UNIVERSITÀ DEGLI STUDI DI PADOVA

Social Network Analysis

A.Y. 23/24

Communication Strategies

Closeness and Harmonic centralities

importance of nodes as spreaders of information

Closeness centrality

a definition

Closeness centrality

From Wikipedia, the free encyclopedia

In a connected graph, **closeness centrality** (or **closeness**) of a node is a measure of centrality in a network, calculated as the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the graph. Thus, the more central a node is, the *closer* it is to all other nodes.

Closeness was defined by Bavelas (1950) as the reciprocal of the **farness**,[1][2] that is:

$$C(x) = \frac{1}{\sum_y d(y,x)}.$$

where $d(y,x)$ is the distance between vertices x and y. When

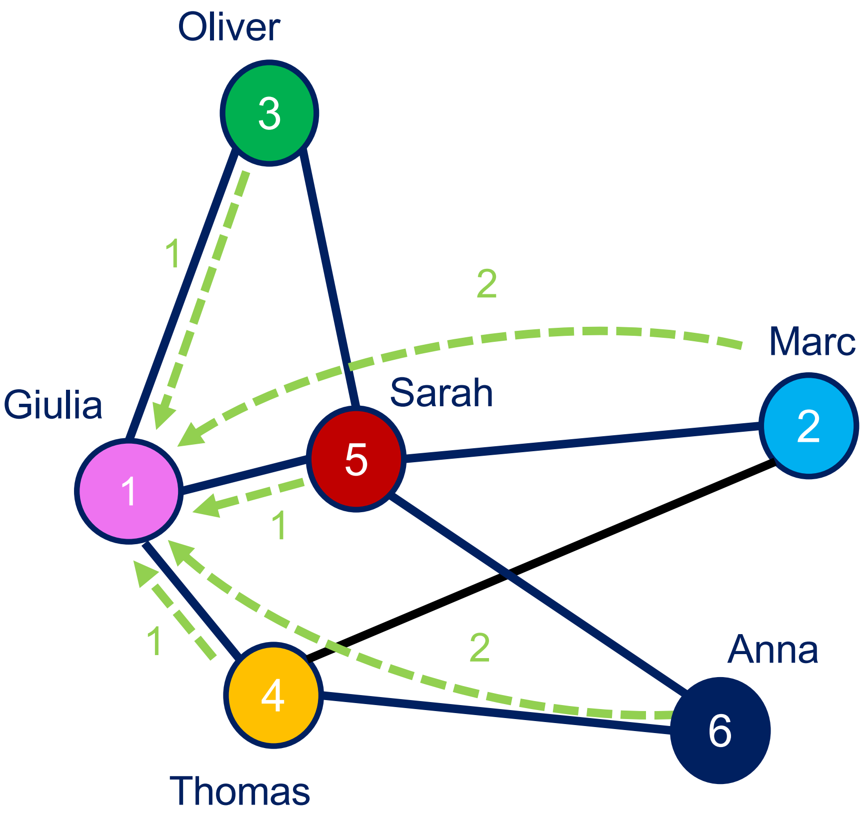
Rationale: the node which is the easiest to reach, the one which is the best for spreading information

An example

on how to calculate closeness centrality

count the lengths of the shortest paths leading to Giulia

1 + 2 + 1 + 2 + 1 = 7



Closeness

0.1429	Giulia
0.1250	Marc
0.1250	Oliver
0.1429	Thomas
0.1667	Sarah
0.1250	Anna

Sarah is the preferred node for spreading information

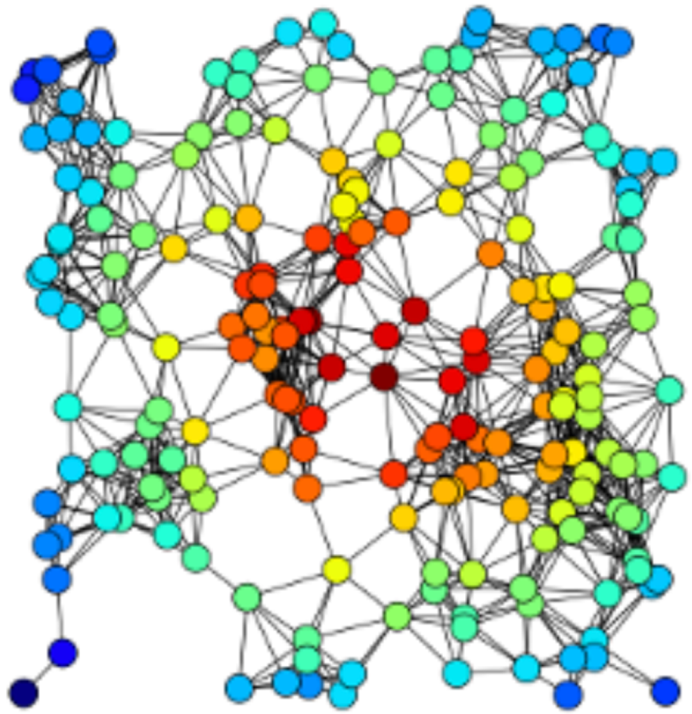
$$C(\text{Giulia}) = 1/7 = 0.1429$$

UNIVERSITÀ DEGLI STUDI DI PADOVA

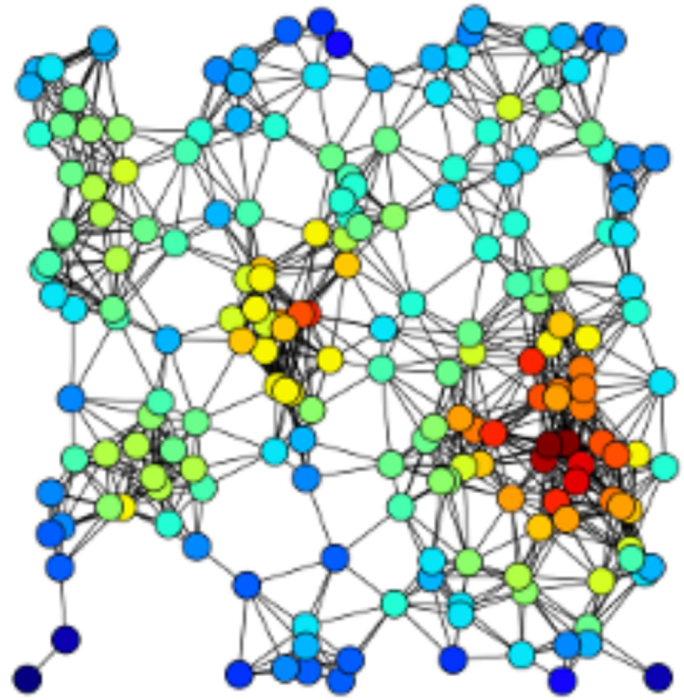
Closeness versus degree centrality

a graphical interpretation

Closeness



Degree



Harmonic centrality

a definition

In disconnected graphs [edit]

When a graph is not strongly connected, a widespread idea is that of using the sum of reciprocal of distances, instead of the reciprocal of the sum of distances, with the convention $1/\infty = 0$:

$$H(x) = \sum_{y \neq x} \frac{1}{d(y,x)}.$$

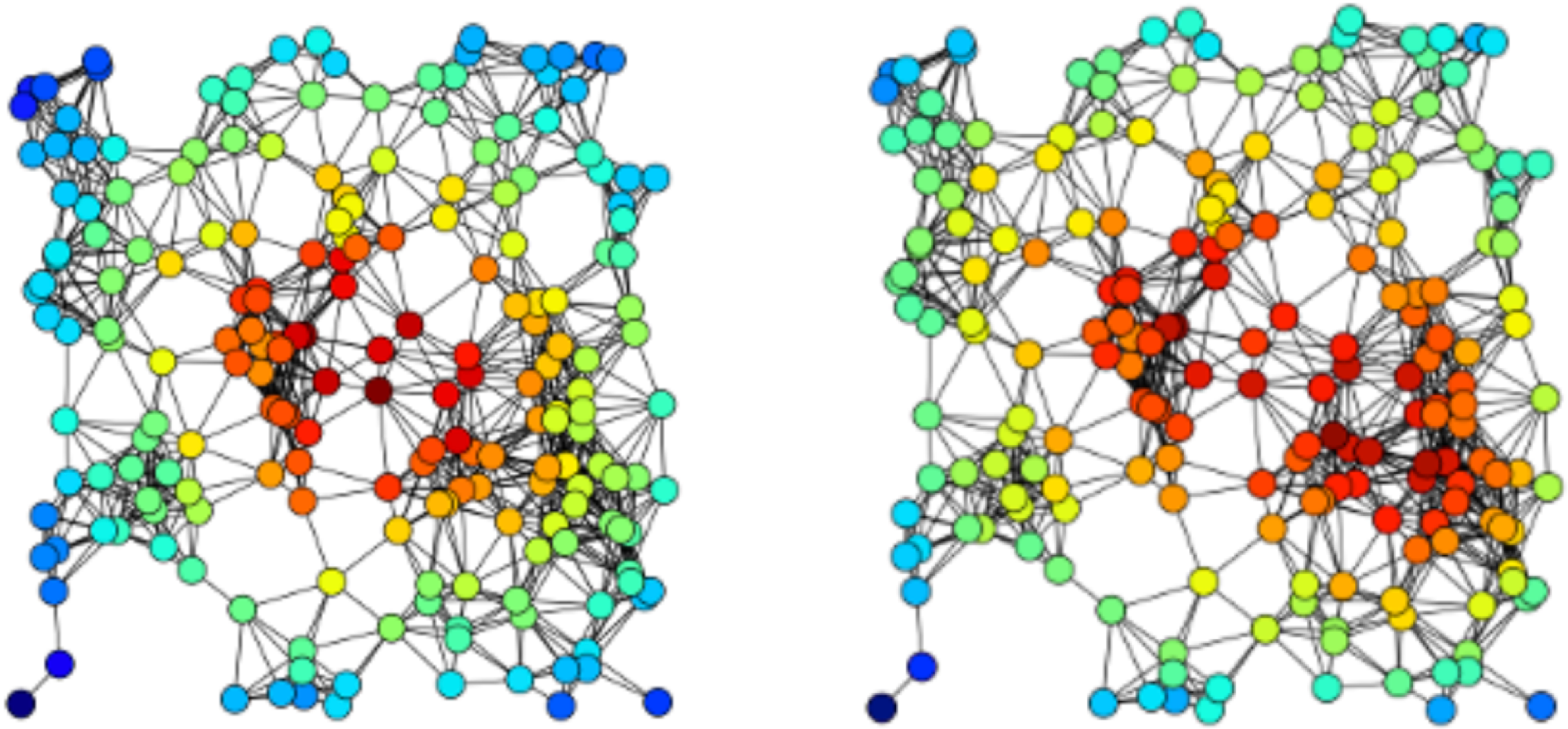
The most natural modification of Bavelas's definition of closeness is following the general principle proposed by Marchiori and Latora (2000)[3] that in graphs with infinite distances the harmonic mean behaves better than the arithmetic mean. Indeed, Bavelas's closeness can be described as the denormalized reciprocal of the arithmetic mean of distances, whereas harmonic centrality is the denormalized reciprocal of the harmonic mean of distances.

Closeness versus harmonic centrality

a graphical interpretation

Closeness

Harmonic



Betweenness centrality

importance of nodes as bridges or brokers

Betweenness centrality
a definition

Betweenness centrality

From Wikipedia, the free encyclopedia

In graph theory, **betweenness centrality** is a measure of centrality in a graph based on shortest paths. For every pair of vertices in a connected graph, there exists at least one shortest path between the vertices such that either the number of edges that the path passes through (for unweighted graphs) or the sum of the weights of the edges (for weighted graphs) is minimized. The betweenness centrality for each vertex is the number of these shortest paths that pass through the vertex.

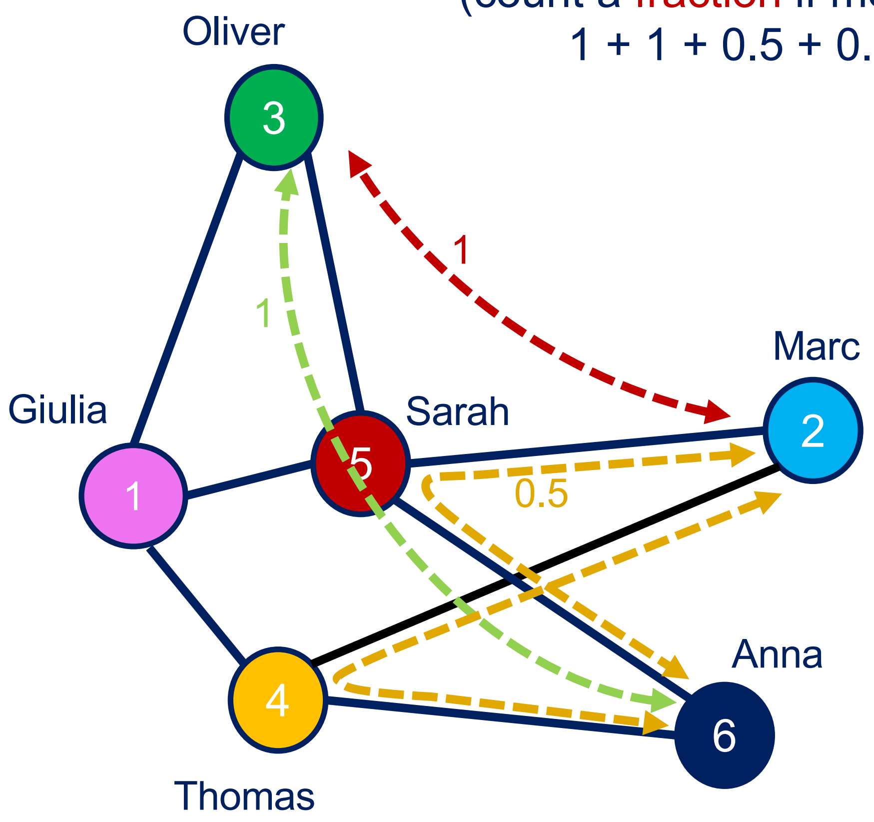
Betweenness centrality was devised as a general measure of centrality:[1] it applies to a wide range of problems in network theory, including problems related to social networks, biology, transport and scientific cooperation. Although earlier authors have intuitively described centrality as based on betweenness, Freeman (1977) gave the first formal definition of betweenness centrality.

