Friendly Graph Theory

Introduction to Network Science

Dr. Michele Starnini — https://github.com/chatox/networks-science-course

Universitat Pompeu Fabra Barcelona

Contents:

- Degree correlations
- Clustering
- Homophily

all related to friendship in social networks!

Degree correlations

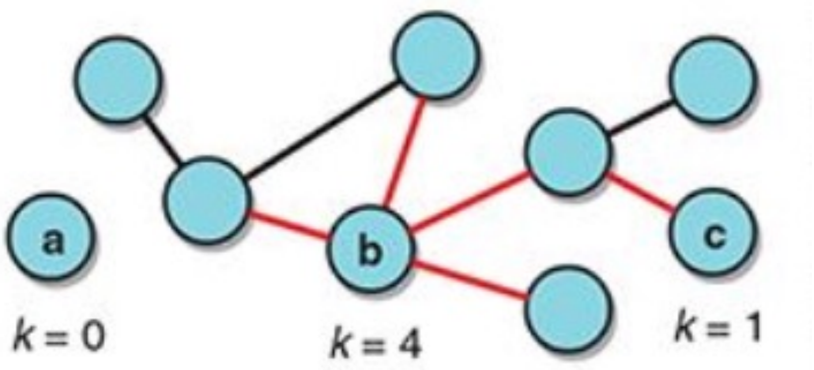
Who is a friend? [Assortativity]

- Degree is the main feature of nodes
- In social networks, degree correlations can determine connections: **assortativity**
- Example: very famous people (with millions of followers) follow each other

Degrees

Node i has degree k_i:
number of links incident on this node

High-degree nodes are called **hubs, k >>> 1**



Prob. that a random node
has degree k (k-node)

$$p(k) \quad \propto \quad N_k$$

\# of k-nodes

Normalization

$$\sum_k N_k = N$$

Prob. of finding a k-node

$$p(k) = \frac{N_k}{N}$$

Average degree

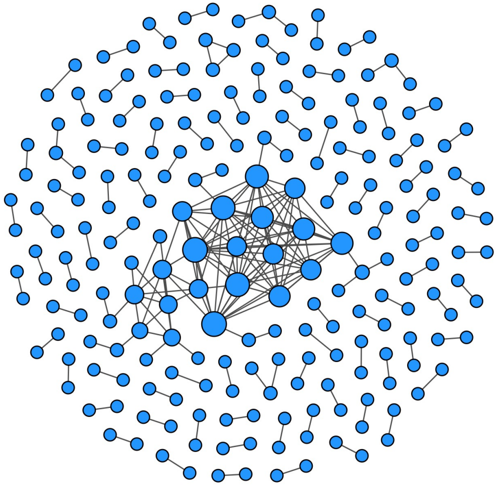
$$\langle k \rangle = \frac{\sum_k kp(k)}{N}$$

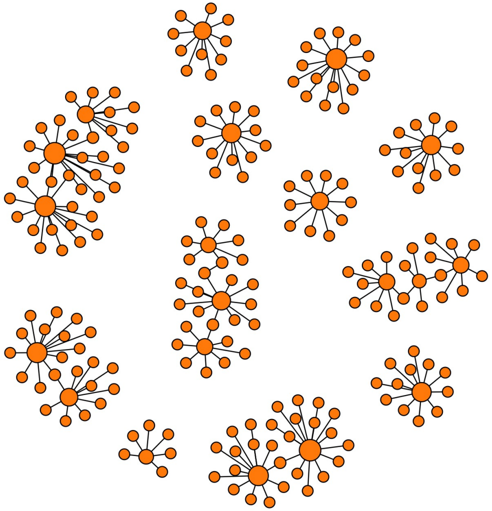
Degree assortativity

A.k.a. **degree correlation:**

- Assortative networks have a **core-periphery** structure with hubs in the core
  (Ex: social networks)

