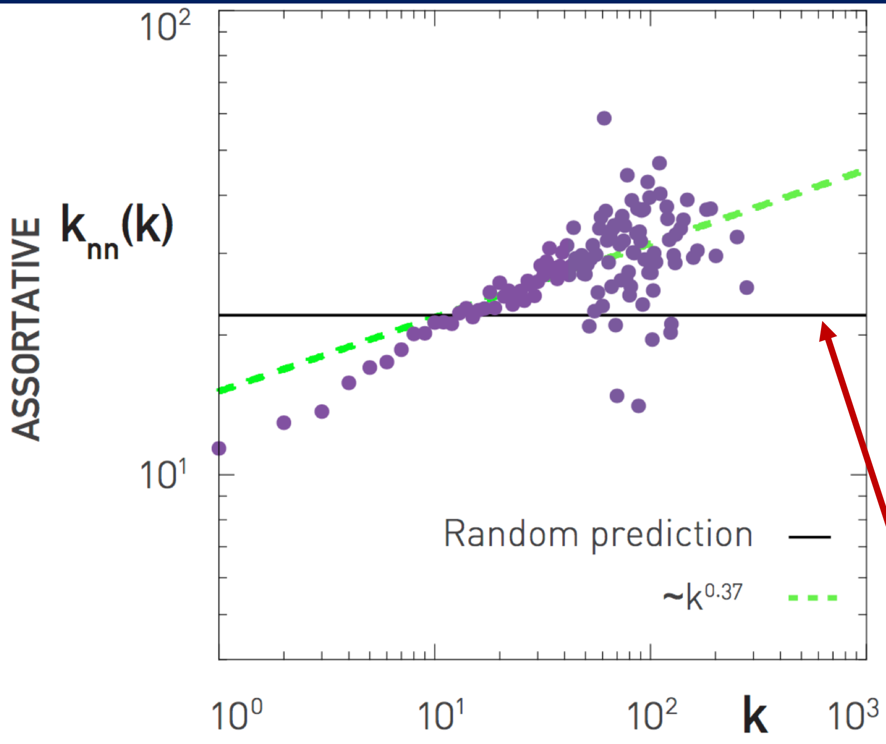
Nearest neighbour degree

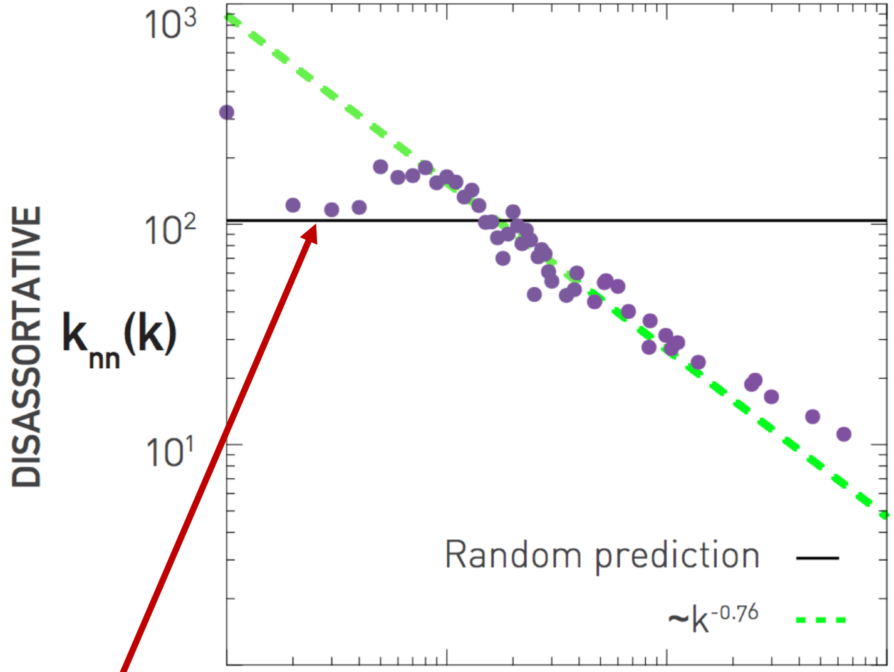
- **Idea** : inspect the degrees of the **neighbouring** nodes (easier than matrices)