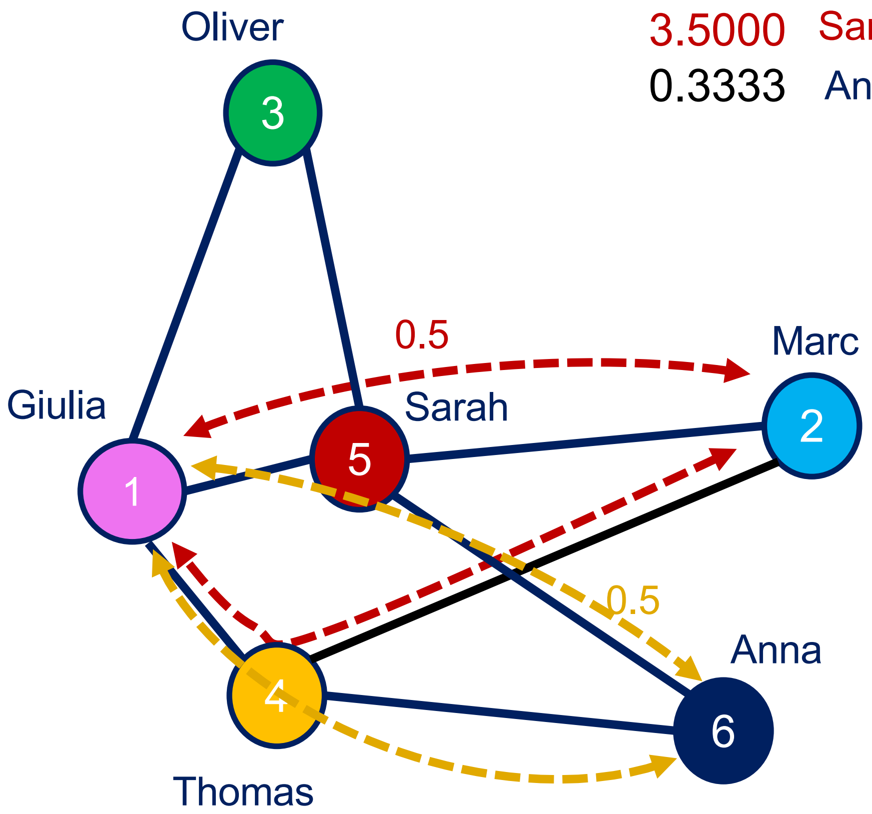
Rationale: the node which takes you elsewhere (bridge, broker)

An example

on how to calculate betweenness centrality

count the # of shortest paths
passing through Sarah
(count a fraction if more than one path)
1 + 1 + 0.5 + 0.5 + 0.5 = 3.5





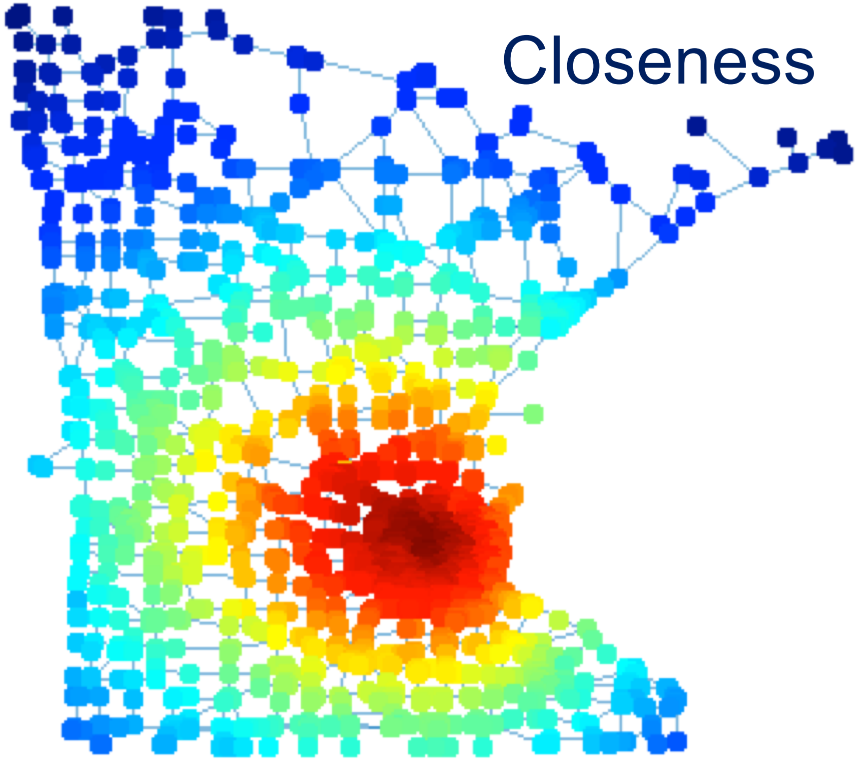
Betweenness

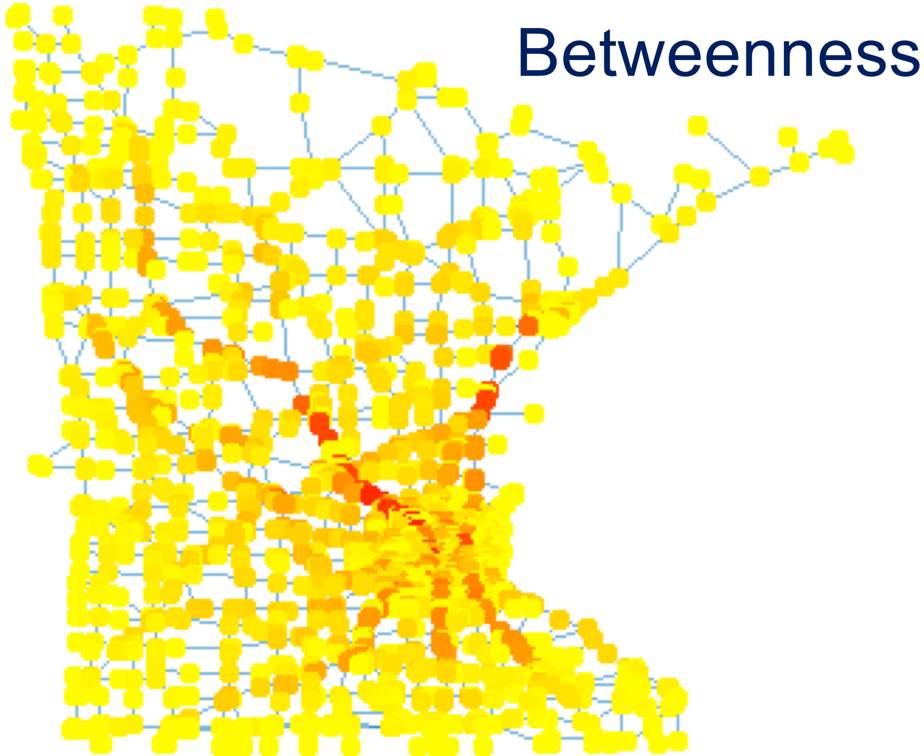
1.3333	Giulia
0.3333	Marc
0	Oliver
1.5000	Thomas
3.5000	Sarah
0.3333	Anna

Closeness vs betweenness centrality

a graphical interpretation

Minnesota road network





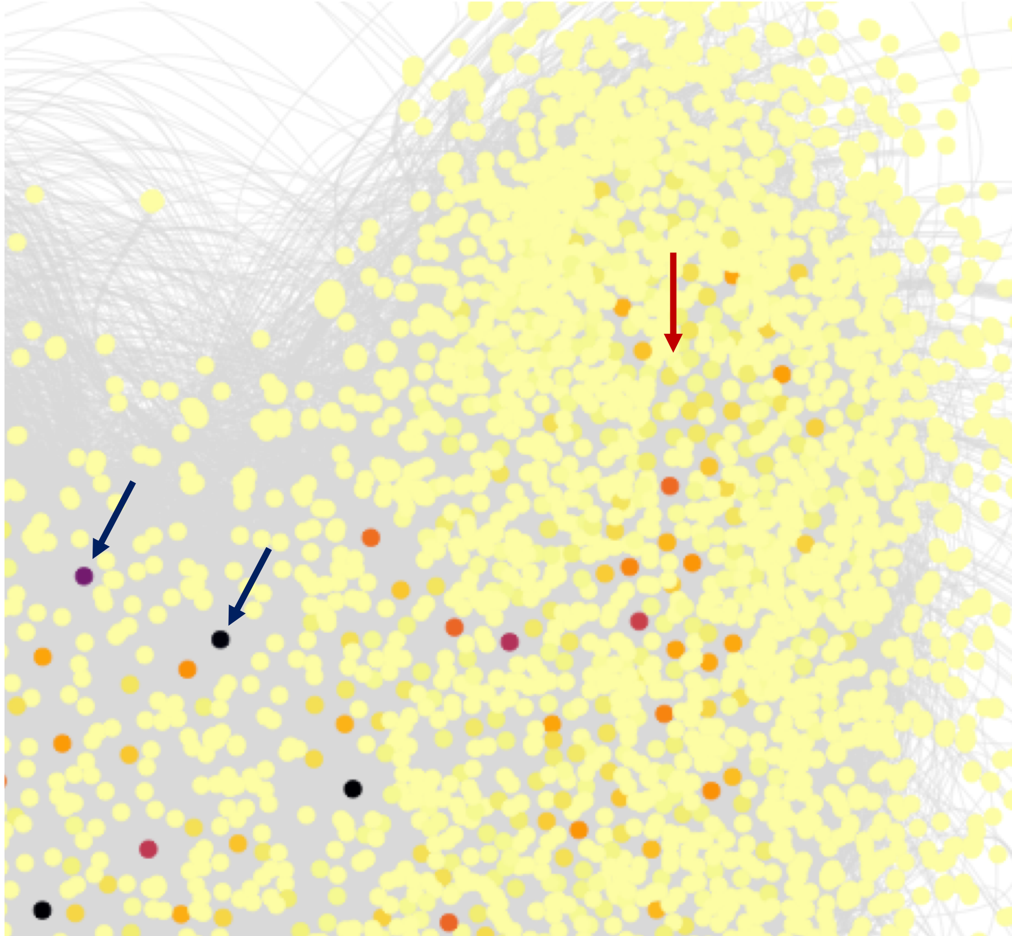
Closeness is a measure of center of gravity (best node to spread info)

Betweenness is a measure of brokerage (i.e., being a bridge)

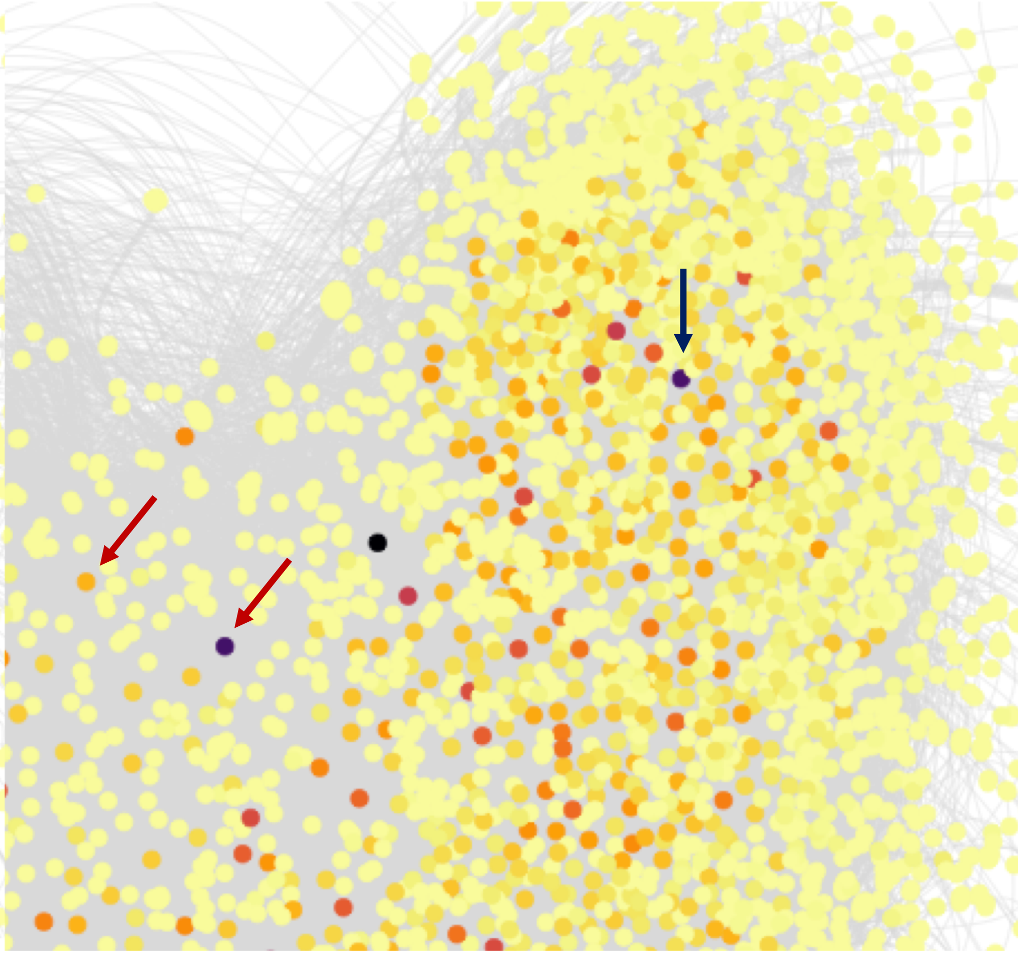
Betweenness vs PageRank centrality

wiki vote network

Betweenness

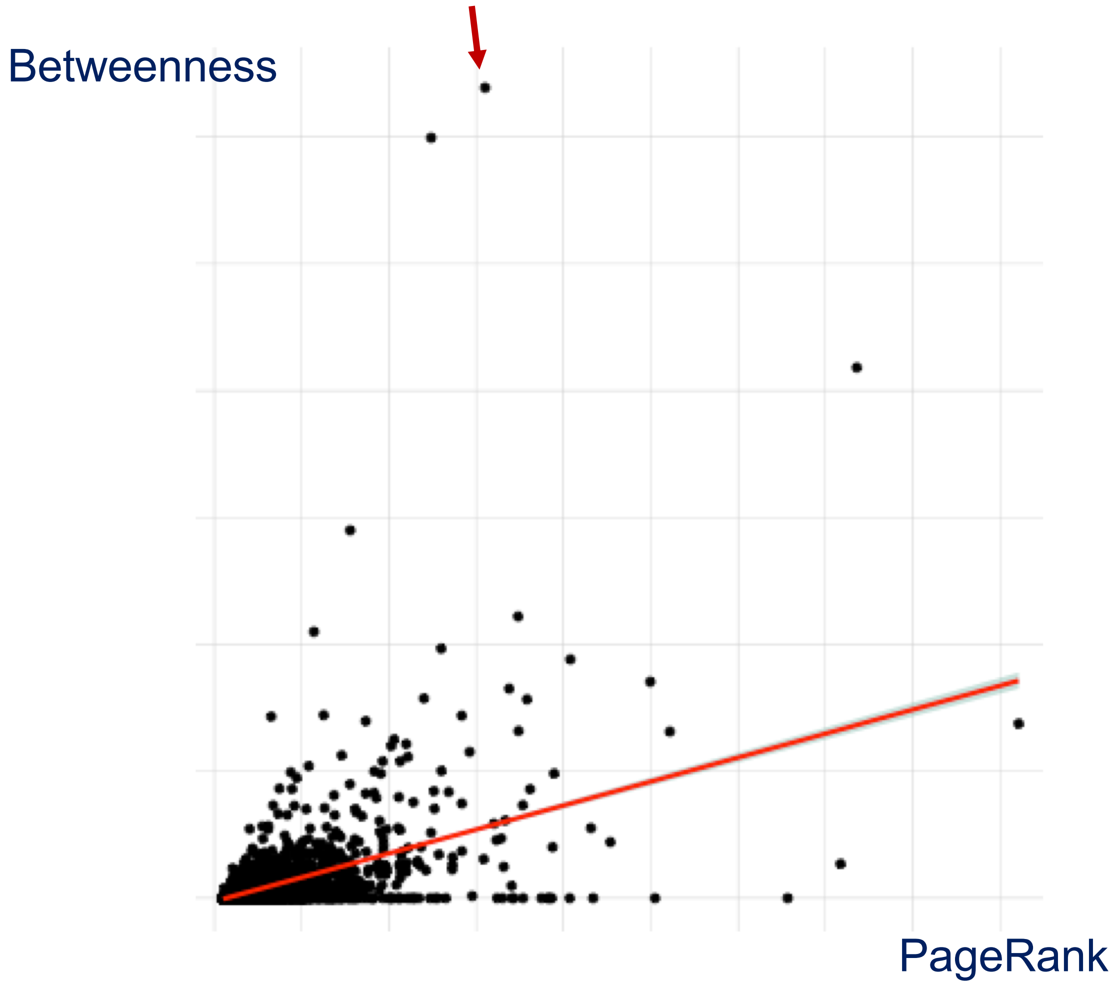


PageRank



Betweenness vs PageRank centrality

a correlation view



Clustering coefficient

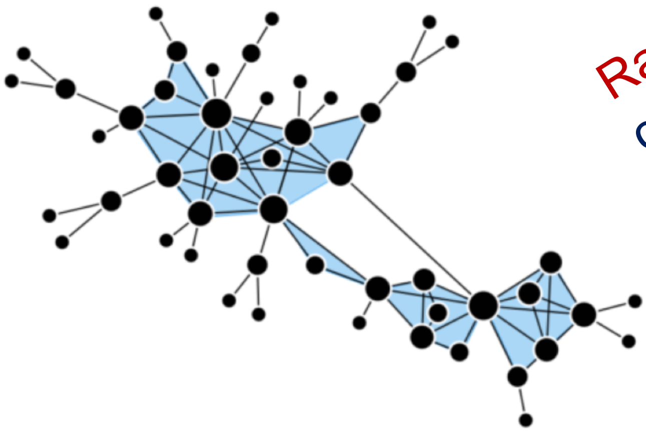
how tightly linked is the network locally