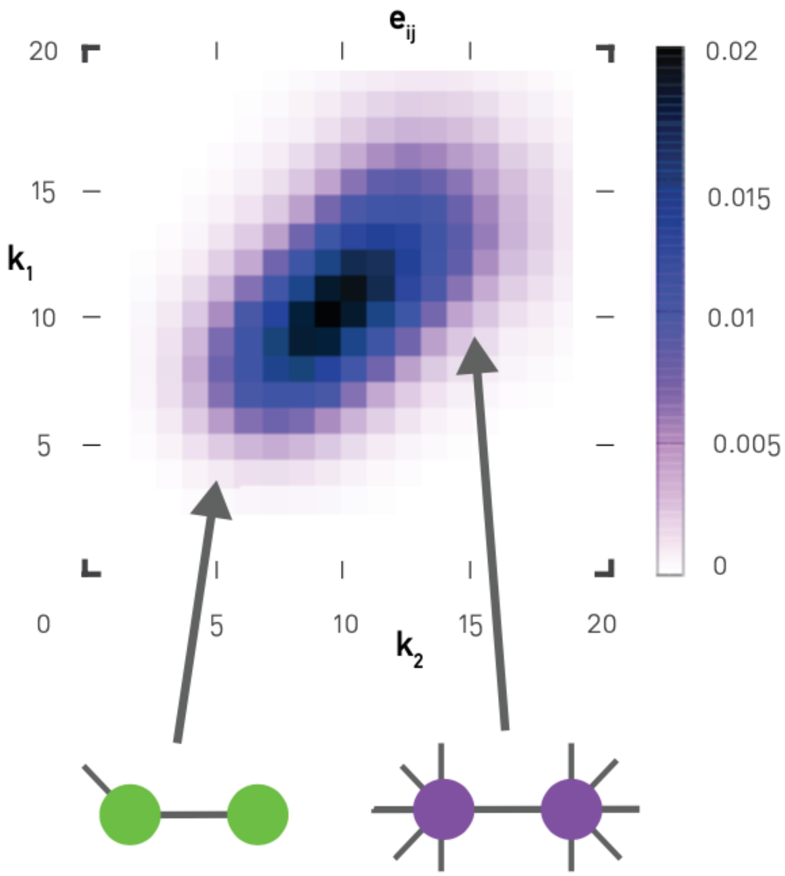
- Disassortative networks have **hub-and-spoke** (or **star**) structure
  (Ex: Web, Internet, food webs, bio networks)





Assortative networks

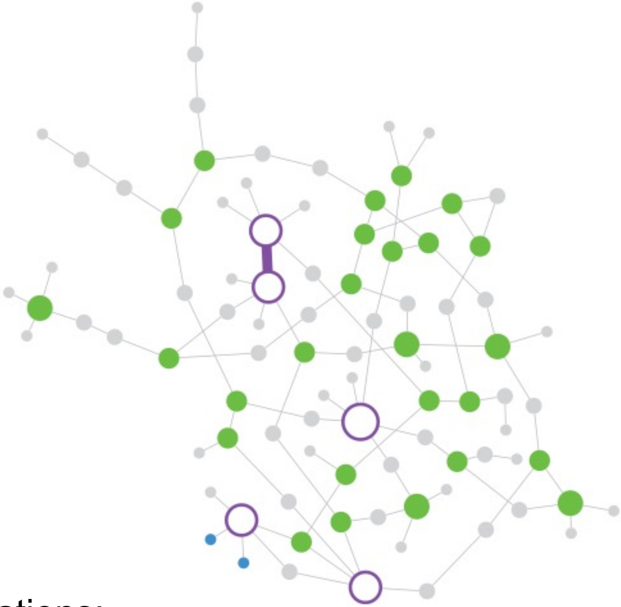
$E_{k,k'}$ = #links between k-nodes & k'-nodes