average neighbour degree of node *i* is ¼ (4 + 3 + 1 + 3) = 2.75

Examples



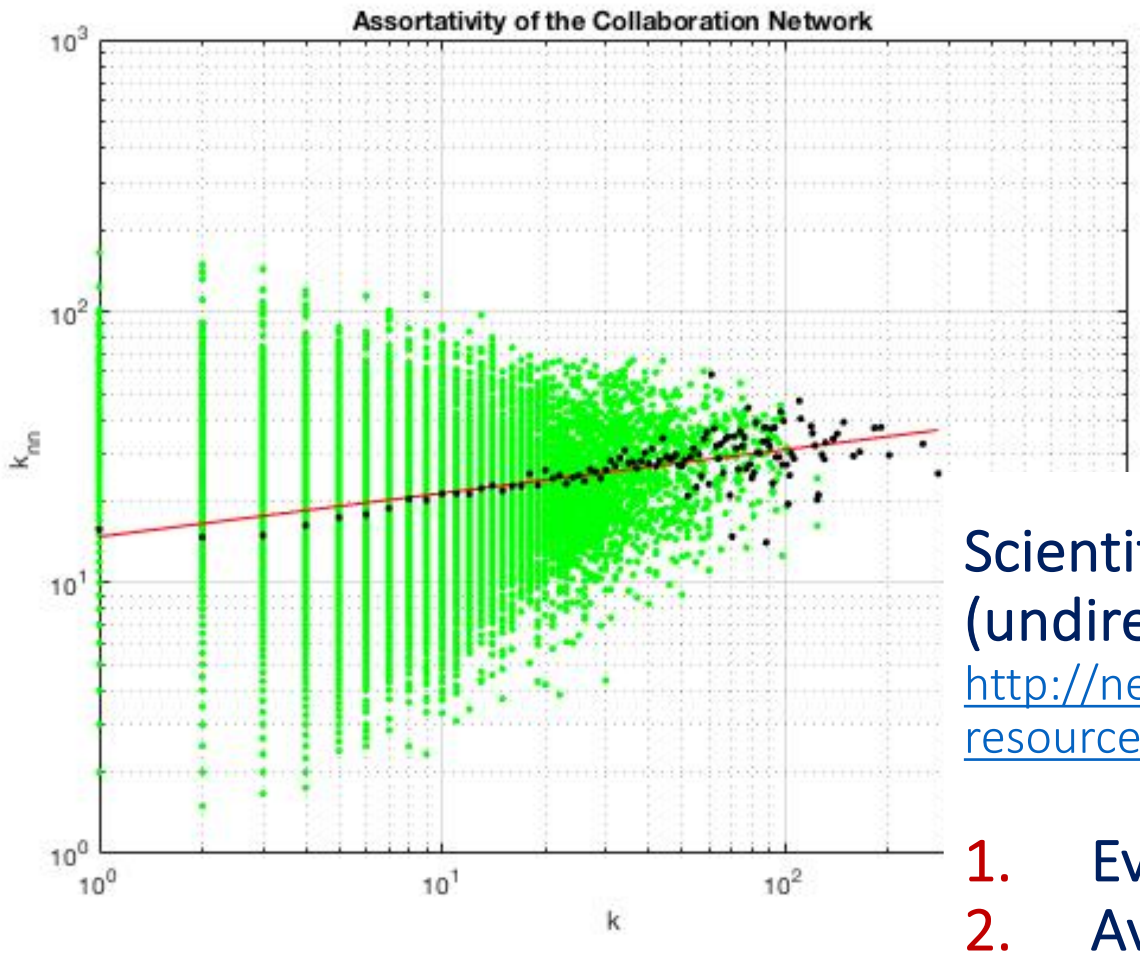


constant = independent of the degree (i.e., random = neutral)

$$\ln (k_{nn}) = \mu \ln(k_i) \quad \rightarrow \quad \mu > 0 = \text{assortative}$$