Clustering coefficient

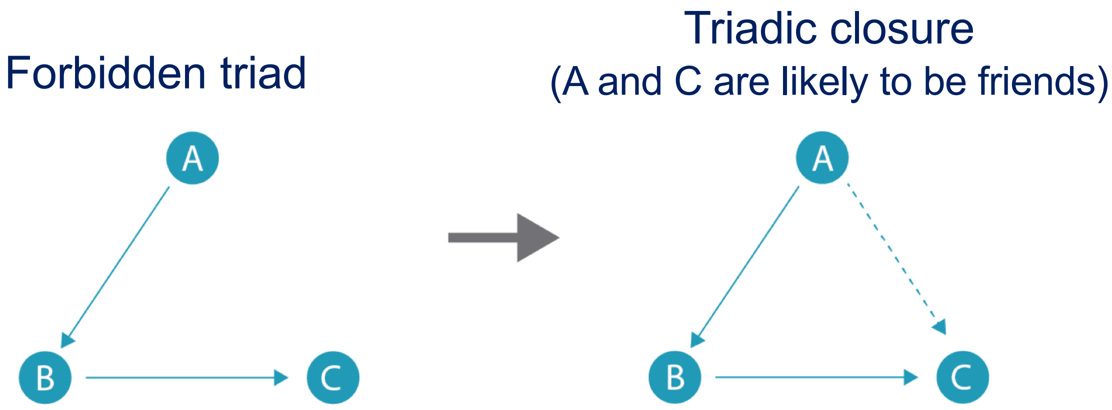
a definition

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



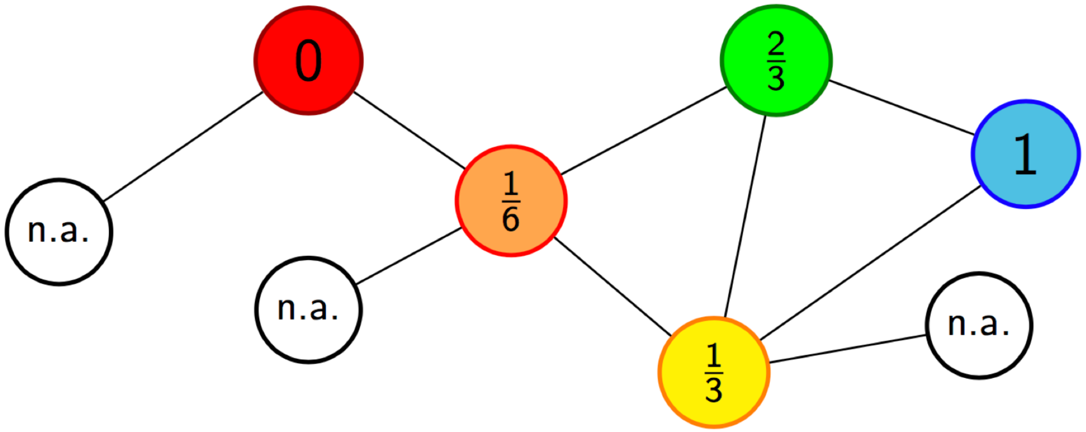
Triadic closure

- 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

Granovetter, The strength of weak ties [1973]
https://www.jstor.org/stable/pdf/2776392.pdf

Local clustering coefficient

a measure of triadic closures



Local Clustering coefficient C_i counts the fraction of pairs of neighbours N_i which form a triadic closure with node i

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

equal to diag($\mathbf{A}^3$)

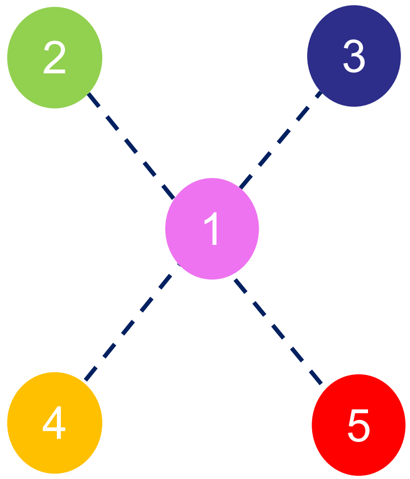
where $tc_{ijk} = 1$ if the triplet (i,j,k) forms a triadic closure, and zero otherwise

Local clustering coefficient

examples

not connected neighbourhood

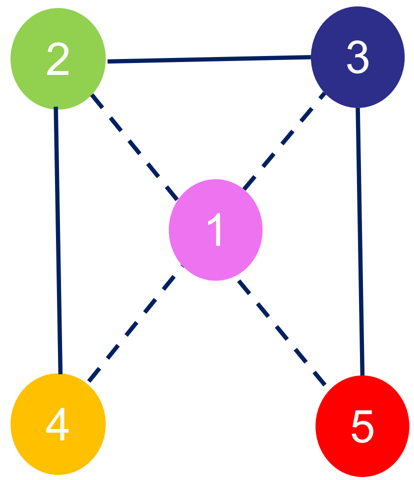
$\langle C \rangle = 0$



$C_1 = 0$

weakly connected neighbourhood

$\langle C \rangle = 0.766$



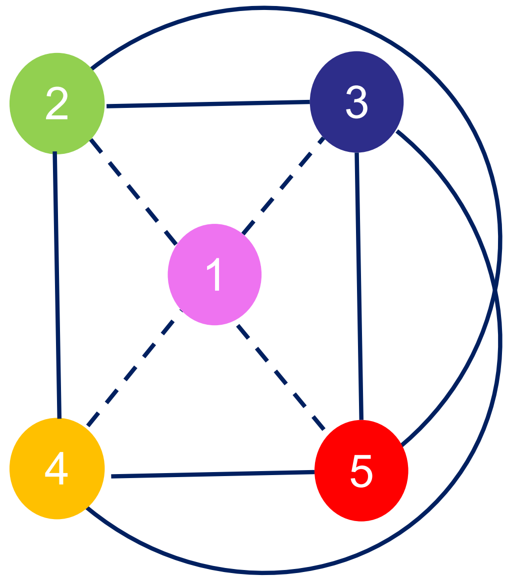
$C_1 = \frac{1}{2} = 3 / (4 \times 3/2)$

$C_2 = C_3 = \frac{2}{3}$

$C_4 = C_5 = 1$

strongly connected neighbourhood

$\langle C \rangle = 1$

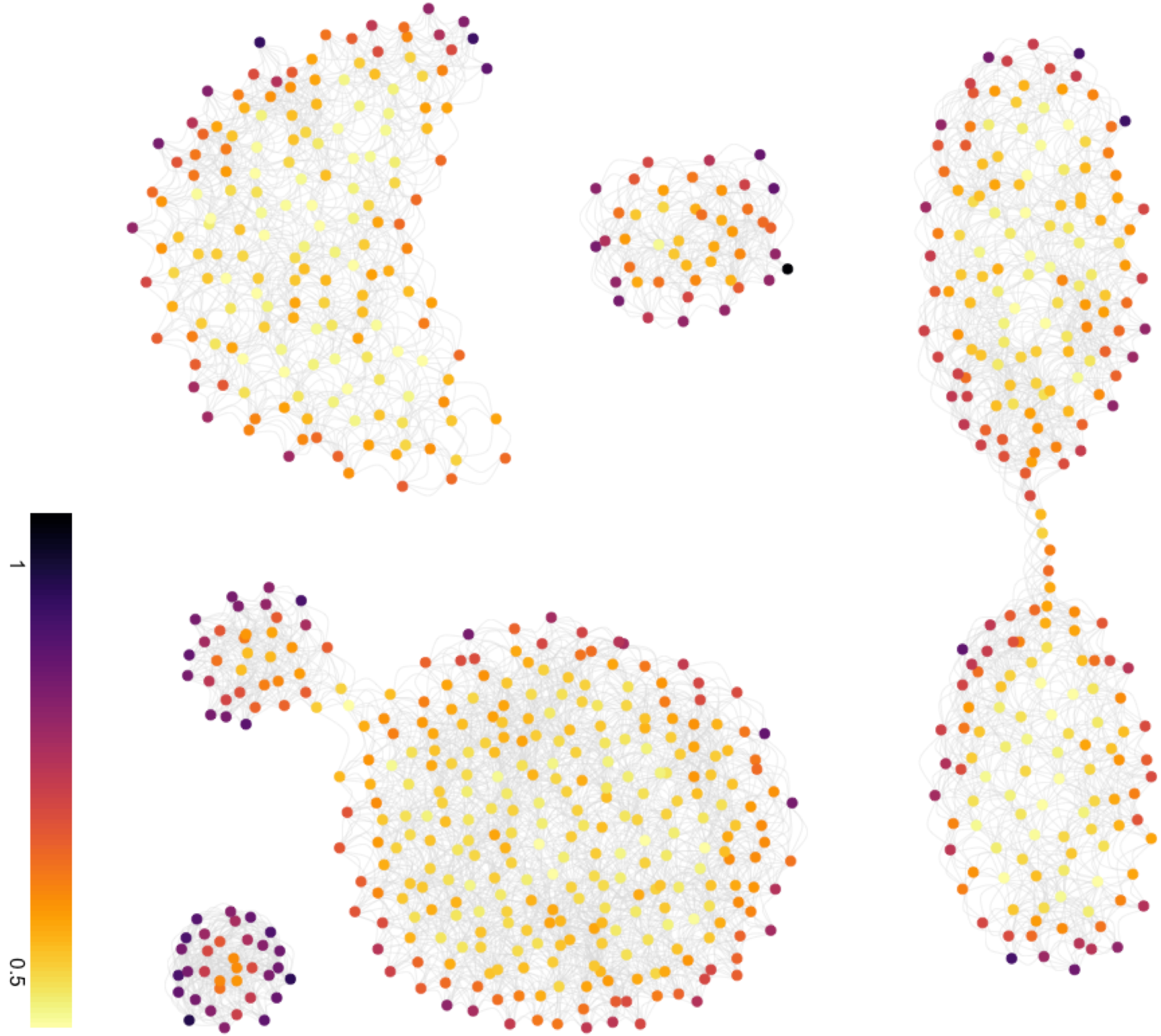


$C_1 = 1 = 6 / (4 \times 3/2)$

Warning

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

Visual example



Wrap-up

on centrality measures

Takeaways

for Closeness, Betwenness and Clustering coefficient

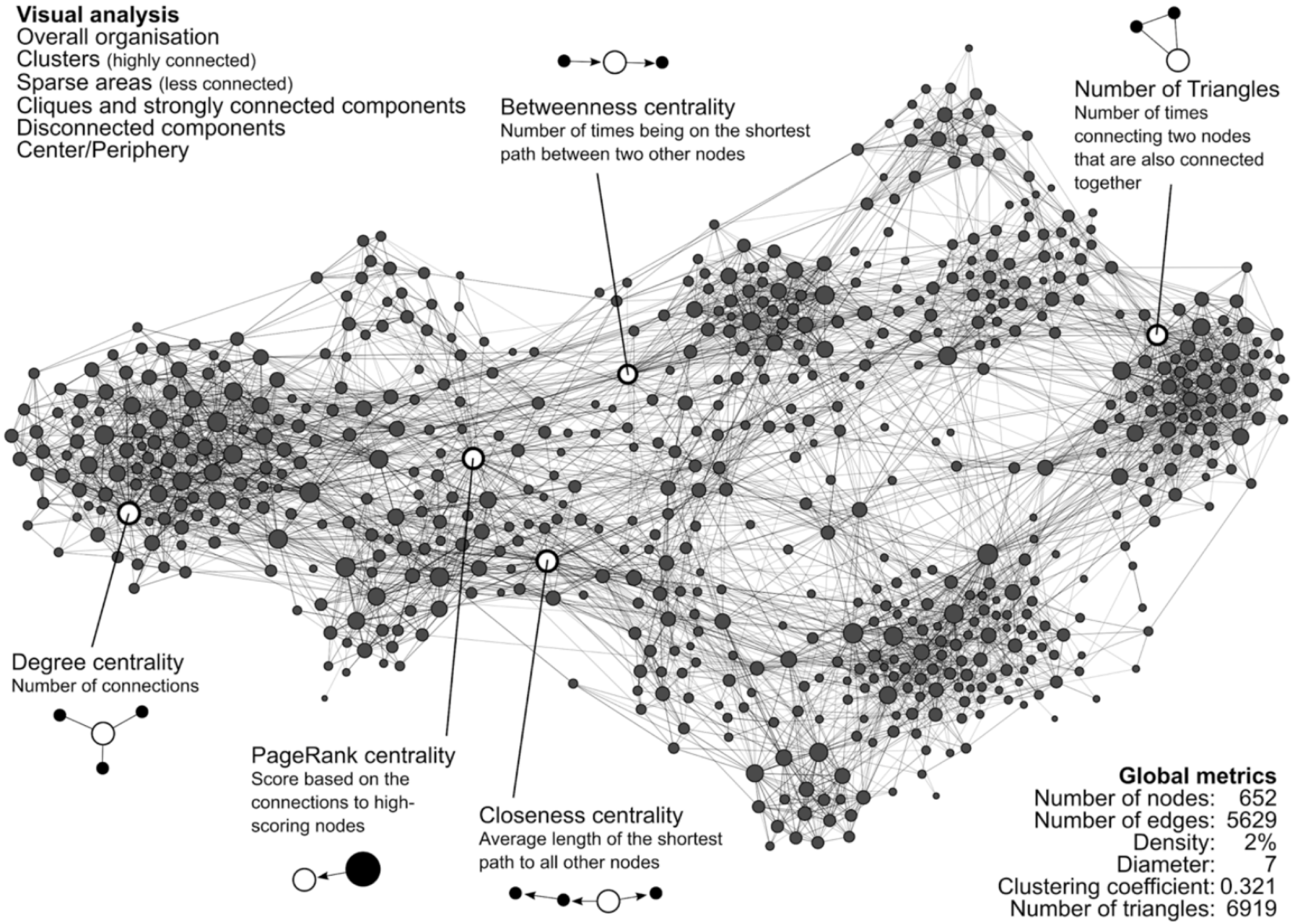
- Closeness, betweenness and clustering coefficient are alternative centrality measures that have a different view wrt PageRank
- They provide useful insights especially in social networks, as they are linked to sociology concepts
- Closeness and betweenness are based on distances, that require algorithms that are less scalable than PageRank
- Exploit their potential at your best

Takeaways
on centrality measures

Centrality measure	Technical property	Meaning
Degree (in/out)	Measures number (and quality) of connections	Cohesion Entrepreneurship
PageRank (authorities/hubs)	Measures number (and quality) of direct and indirect connections	Cohesion Entrepreneurship Closeness/Similarity/Friendship (with a direction) Dependence
Closeness	Measures length of min paths	Visual centrality Significant spreading points Outliers
Betweenness	Measures number of min paths	Brokerage Structural holes Ostracism
Clustering coeff.	Measures number of triadic closures	Centrality in a community Cohesion of the neighbourhood