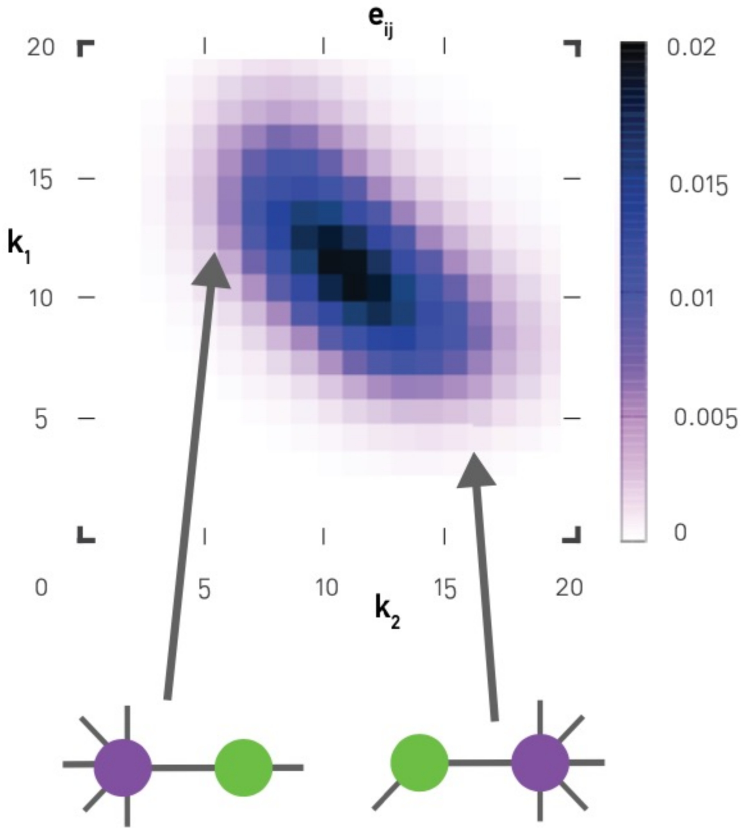


Positive degree correlations:
small-degree nodes to small-degree nodes,
hubs to hubs

$E_{k,k'}$ = #links between k-nodes & k'-nodes

Dis-assortative networks





Negative degree correlations:
small-degree nodes to hubs

Degree correlations

$E_{k,k'}$ = #links between k-nodes & k'-nodes

$$\sum_{k,k'} E_{k,k'} = N\langle k \rangle = 2E$$

Sum over all nodes twice

Joint prob. that a random link is connected to a k-node & a k'-node

$$p(k',k) \propto E_{k,k'}$$

Normalization

$$\sum_{k,k'} p(k',k) = 1$$

Joint prob. that a random link connects a k-node & a k'-node

$$p(k,k') = \frac{E_{k,k'}}{\sum_{k,k'} E_{k,k'}} = \frac{E_{k,k'}}{N\langle k \rangle}$$

Degree correlations

$E_{k,k'}$ = #links between k-nodes & k'-nodes

$$\sum_{k,k'} E_{k,k'} = N\langle k\rangle = 2E$$

Sum over all nodes twice

Prob. that a random link is connected to a k-node

Total # links from k-nodes

$$q_k \propto \sum_{k'} E_{k,k'} = kN_k$$

Normalization

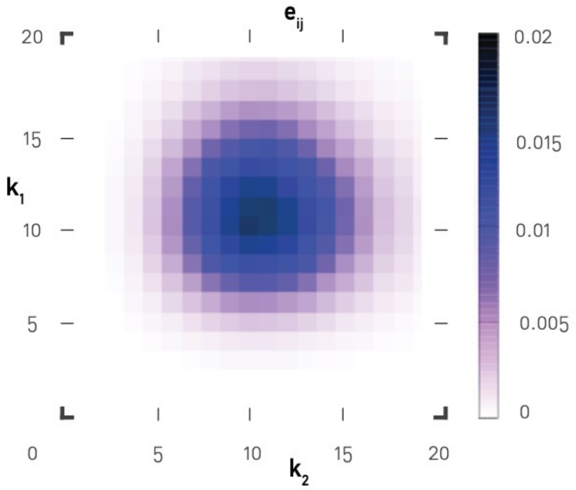
$$\sum_k q_k = 1$$

Prob. that a random link is connected to a k-node

$$q_k = \sum_{k'} p(k',k) = \frac{\sum_{k'} E_{k',k}}{N\langle k\rangle} = \frac{kp(k)}{\langle k\rangle}$$