$$\mu < 0 = \text{disassortative}$$

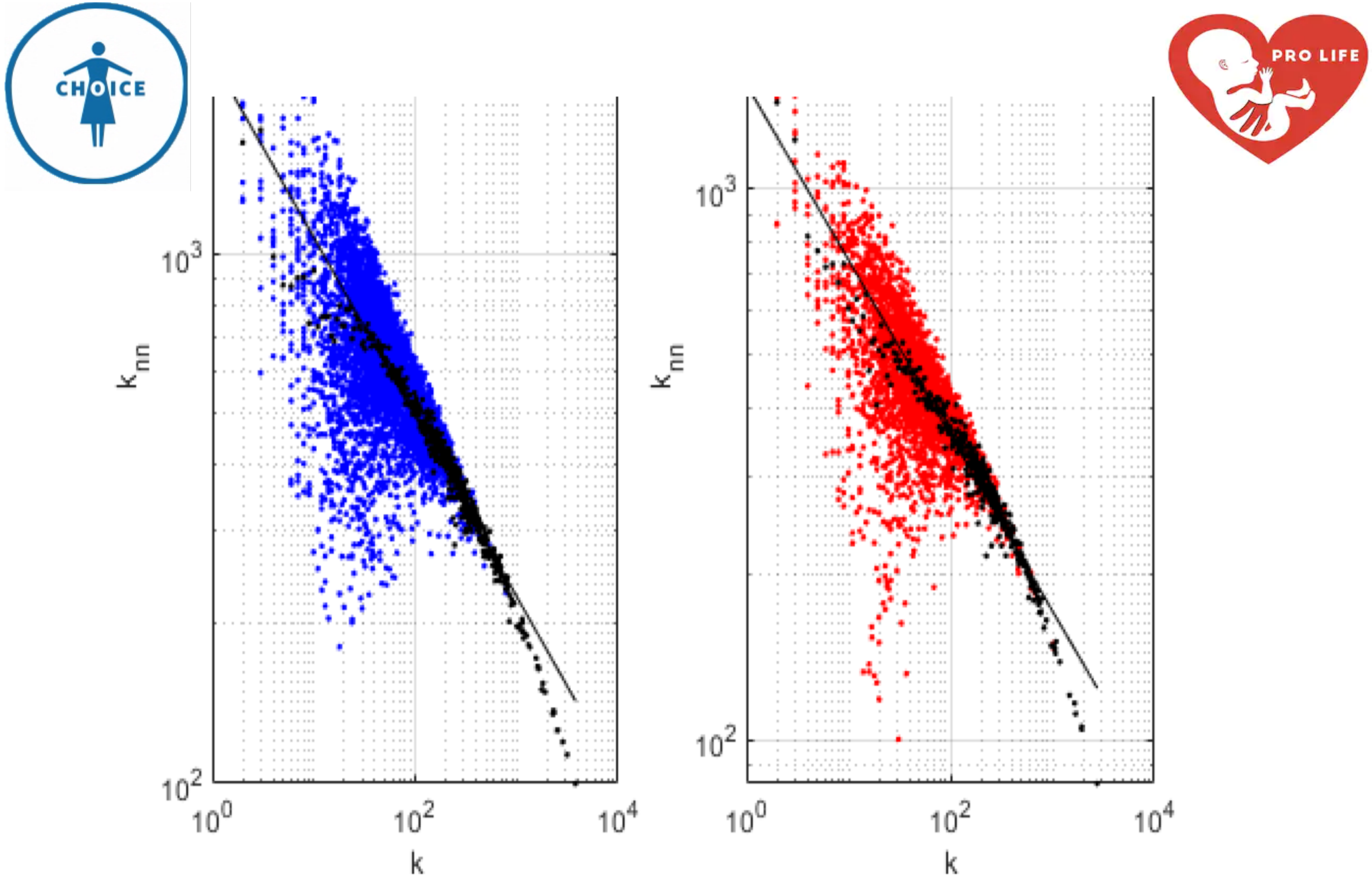
Scientific collaboration network



Scientific collaboration network (undirected, assortative)
http://networksciencebook.com/translations/en/resources/data.html