More on the meaning

https://reticular.hypotheses.org/1745

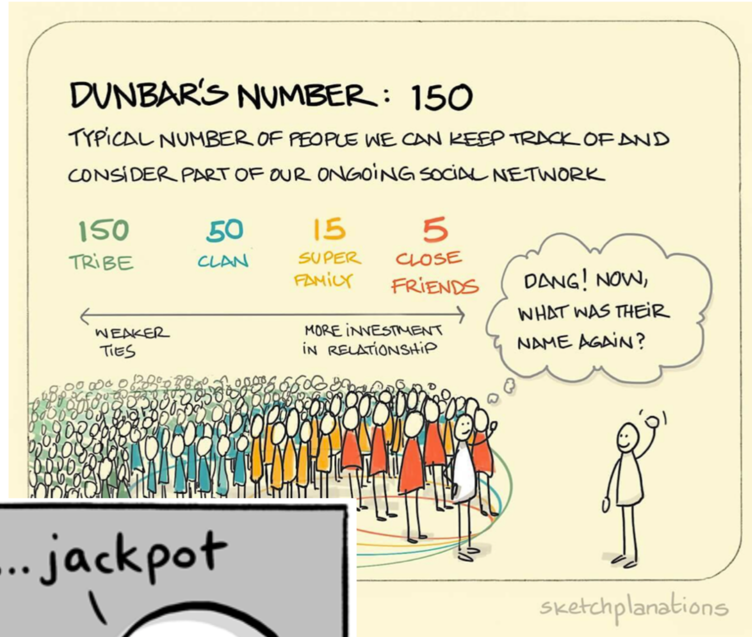


Homophily and Polarization

an overview

Humans and social media

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



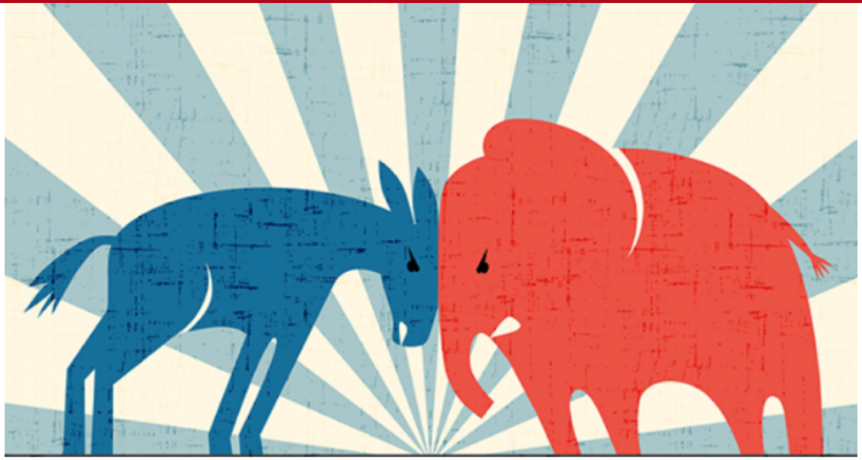
Because we are…

Bounded

Biased

Effects on online behaviour

Polarization



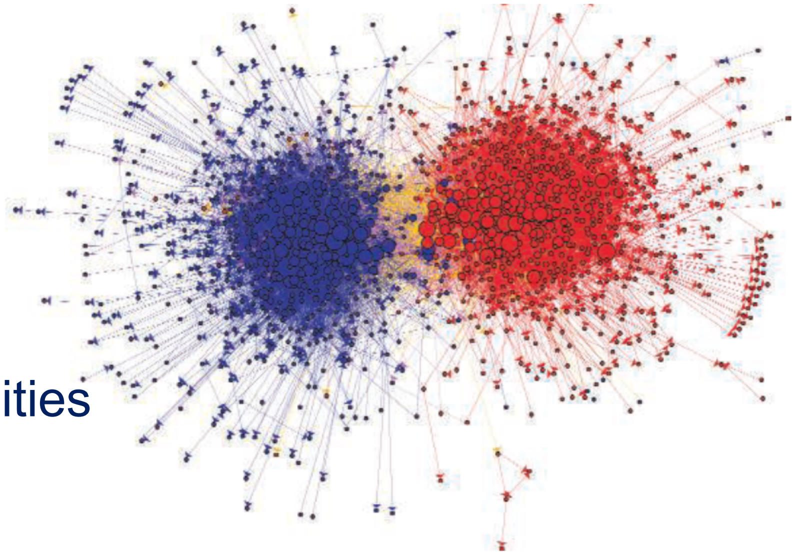
Homophily

Selective exposure

Homophily

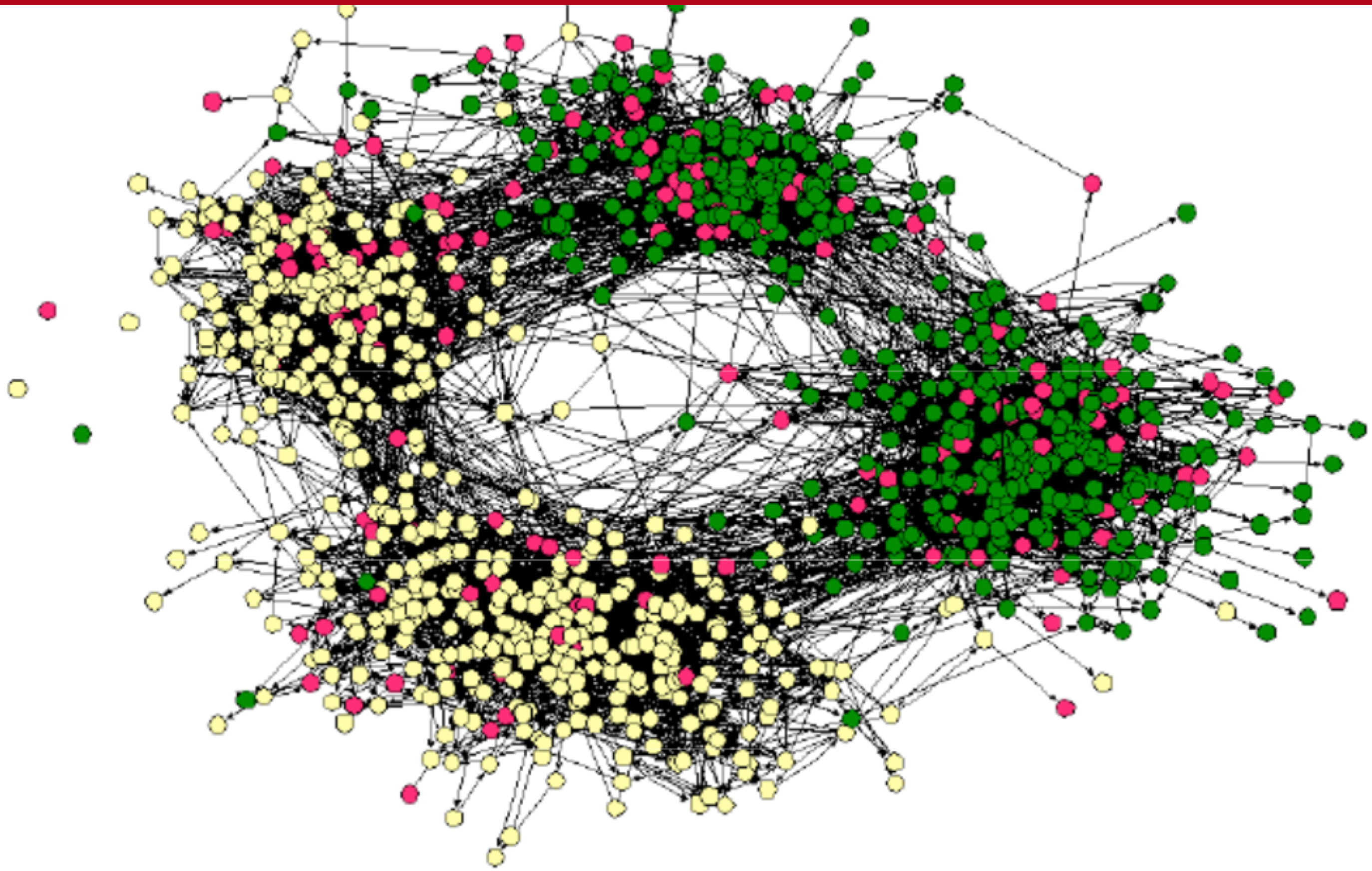
Homophily (from Ancient Greek: *homoû*, 'together' + *philíē*, 'friendship, love') is the tendency of individuals to associate and bond with similar others, as in the proverb "birds of a feather flock together."[1] The presence of homophily has been discovered in a vast array of network studies: over 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 in action

racial segregations

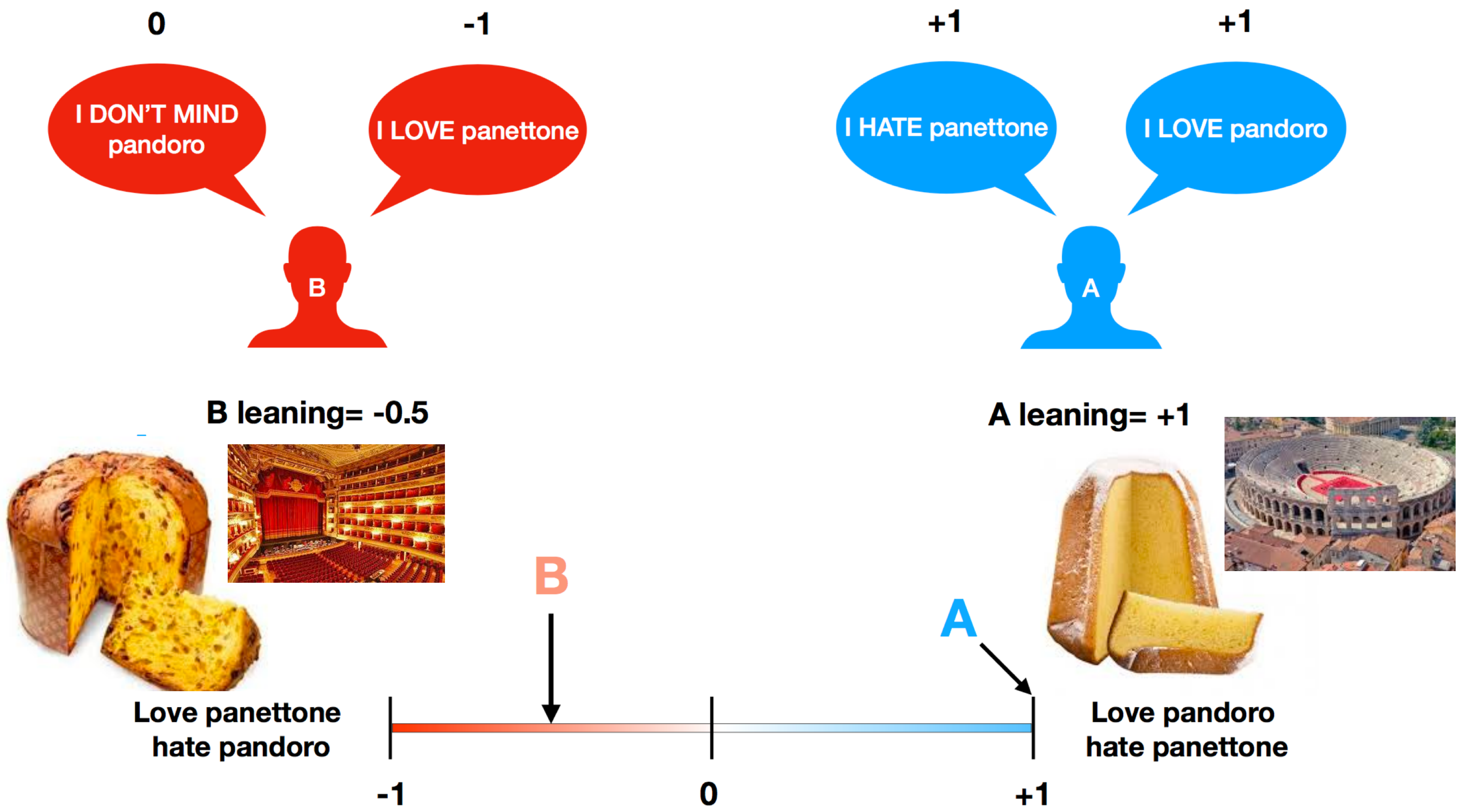


(Easley and Kleinberg, 2010)

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].

UNIVERSITÀ DEGLI STUDI DI PADOVA

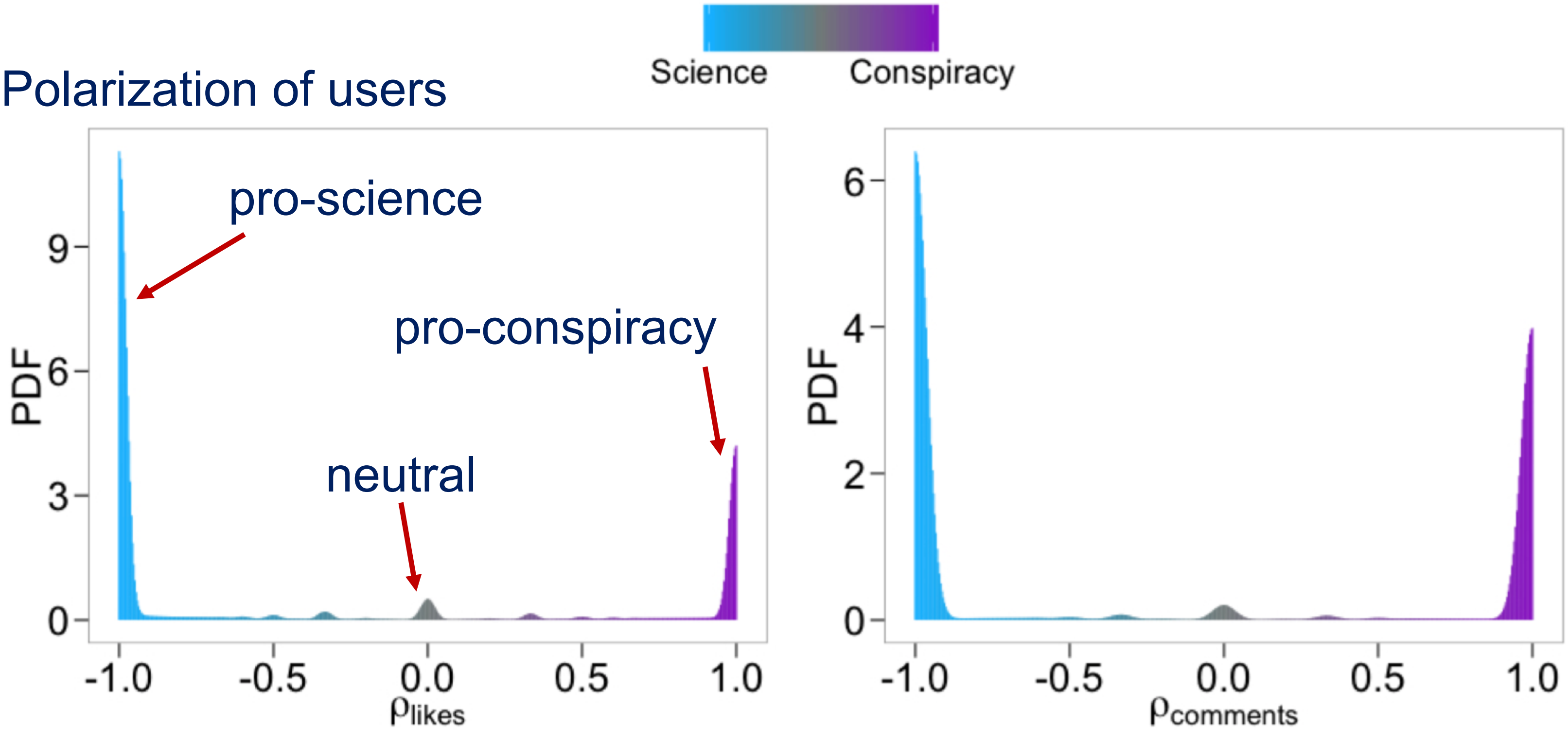
Users leaning
on a controversial topic



Polarization

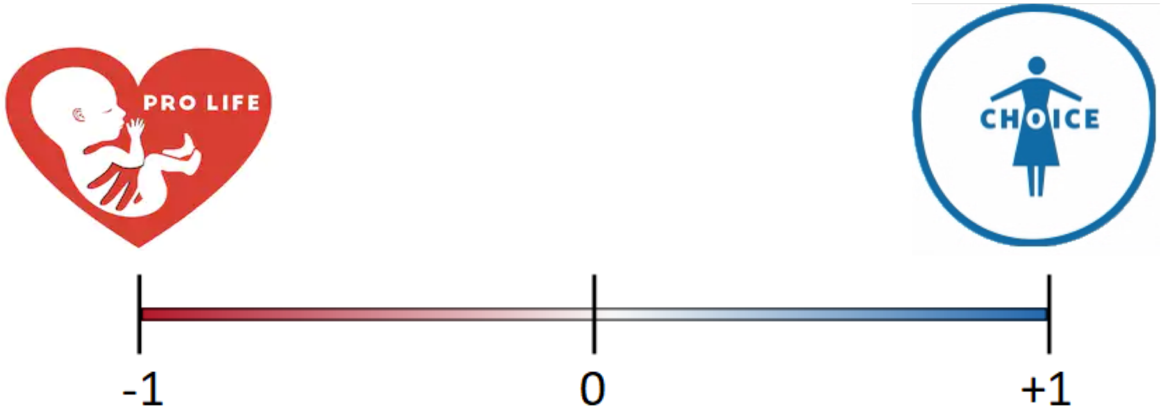
The extreme segregation of users into homogeneous communities based on their opinion on a controversial topic