No degree correlations

No correlations:
degree of a node is **independent** from the degree of others



Joint prob. that a random link is connected to a k-node & a k'-node,
If no degree corralations

$$p_0(k', k) \propto E^0_{k,k'} = kN_k \times k'N_{k'}$$

Total # links from k-nodes

Total # links from k'-nodes

Joint prob. that a random link connects a k-node & a k'-node in uncorrelated networks

$$p_0(k', k) = q_k \times q'_k = \frac{kp(k)k'p(k')}{\langle k \rangle^2}$$

$$q_k = \frac{kp(k)}{\langle k \rangle}$$

Prob. that a random link is connected to a k-node

Measuring degree correlations

Conditional prob. that a random link, **already connected to a k-node,** is also connected to a k'-node $=$ $\dfrac{\text{Joint prob. that a random link is connected to a k-node \& a k'-node}}{\text{Prob. that a random link is connected to a k-node}}$

$$p(k'|k) = \frac{p(k',k)}{\sum_{k'} p(k',k)}$$

Remember:

$$p(k',k) = \frac{E_{k,k'}}{N\langle k\rangle}$$

$$\sum_{k'} p(k',k) = q_k = \frac{kp(k)}{N\langle k\rangle}$$

$\Rightarrow$

Conditional prob. that a random link is connected to a k'-node, given that it is connected to a k-node

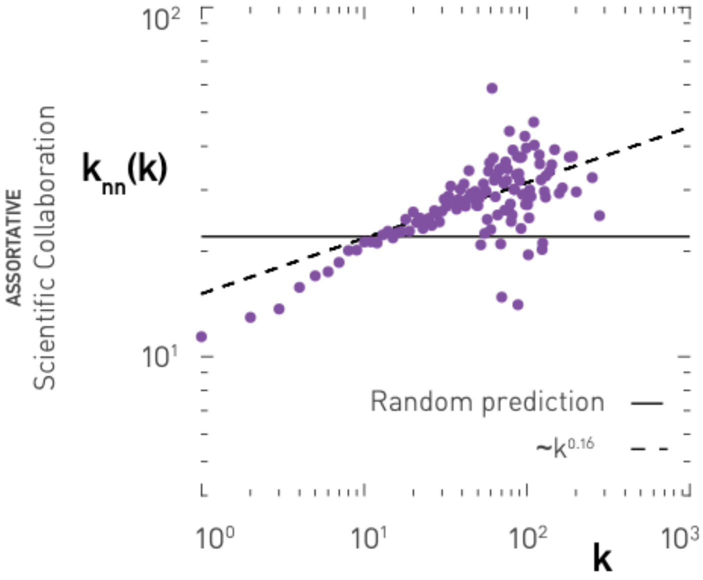
$$p(k'|k) = \frac{E_{k,k'}}{kp(k)}$$

Measuring degree correlations

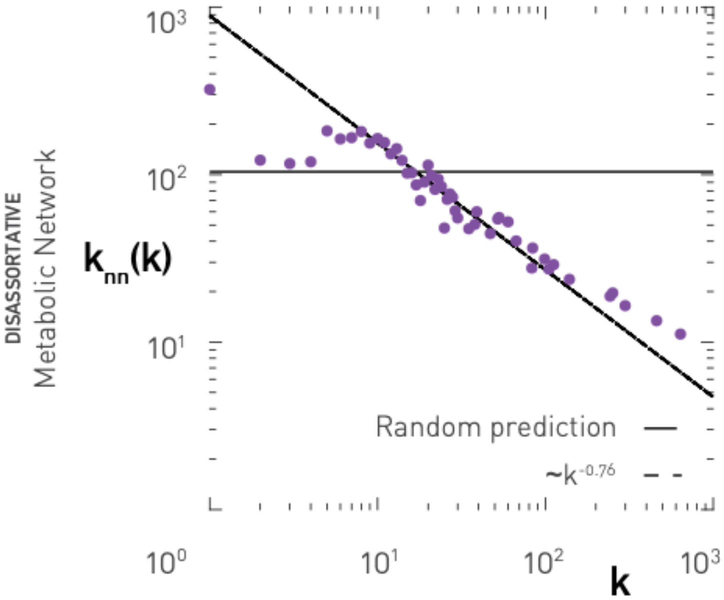
Average degree of the nearest neighbors (NN) of nodes with degree k:

$$k_{NN}(k) \equiv \sum_{k'} k' p(k'|k)$$

Assortative networks: Increasing $k_{NN}(k)$



Disassortative networks: Decraesing $k_{NN}(k)$



Exercise

Find $k_{NN}(k)$ if there are no degree correlations!