1. Evaluate average neigh. deg. k_{nn}
2. Average w.r.t. k
3. Extract the assortativity value μ=0.16

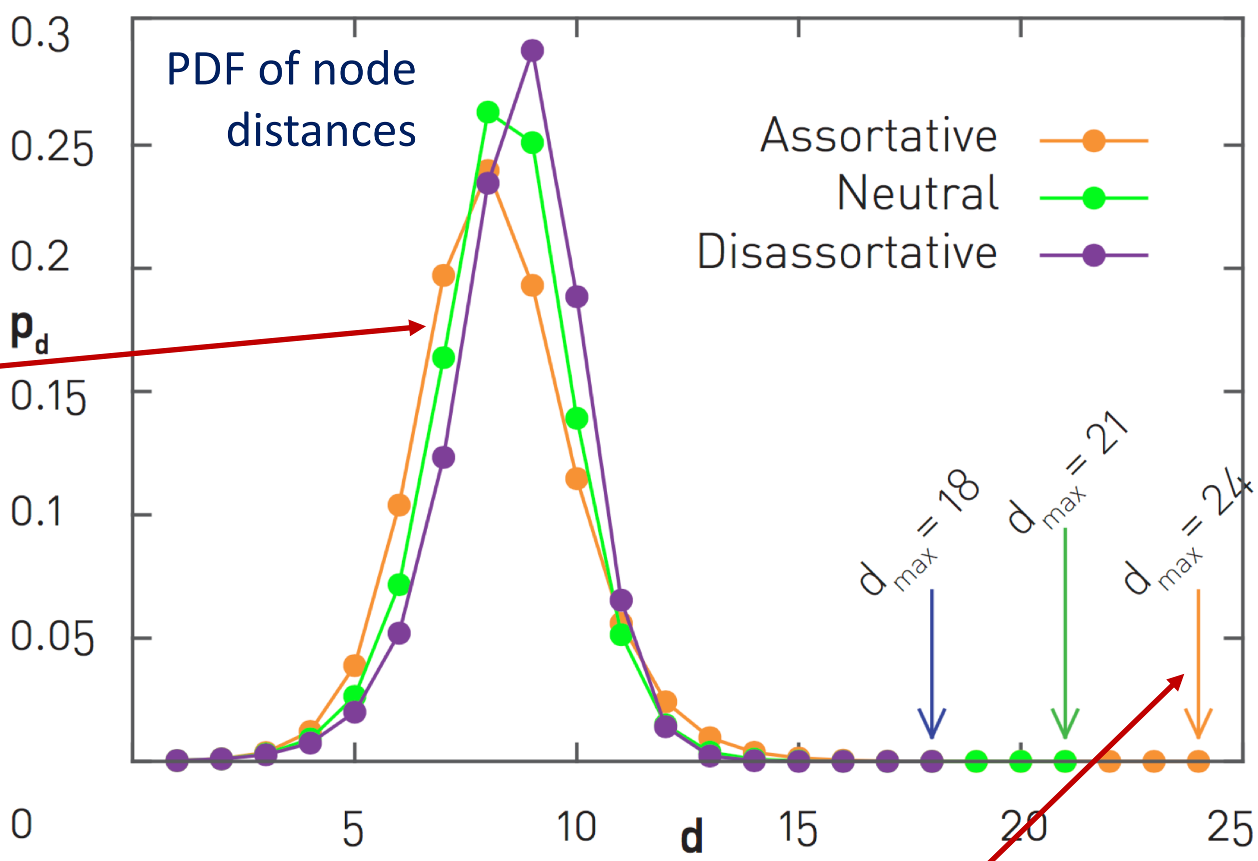
Hashtag network disassortativity



Implications of assortativity

(dis)assortativity influences the path length and the network diameter

In assortative networks the average path length decreases

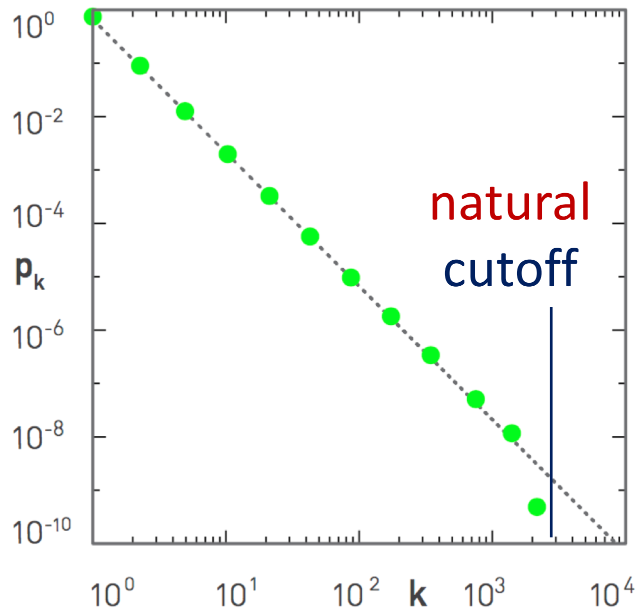


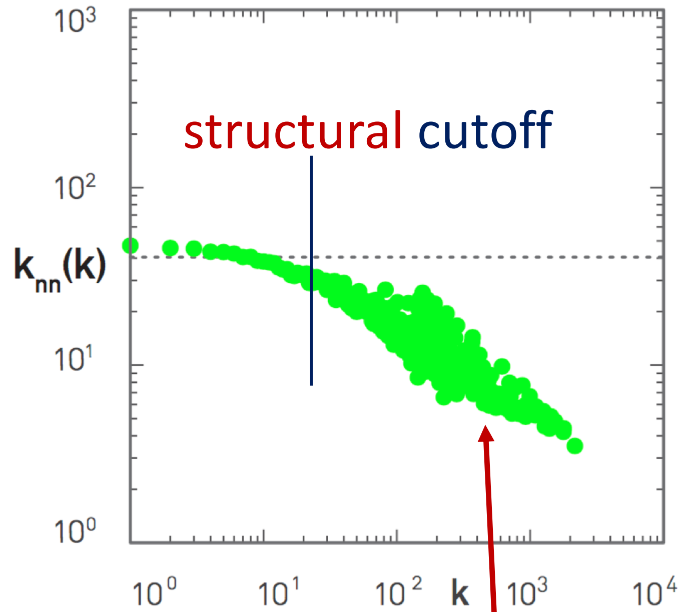
… and the diameter increases!!! (nodes with low degree tend to interconnect)