Polarization of users



Hashtag polarization

polarization in pro-life/pro-choice networks IP (2019)



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

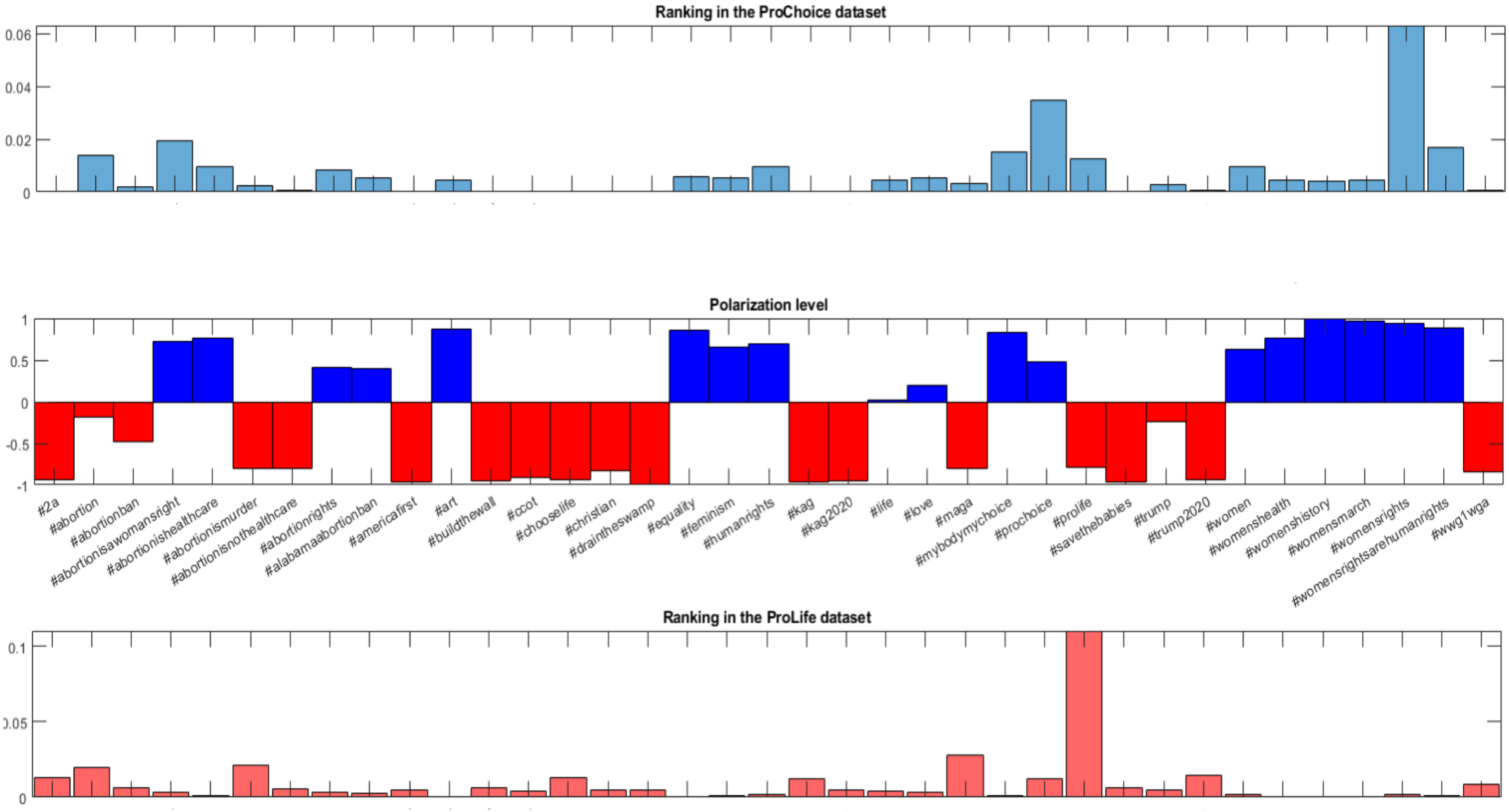
$$P_i = \frac{W_{pc_i} - W_{pl_i}}{W_{pc_i} + W_{pl_i}}$$

ranking values for word *i*

prestige mapping

UNIVERSITÀ DEGLI STUDI DI PADOVA

Hashtag polarization

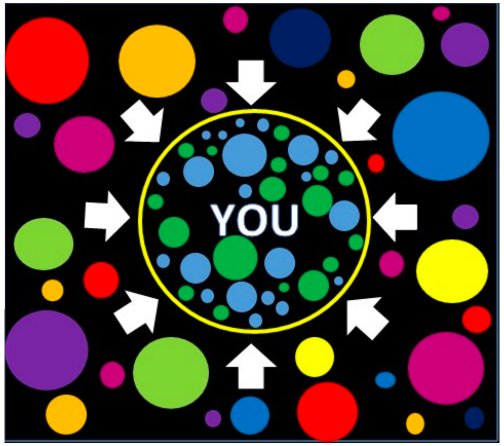


Echo chambers

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]



Università degli Studi di Padova

Echo chamber

a formalization

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

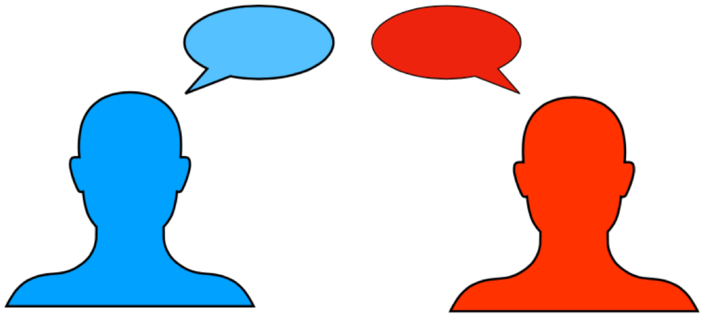
Coexistence of

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

To assess **polarization**

↓

Measure **individual leaning**

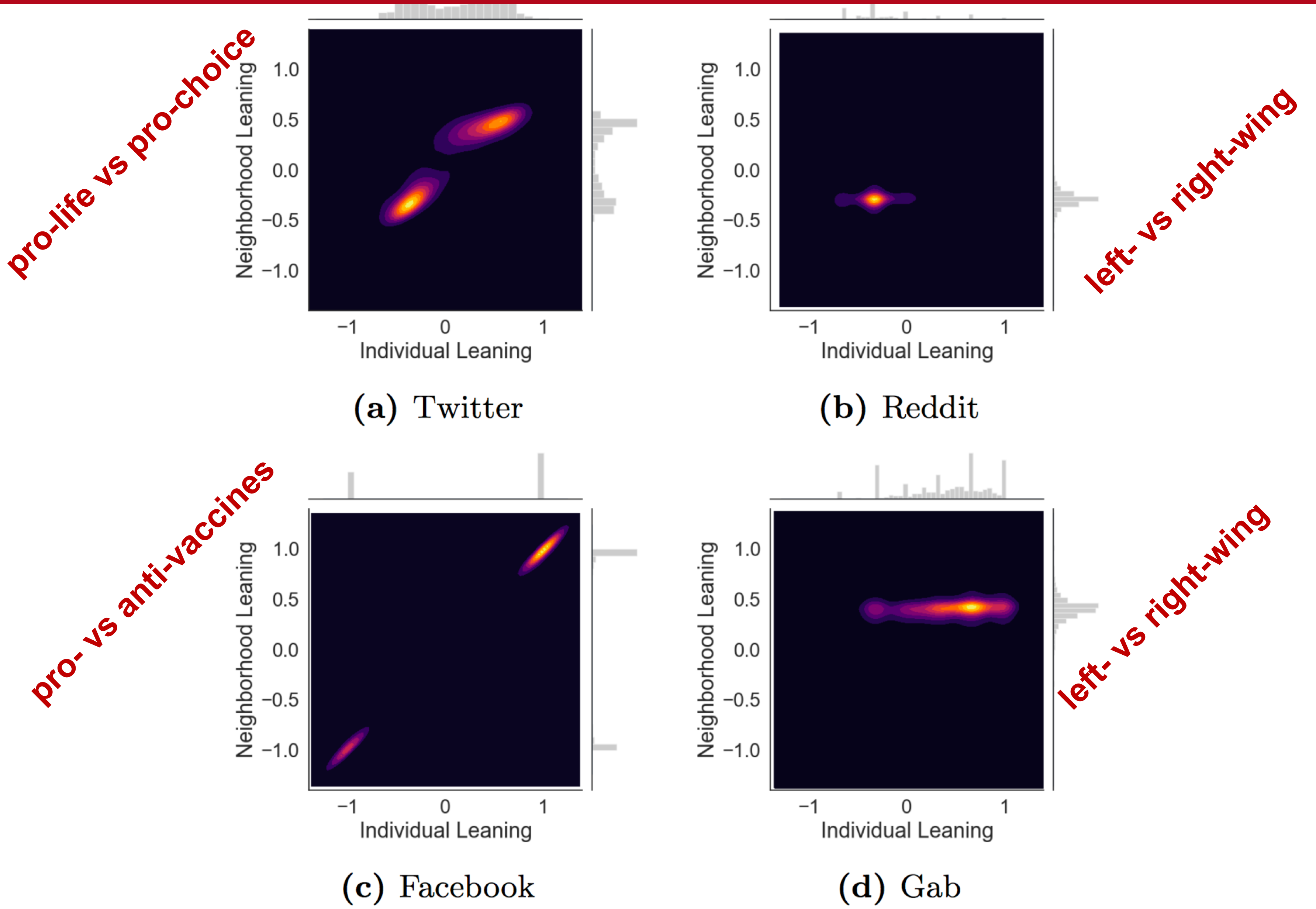


To quantify **homophily**

↓

Build **interaction network**

Echo chamber effect
in social networks



Filter bubble

From Wikipedia, the free encyclopedia

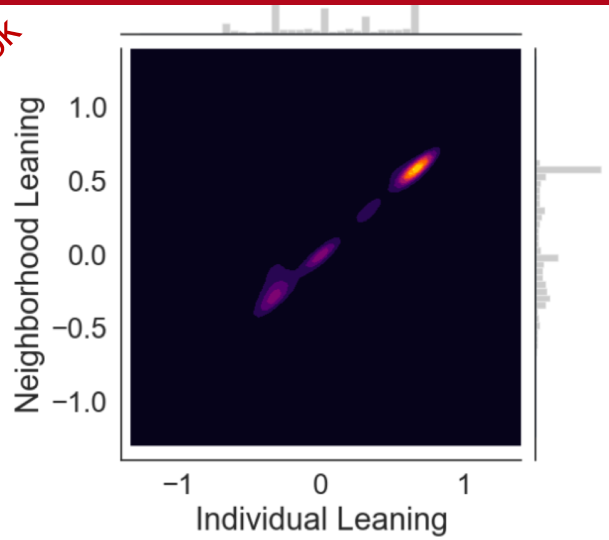
A **filter bubble** – a term coined by internet activist Eli Pariser – is a state of intellectual isolation[1] that allegedly can result from personalized searches when a website algorithm selectively guesses what information a user would like to see based on information about the user, such as location, 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

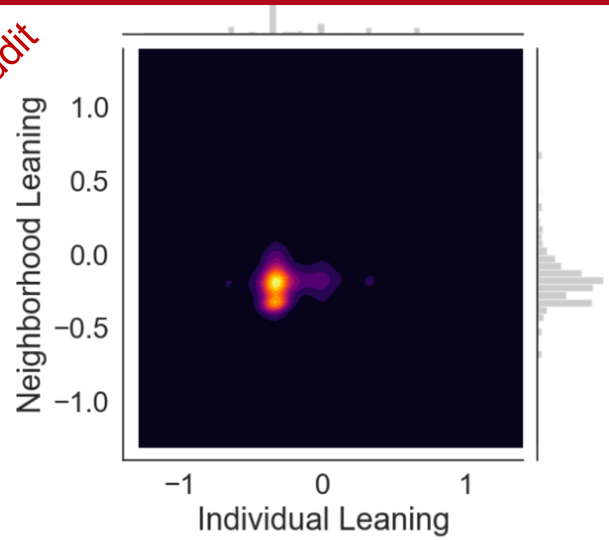
UNIVERSITÀ
DEGLI STUDI
DI PADOVA

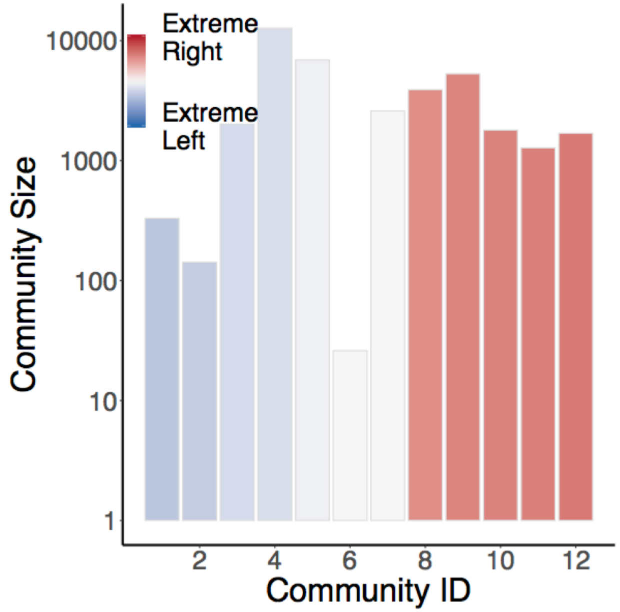
Filter bubbles
in social networks

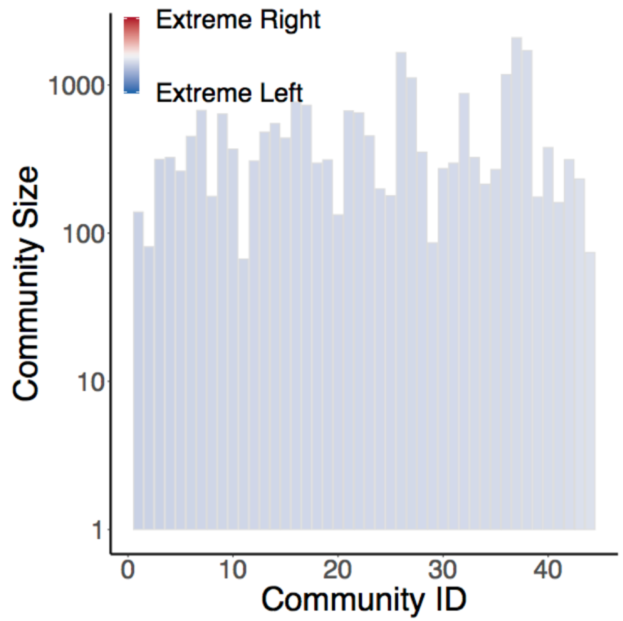
FaceBook



Reddit







- 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

Assortativity

i.e., degree homophily

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

Assortativity

(dis)assortativity 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

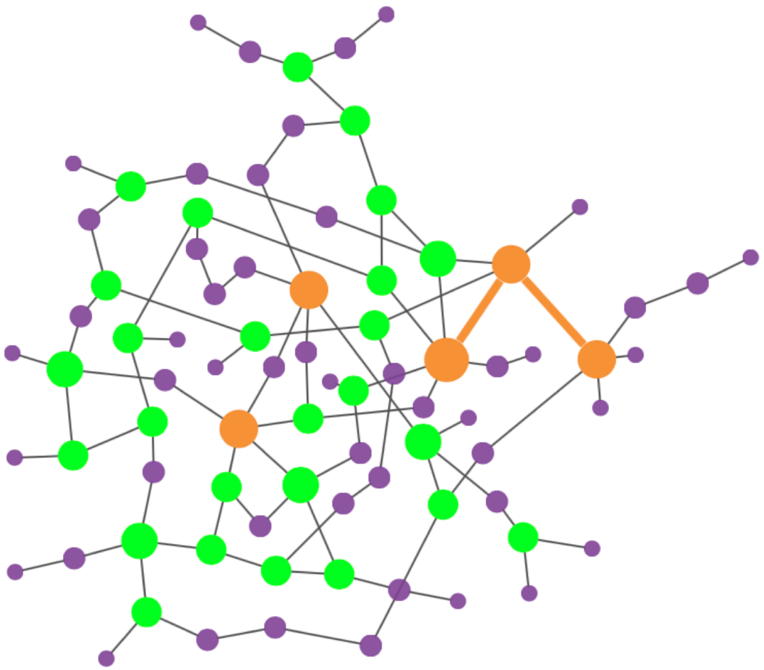
i.e., we expect social networks to be assortative