$$k_{NN}(k) \equiv \sum_{k'} k' p(k'|k)$$

$$p(k'|k) = \frac{p(k',k)}{\sum_{k'} p(k',k)}$$

$$p_0(k',k) = q_k \times q'_k = \frac{kp(k)k'p(k')}{\langle k \rangle^2}$$

Pin board: https://upfbarcelona.padlet.org/chato/v0apheshv2l4hbot

No degree correlations

$$k_{NN}(k) \equiv \sum_{k'} k' p(k'|k)$$

$$p(k'|k) = \frac{p(k',k)}{\sum_{k'} p(k',k)}$$

$$p_0(k',k) = q_k \times q_k' = \frac{kp(k)k'p(k')}{\langle k \rangle^2}$$

$$p_0(k'|k) = \frac{q_k q_{k'}}{q_k} = q_{k'}$$

Definition of no-correlations:
p_0(k'|k) is **independent** of k!!

$$k^0_{NN}(k) = \sum_{k'} k' p_0(k'|k) = \sum_{k'} k' q_{k'} = \frac{\sum_{k'} k'k'p(k')}{N\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

Friendship paradox

Average degree (number of friends)
of a randomly chosen node (you): **<k>**

<

Average degree (number of friends)
of a k-node (your friend with k friends): $\mathbf{k_{NN}(k)}$

If no degree correlations:
k_{NN}(k) = <k^2> / <k>

$$\langle k \rangle = \sum_k k p(k)$$

average of
degree distribution

$$\langle k^2 \rangle = \sum_k k^2 p(k)$$

$$\sigma^2 = \langle k^2 \rangle > \langle k \rangle^2 > 0$$

Variance of
degree distribution

$$\langle k^2 \rangle > \langle k \rangle^2$$

Your friends' average
number of friends

Your average
number of friends

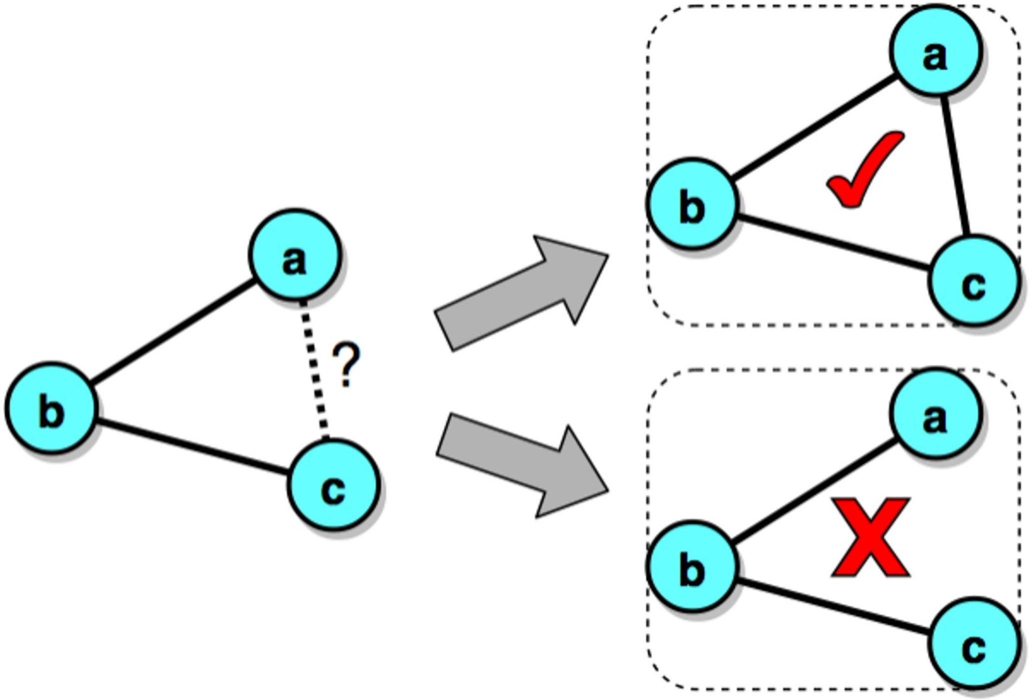
$$\frac{\langle k^2 \rangle}{\langle k \rangle} > \langle k \rangle$$

Your friends have more friends than you have!