Structural Disassortativity

Rationale for (dis)assortativity

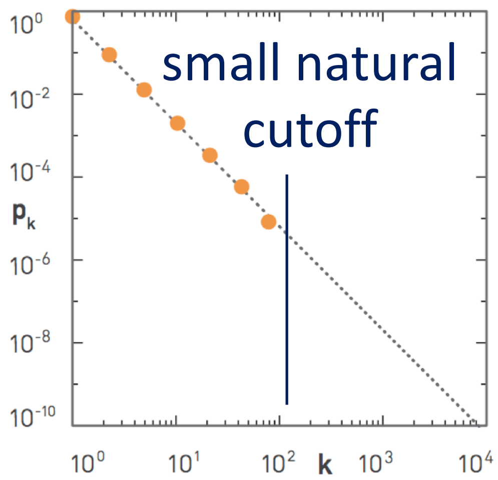
(dis)Assortativity can be linked to structural network properties

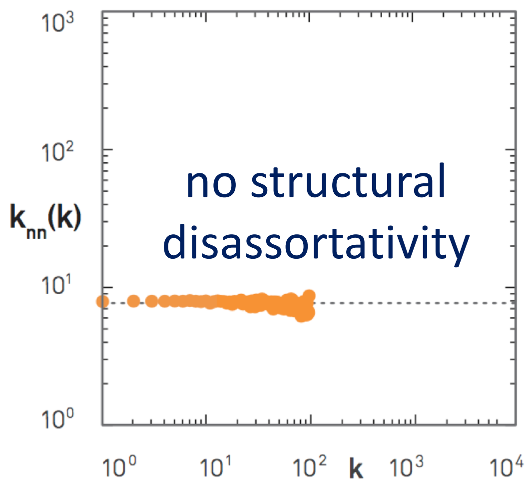




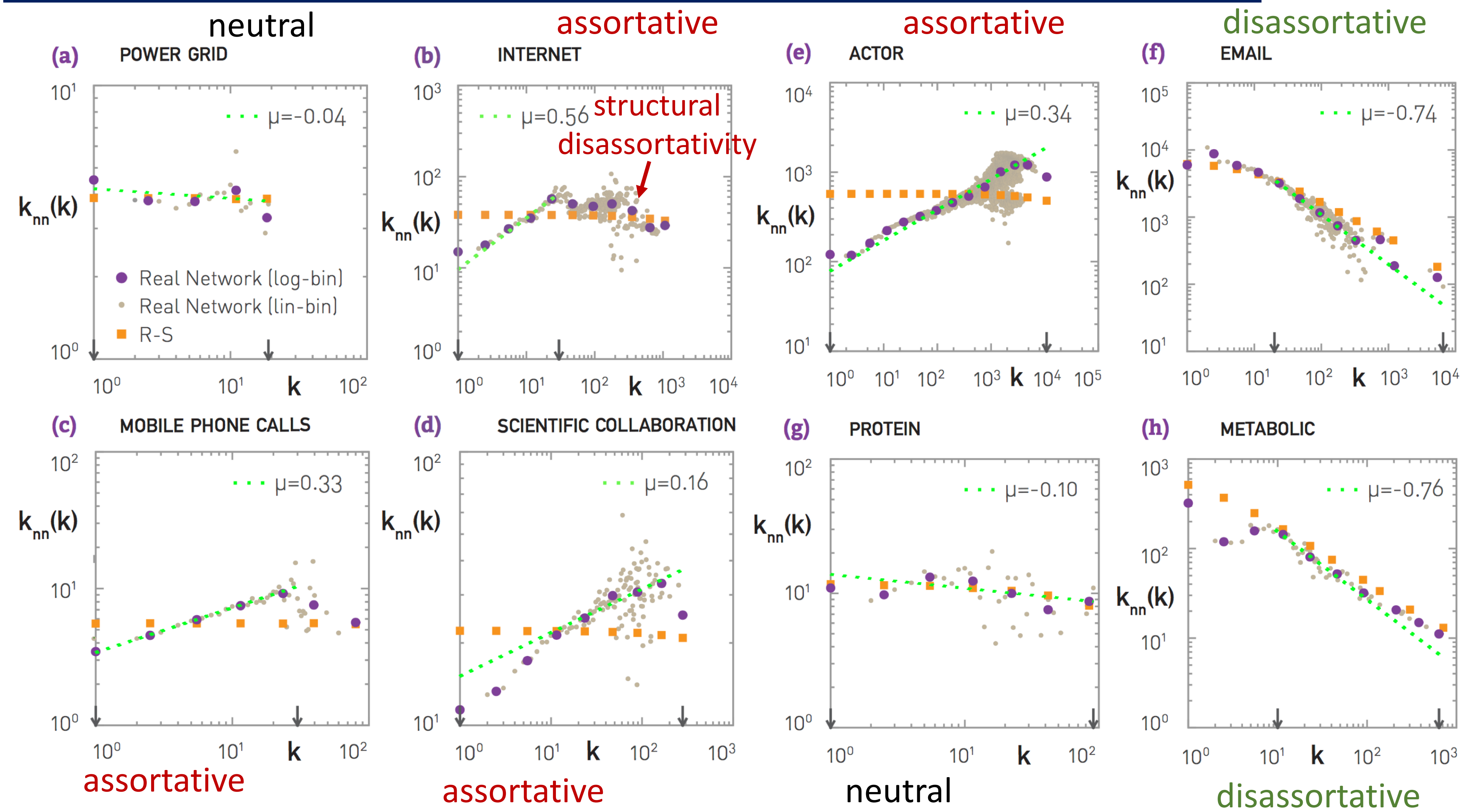
structural disassortativity

large degrees cannot be supported by a neutral network





Real networks



Social networks are assortative, most with a structural cutoff

Randomization curve

Real networks may look as disassortative because

- they really involve disassortative effects
- they do not but just have it as structural

Check with the yellow R-S curve (null model/unbiased):

- it is a degree preserving randomization