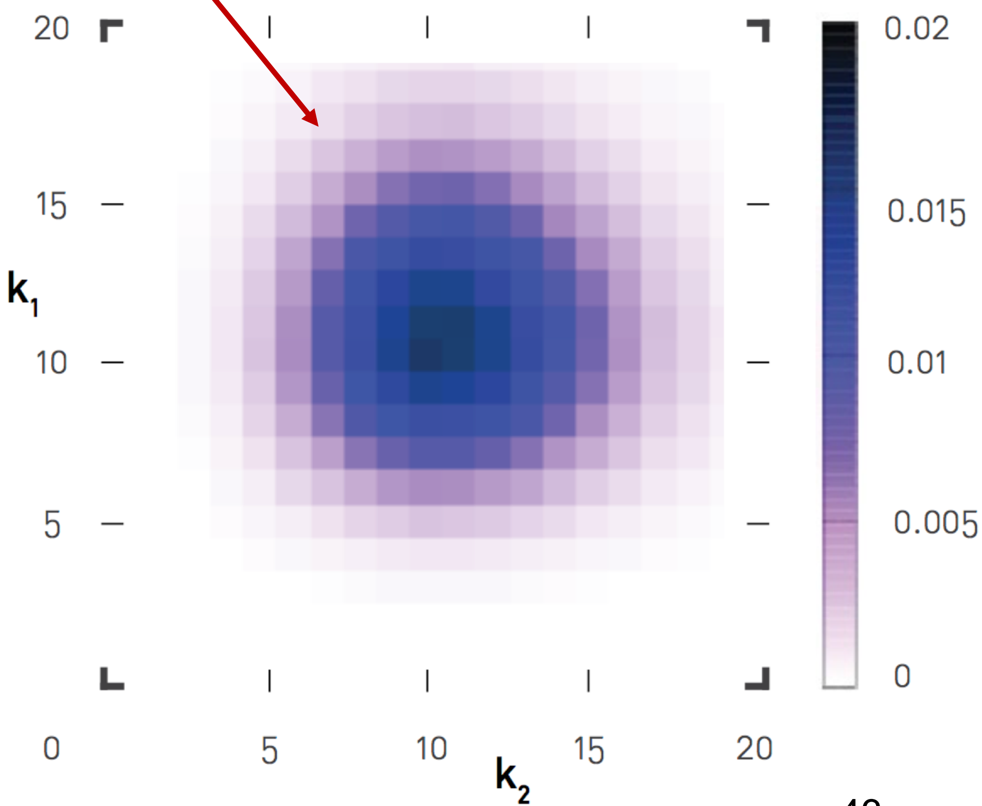
Neutral networks

The degree correlation is visually centred around the average degree

in the neutral case we expect circular symmetry, i.e., independence

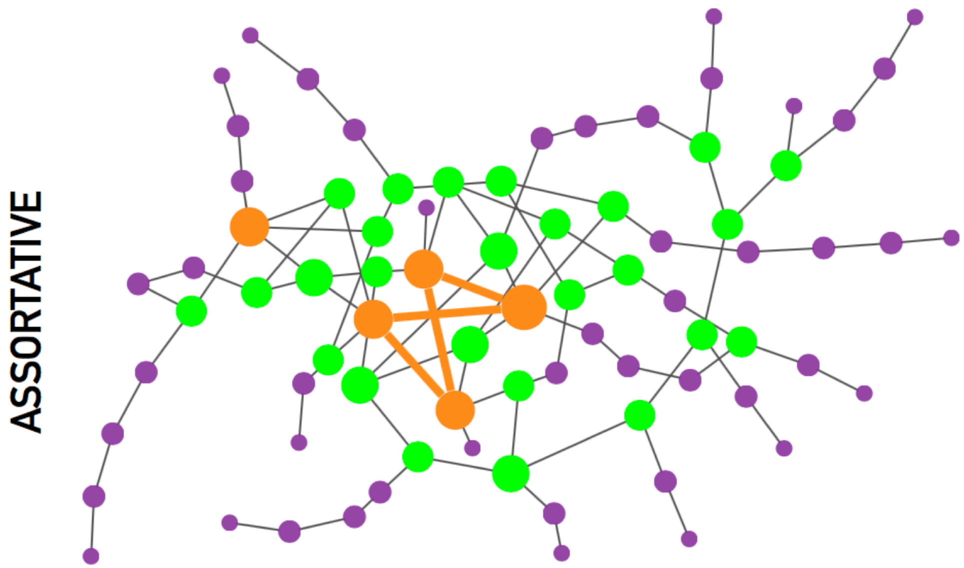
NEUTRAL

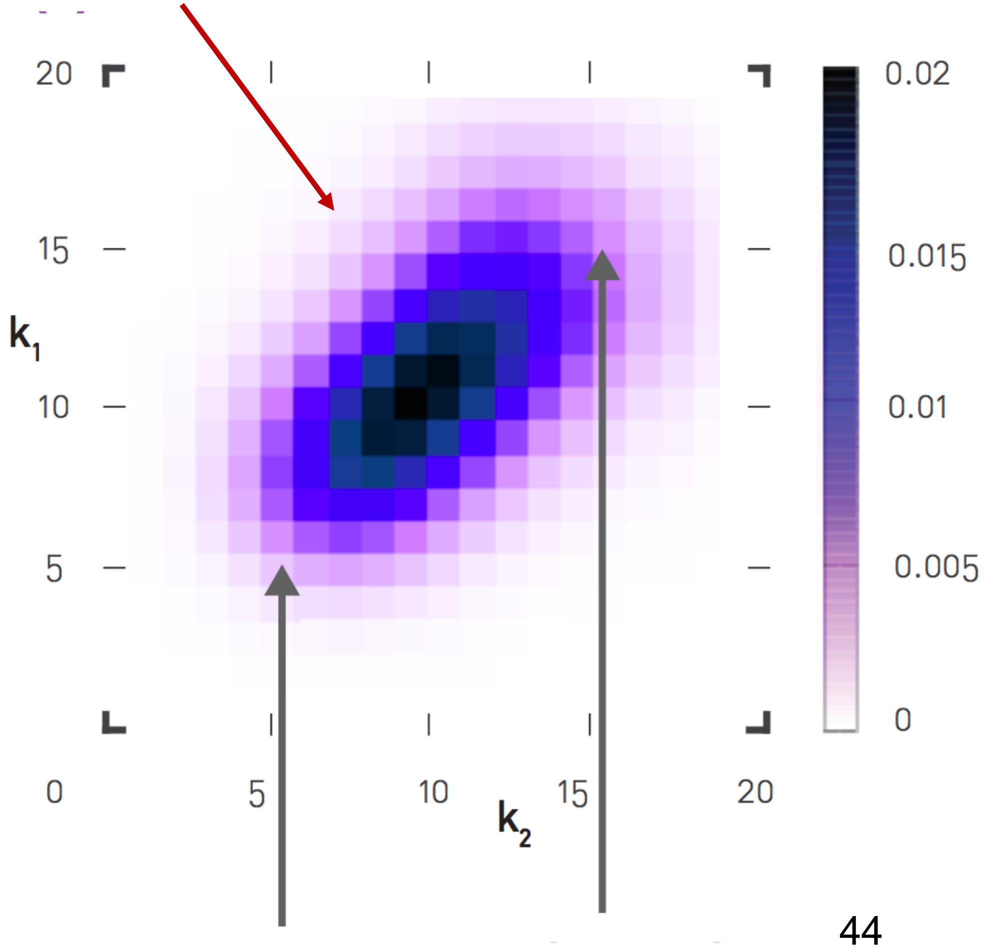




Assortative networks

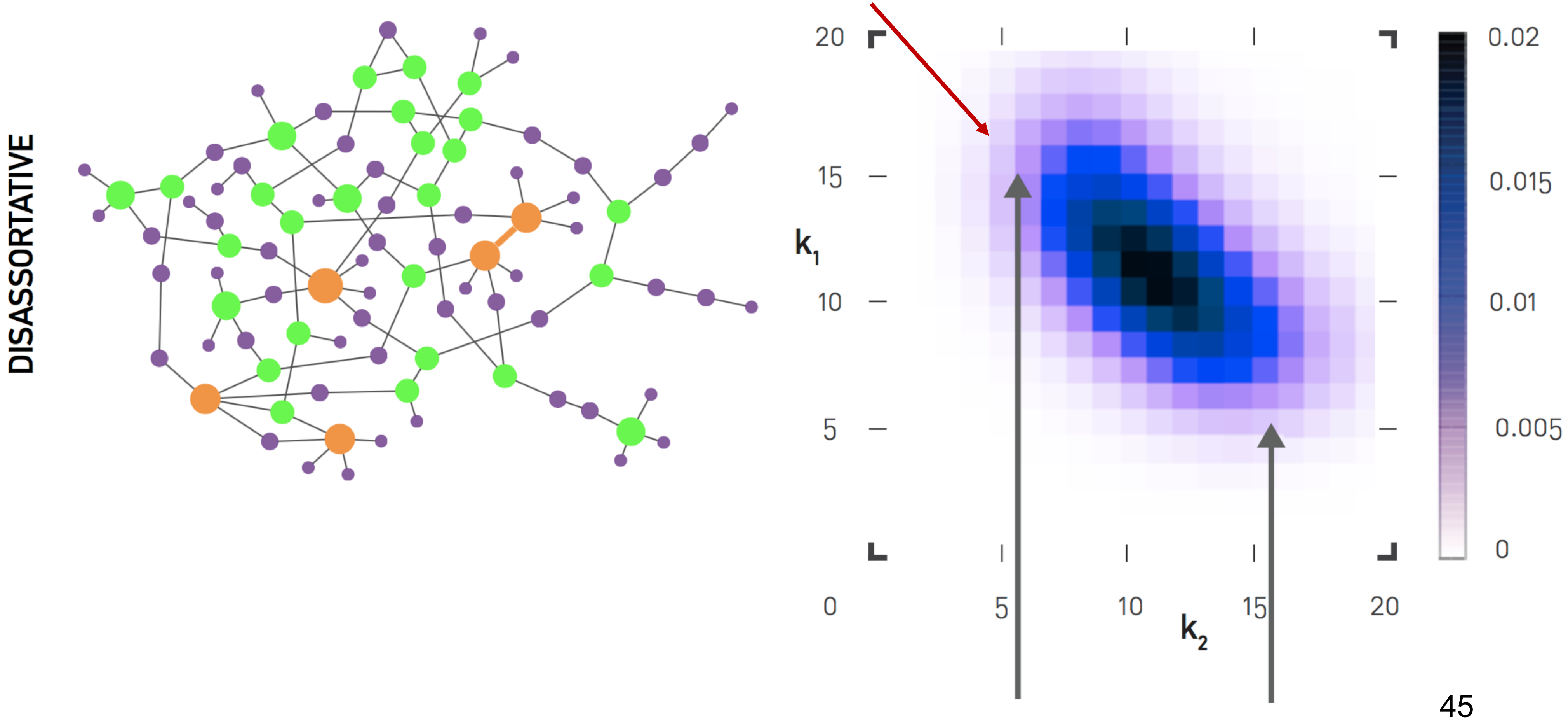
The degree correlation is turning to the right





Disassortative networks

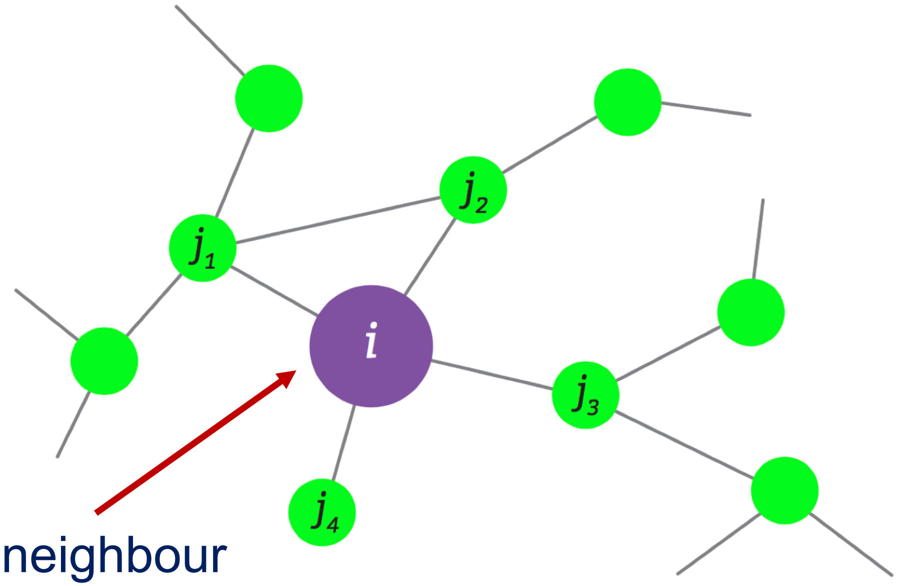
The degree correlation is turning to the left



Nearest neighbour degree

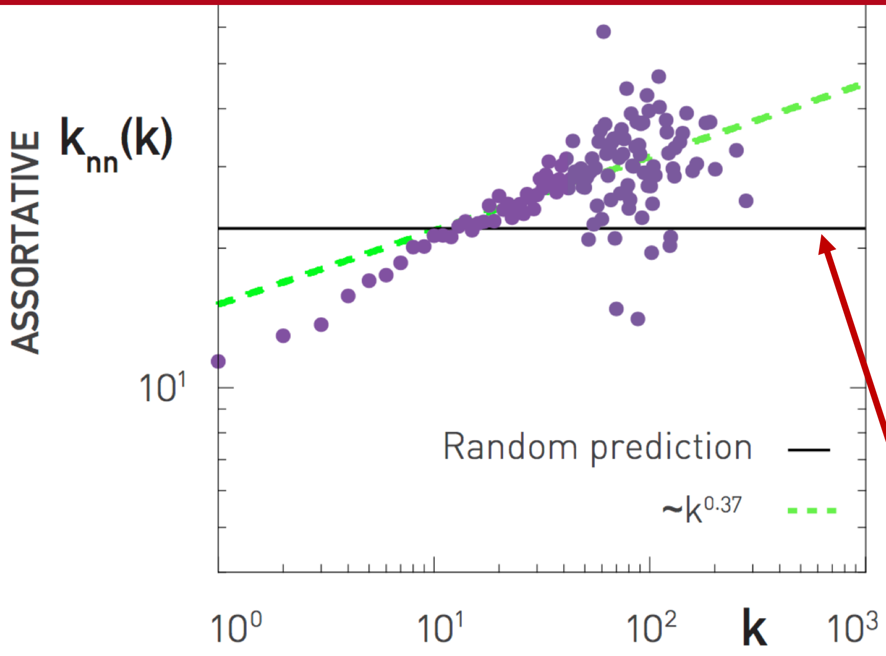
how to simplify plots from 2D to 1D

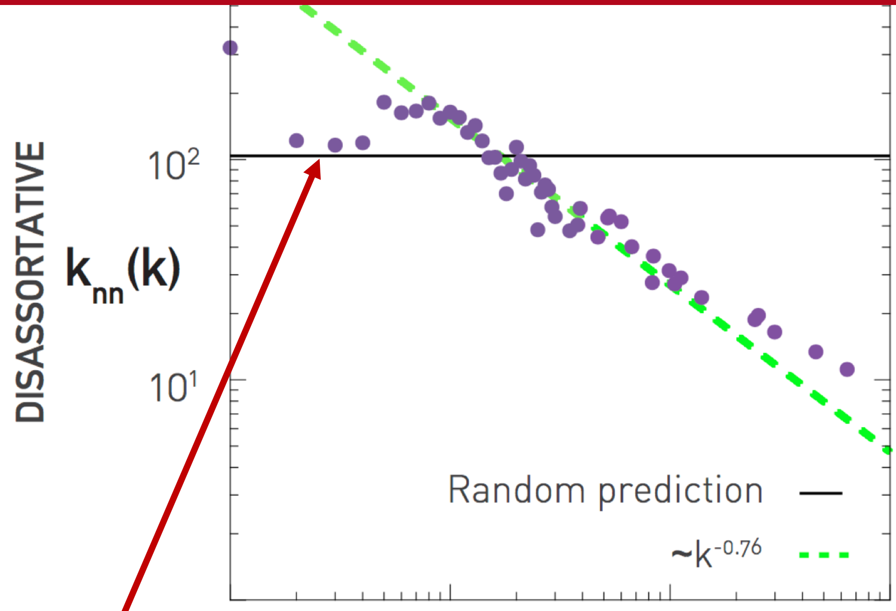
- Idea : inspect the degrees of the neighbouring nodes (easier than matrices)