Clustering

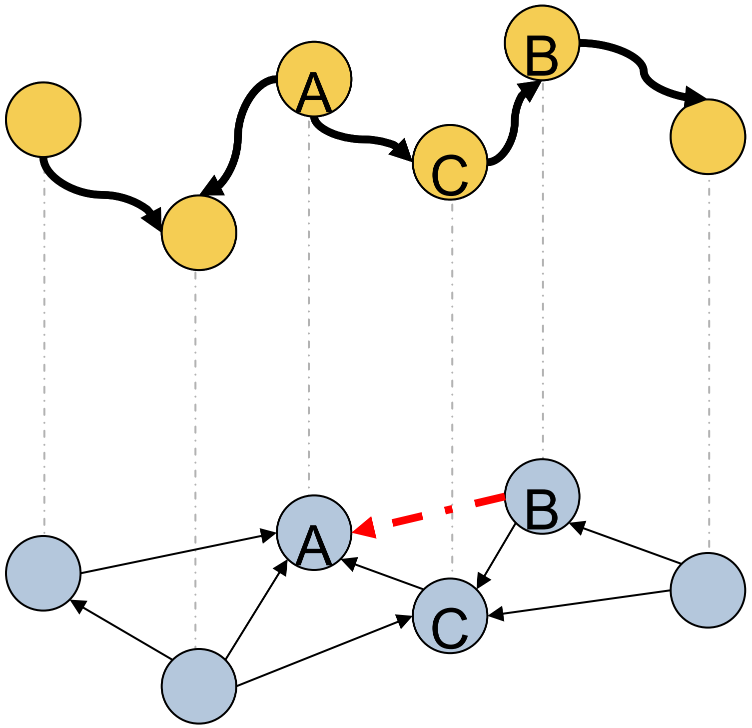
Who is a friend? [Triangle closure]

A prevalent way in which we form friendships is by befriending **friends of friends**



Source: Menczer, Fortunato, David: [A First Course on Networks Science](). Cambridge, 2020.

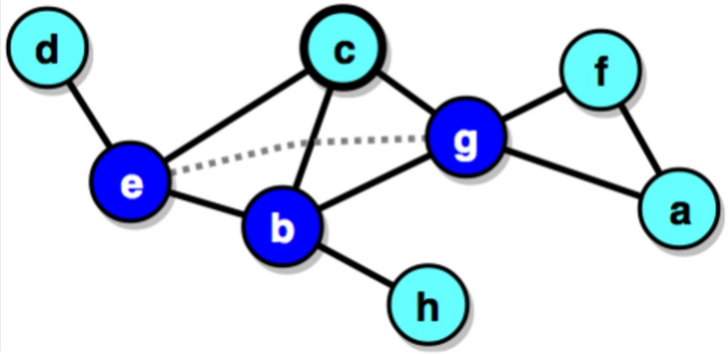
Tendency to form triangles



- The dynamics **on** the network, i.e., information diffusion, affect the dynamics **of** the network, i.e., the creation of links
- *B* is more likely to start following *A* after seeing content posted by *A* and re-posted by an account *C* that *B* already follows

https://doi.org/10.1145/2487575.2487607

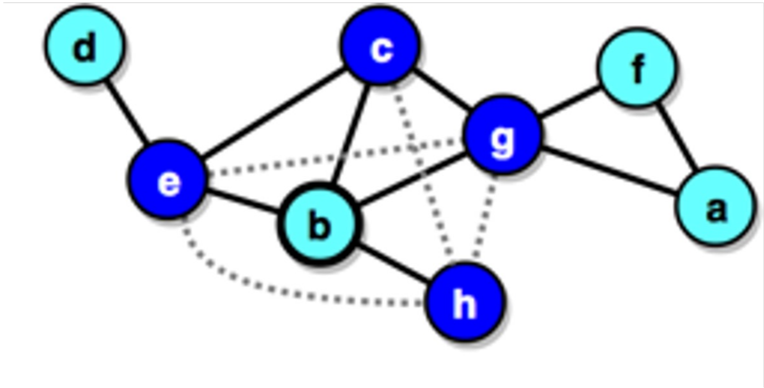
Example 1



Node c has 3 neighbors: e, b, g
They form two triangles out of the possible 3 (the missing one is drawn with a dotted line)

Source: Menczer, Fortunato, David: [A First Course on Networks Science](). Cambridge, 2020.