- at each randomization step we check that we do not have more than one link between any node pairs
- obtained for 100 independent trials
- If k_{nn} does not change → disassortativity is due to a structural reason (i.e., on the degree distribution)
- if something changes → deeper reasons

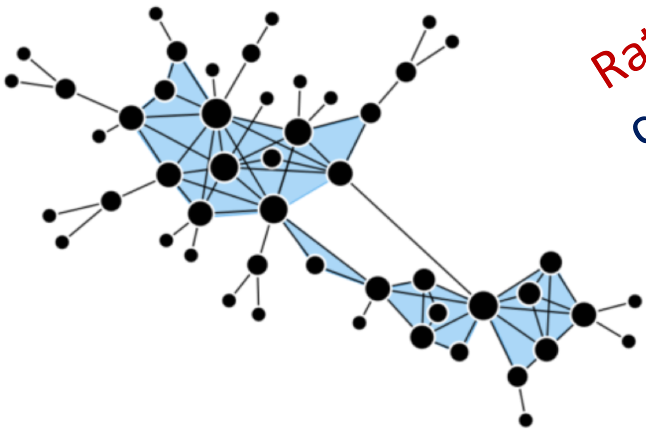
Questions ?

Clustering coefficient

What is the Clustering coefficient?

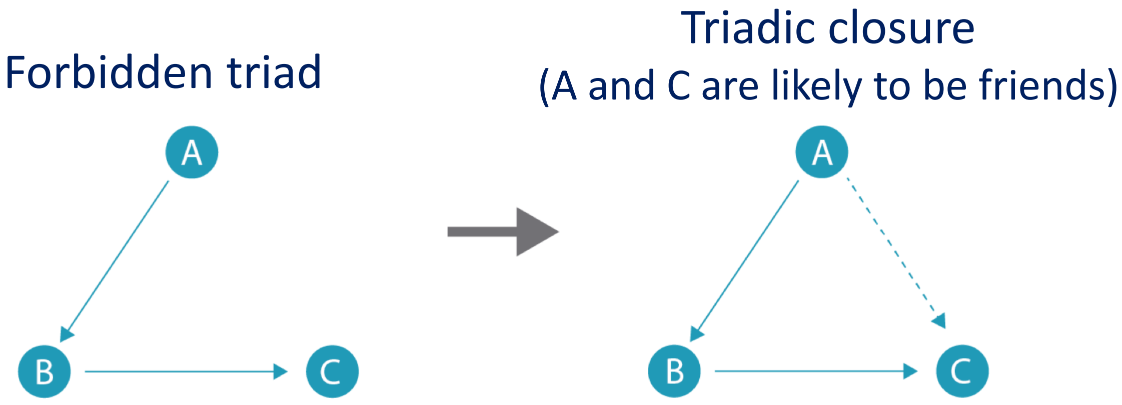
Local clustering coefficient [edit]

The **local clustering coefficient** of a vertex (node) in a graph quantifies how close its neighbours are to being a clique (complete graph). Duncan J. Watts and Steven Strogatz introduced the measure in 1998 to determine whether a graph is a small-world network.