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

Nearest neighbour degree plots





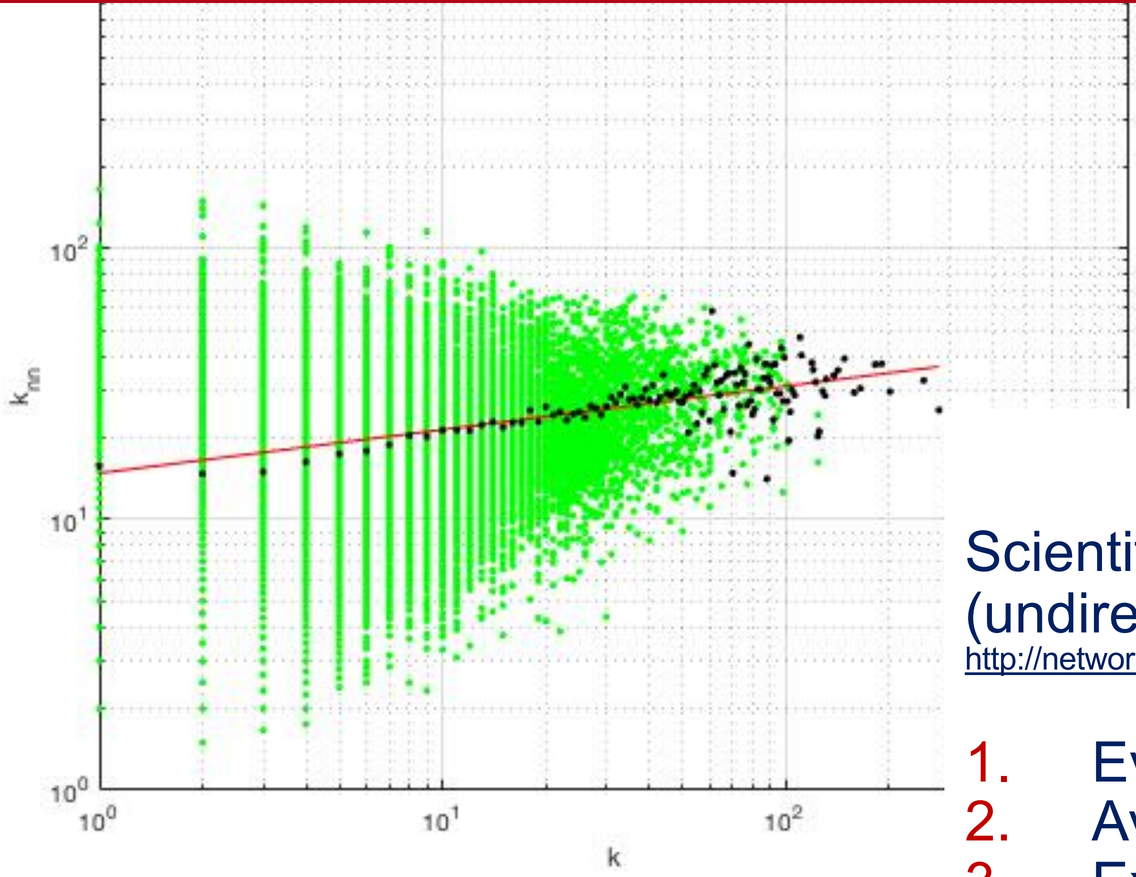
constant = independent of the degree (i.e., random = neutral)

$\ln(k_{nn}) = \mu \ln(k_i)$ → $\mu > 0$ = assortative

$\mu < 0$ = disassortative

A visual example

scientific collaboration network

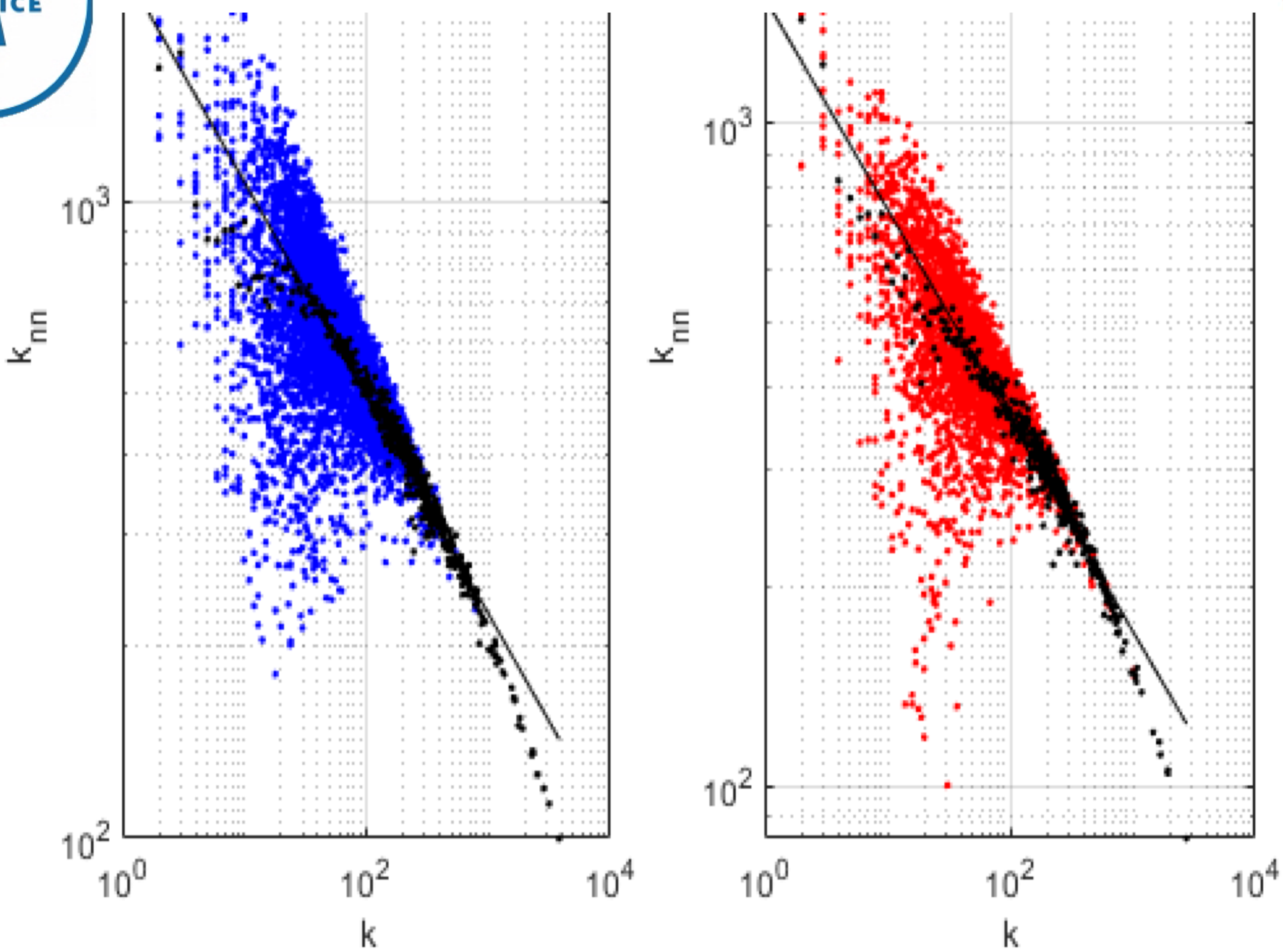


Scientific collaboration network (undirected, assortative)
http://networksciencebook.com/translations/en/resources/data.html