Rationale: how strongly connected is the network locally / general indication of the graph's tendency to be organized into clusters

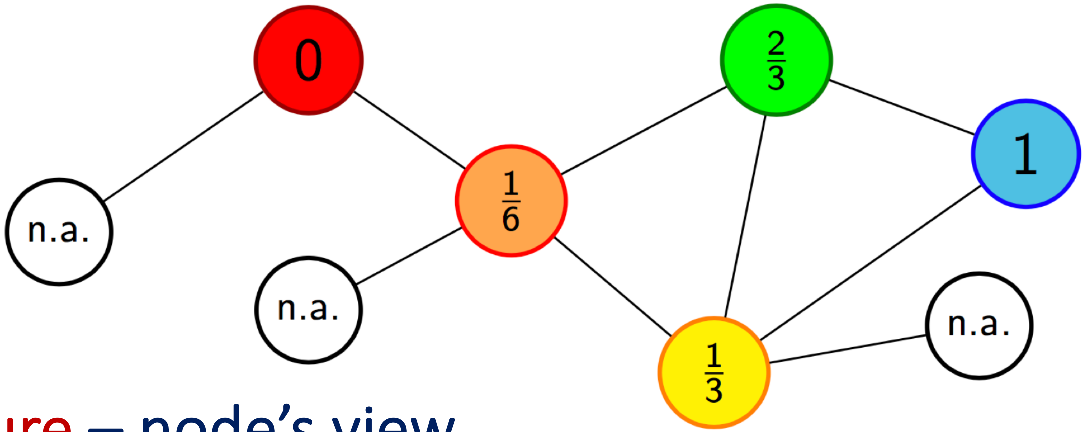
Triadic closure



Triadic closure

- A and C are likely to have the opportunity to meet because they have a common friend B
- The fact that A and C is friends with B gives them the basis of trusting each other
- B may have the incentive to bring A and C together, as it may be hard for B to maintain disjoint relationships

Clustering coefficient and triadic closure



A measure for triadic closure – node's view

- Clustering coefficient C_i
- Counts the fraction of pairs of neighbours which form a triadic closure with node *i*

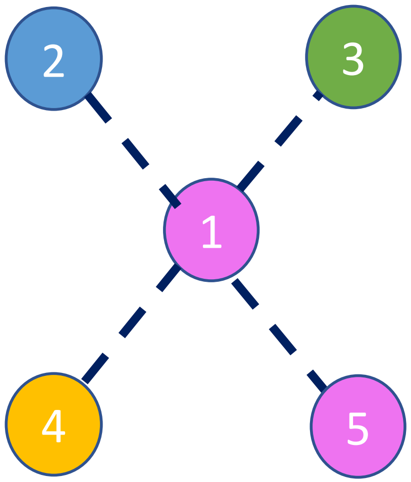
$$C_i = \frac{1}{|\mathcal{N}_i|(|\mathcal{N}_i|-1)} \sum_{\substack{(j,k)\in\mathcal{N}_i^2 \\ j\neq k}} \mathrm{tc}_{i,j,k}$$

where $tc_{ijk} = 1$ if the triplet (i,j,k) forms a triadic closure, and zero otherwise

Examples

not connected neighbourhood

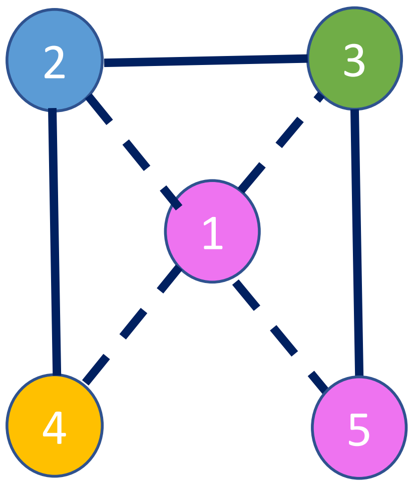
$C_1 = 0$



$\langle C \rangle = 0$

weakly connected neighbourhood

$C_1 = ½ = 3 / (4x3/2)$

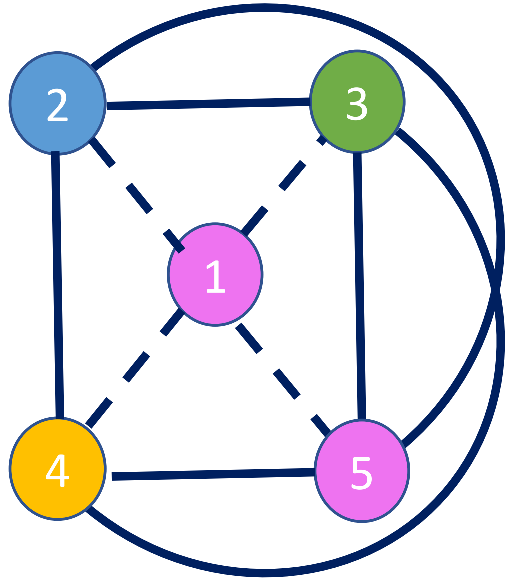


$C_2 = C_3 = ⅔, C_4 = C_5 = 1$

$\langle C \rangle = 0.766$

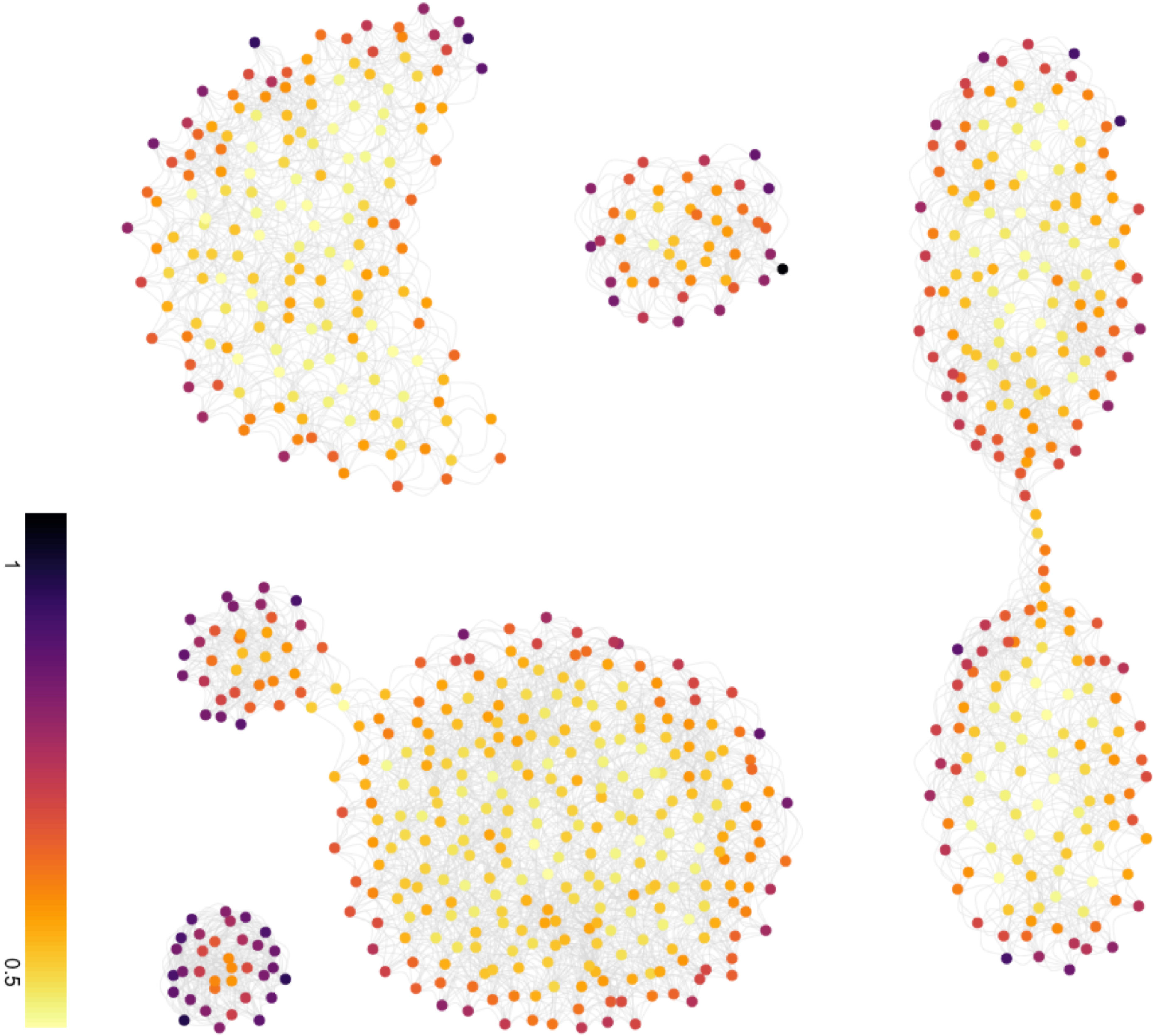
strongly connected neighbourhood

$C_1 = 1 = 6 / (4x3/2)$



$\langle C \rangle = 1$

Visual example 1



Visual example 2

But clustering coefficient is generally hard to see and visual interpretation is considered unreliable