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

Hashtag network disassortativity

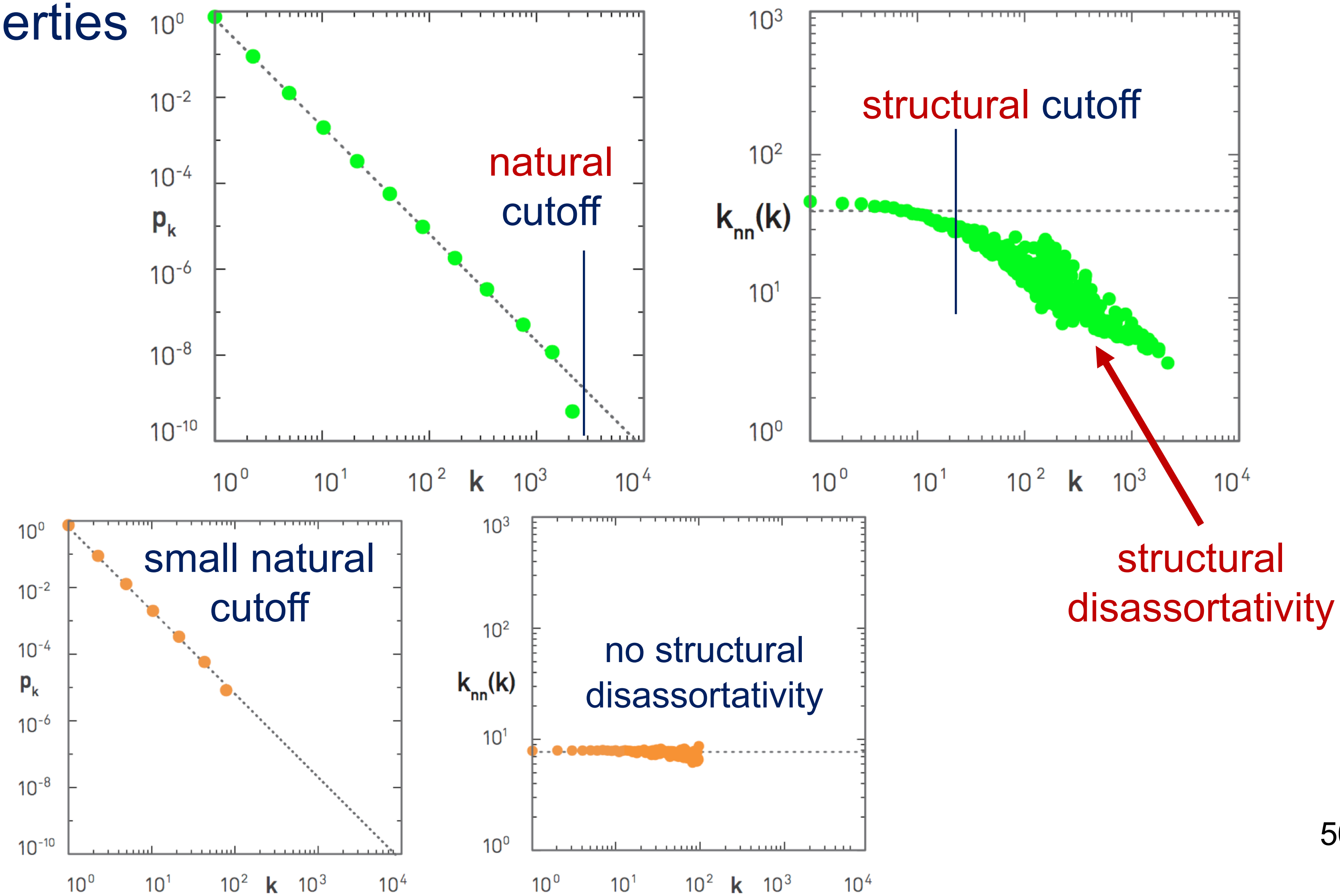
on pro-life/pro-choice data



Structural disassortativity

large degrees cannot be supported by a neutral network

(dis)Assortativity can be linked to structural network properties

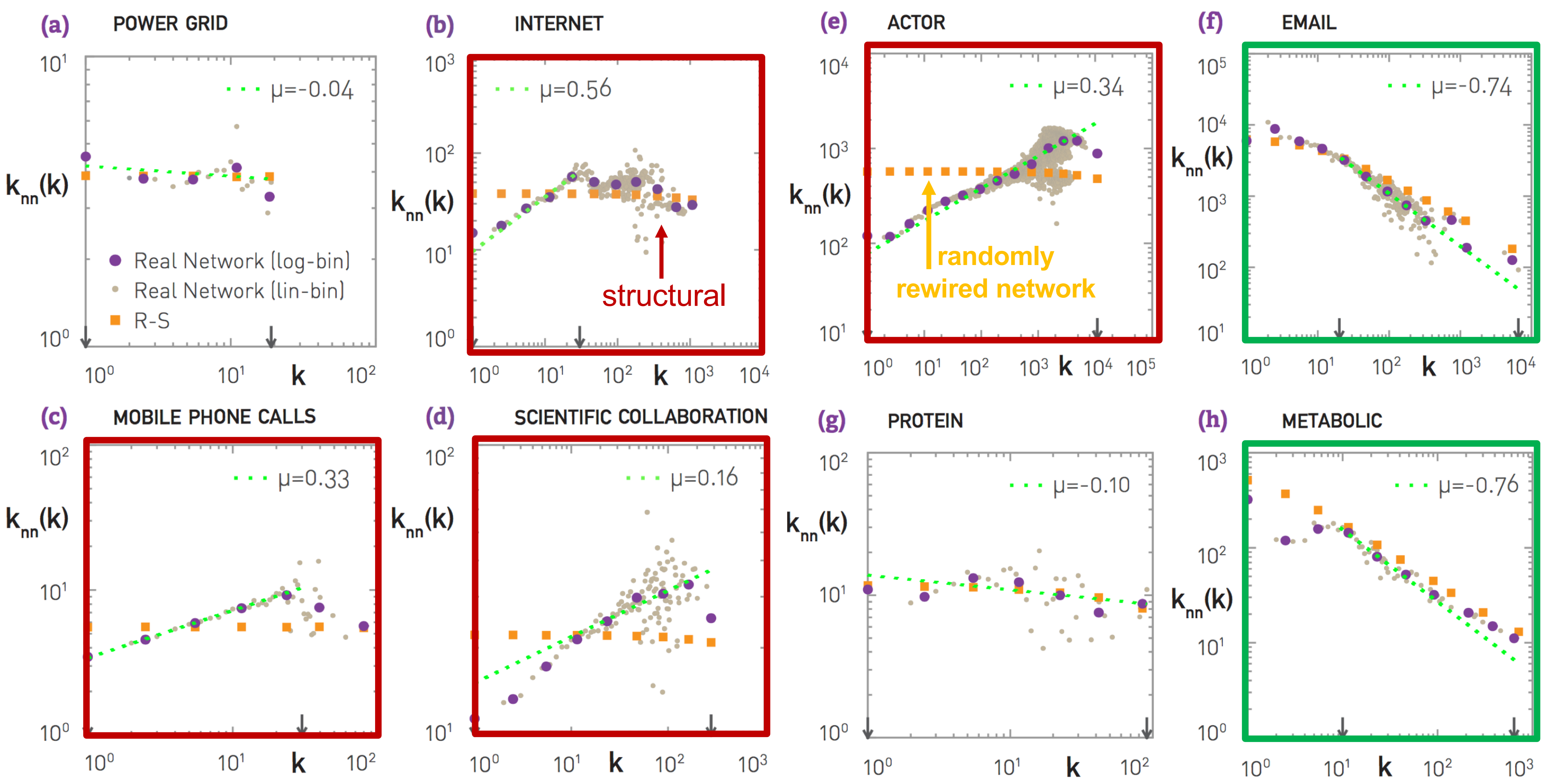


Structural disassortativity in real networks

social networks are assortative, most with a structural cutoff

assortative in red

disassortative in green

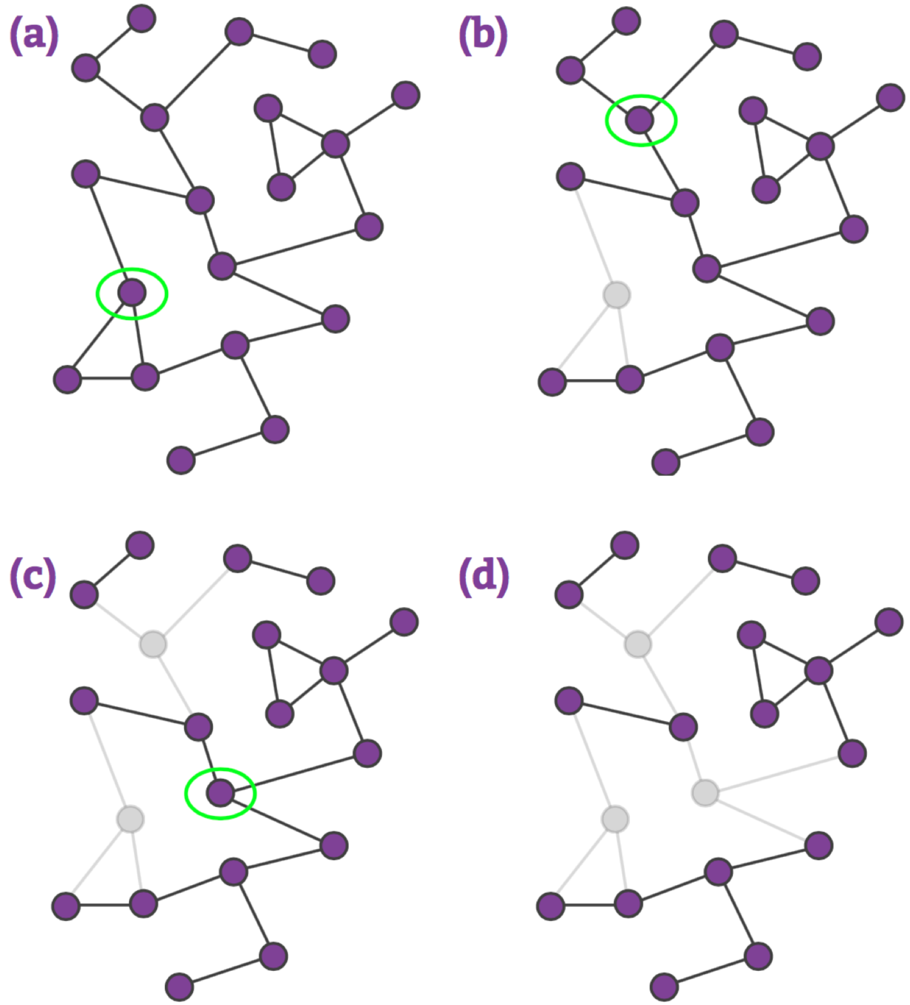


Robustness

of networks to failures

Network robustness

- Would the network still "work" in the presence of missing nodes?
- Failures can lead to either just isolating nodes or breaking the whole network apart
- What is the limit/phase transition?



Applications

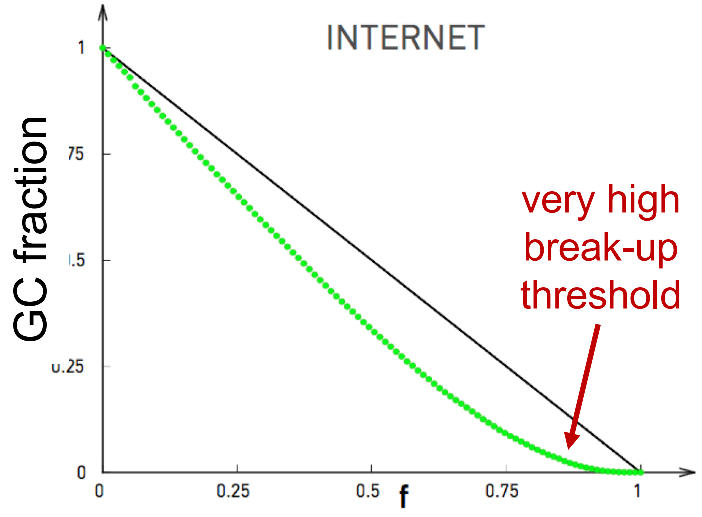
This can serve to identify:

- robustness of air transportation under random strikes
- robustness of social contacts even when someone is off
- possibility of destroying of criminal/terror networks
- eradication of an epidemics
- etc.

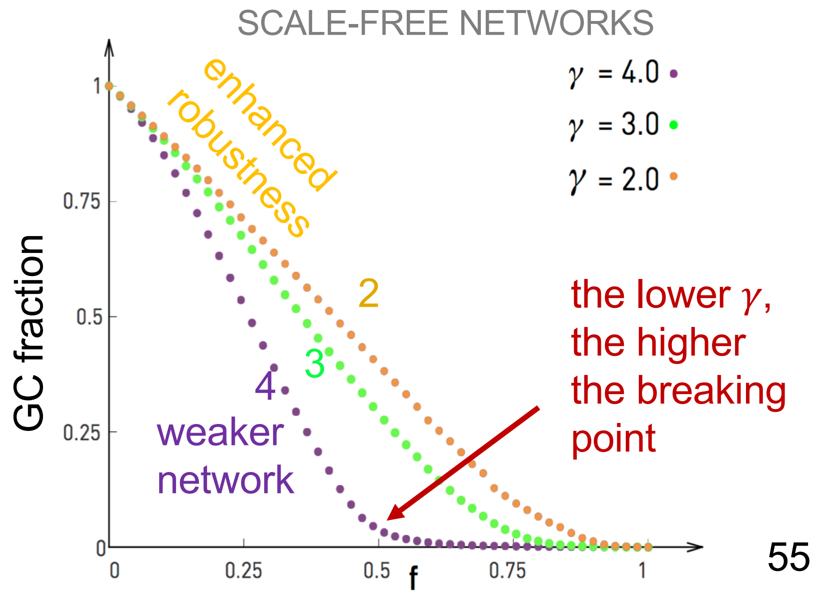
UNIVERSITÀ DEGLI STUDI DI PADOVA

Robustness of scale-free networks
under random node removal

- Robustness of the Internet due to scale-free properties
- Nodes linked to the GC after random removal with rate *f* → still large if *f*<*1*



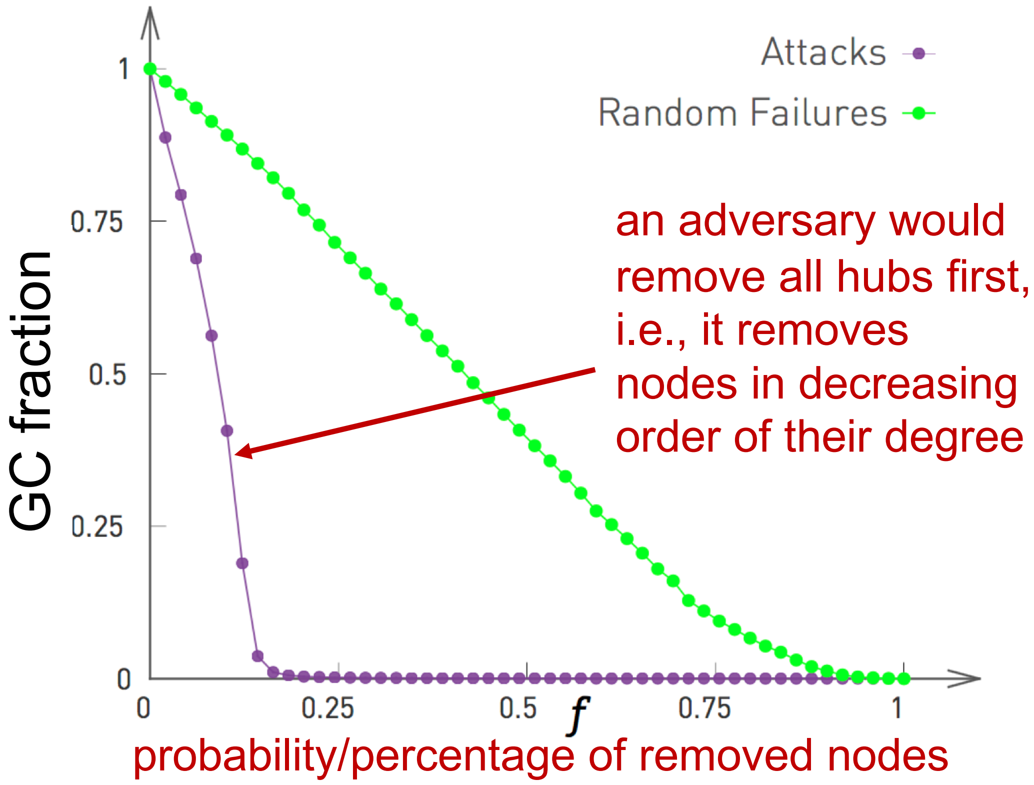
- Experiments aligned with a scale-free model
- Reason: random removal of (many) hubs is very unlikely



Attack tolerance

What if removals are not by chance, but caused by an adversary with sufficient insights on our network?

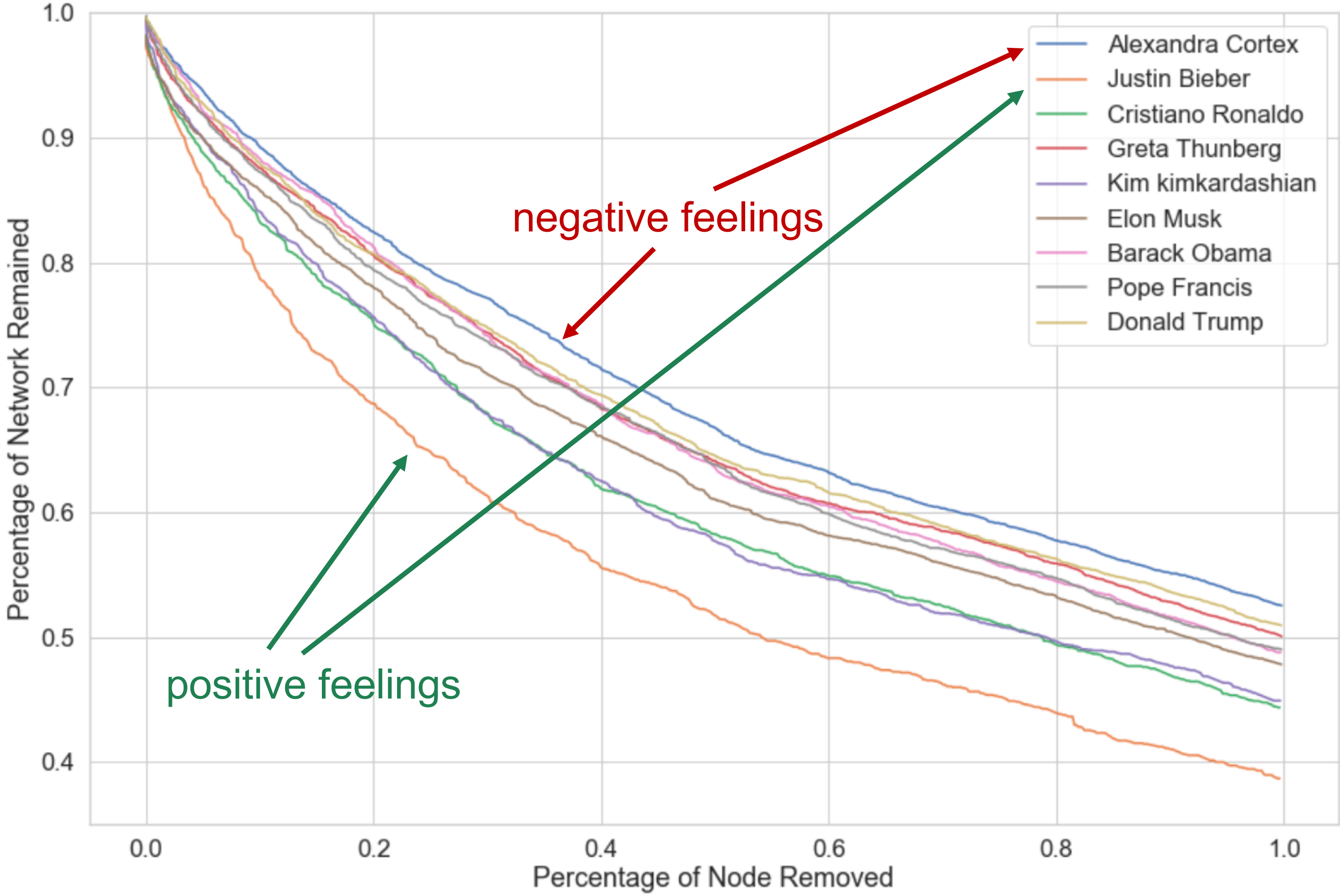
- Scale-free networks are not very robust to targeted attacks exactly because they have vulnerable hubs
- good news in medicine (vulnerability of bacteria) ☺
- bad news for the Internet ☹



Example

network analysis of Tweets' sentiment IP (2019)

robustness of original network to positive node removal



Optimizing robustness

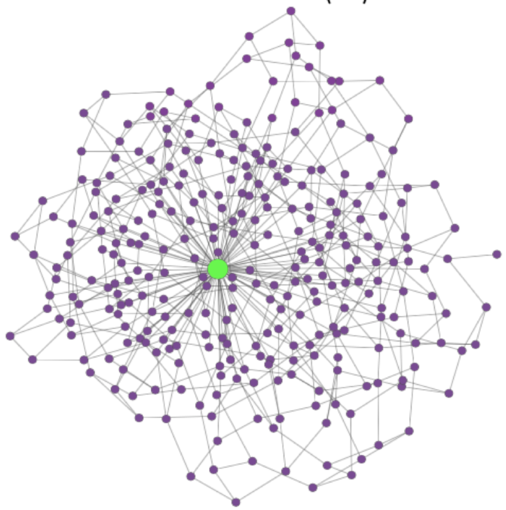
is not an option in real-world networks

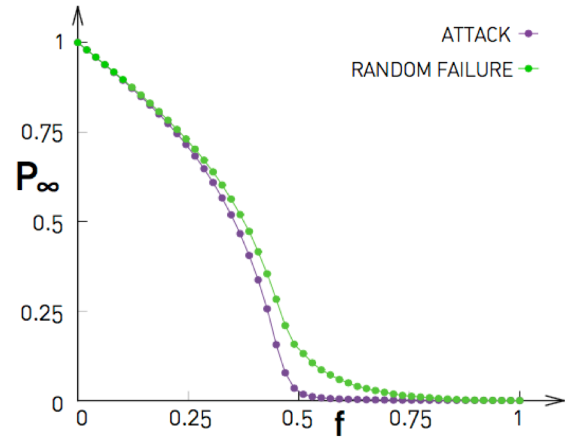
The best option is a bimodal distribution

$$p_k = r\,\delta_{k\text{max}} + (1\text{-}r)\,\delta_{k\text{min}}$$

r = 1/N
k_{max} chosen to maximize the breakpoints

$\langle k \rangle = 3$





$\langle k \rangle = 5$