Example 2



Node b has 4 neighbors: e, c, g, h
They form two triangles out of the possible 6 (the missing ones are drawn with a dotted line)

Source: Menczer, Fortunato, David: [A First Course on Networks Science](). Cambridge, 2020.

Remember

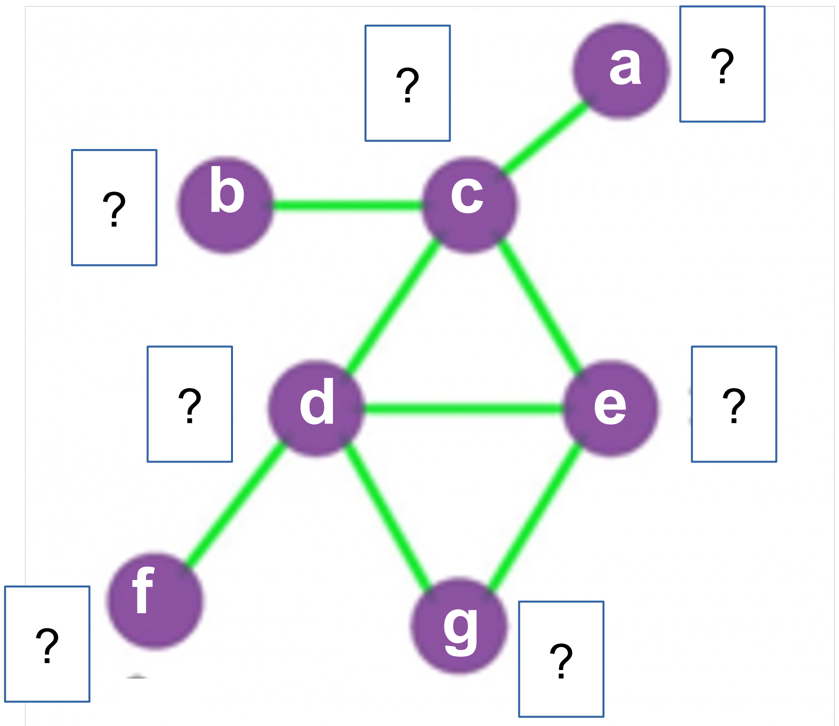
- The maximum number of links between *k* nodes is

$$\frac{k(k-1)}{2}$$

Local clustering coefficient

- The **local clustering coefficient** C_i is a property of a node i
- Let L_i represent the number of links among neighbors of node i

$$C_i = \frac{L_i}{\frac{k_i(k_i-1)}{2}} = \frac{2L_i}{k_i(k_i-1)} \quad C_i \triangleq 0 \text{ if } k_i \leq 1$$



Exercise

What is the local clustering coefficient of each node?

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

$$C_i \triangleq 0 \text{ if } k_i \leq 1$$

Pin board: https://upfbarcelona.padlet.org/chato/v0apheshv2l4hbot

Average clustering coefficient
("global clustering coefficient")

The **average clustering coefficient** is a property of the entire graph

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i$$

Sometimes this is called the *curvature* of a graph

Network clustering coefficient

- Social networks tend to have high clustering coefficients because of **triadic closure**: we meet through common friends
- Other networks, e.g., bipartite and tree-like networks, have low clustering coefficient

Table 2.1 Average path length and clustering coefficient of various network examples. The networks are the same as in Table 1.1, their numbers of nodes and links are listed as well. Link weights are ignored. The average path length is measured only on the giant component; for directed networks we consider directed paths in the giant strongly connected component. To measure the clustering coefficient in directed networks, we ignore link directions.

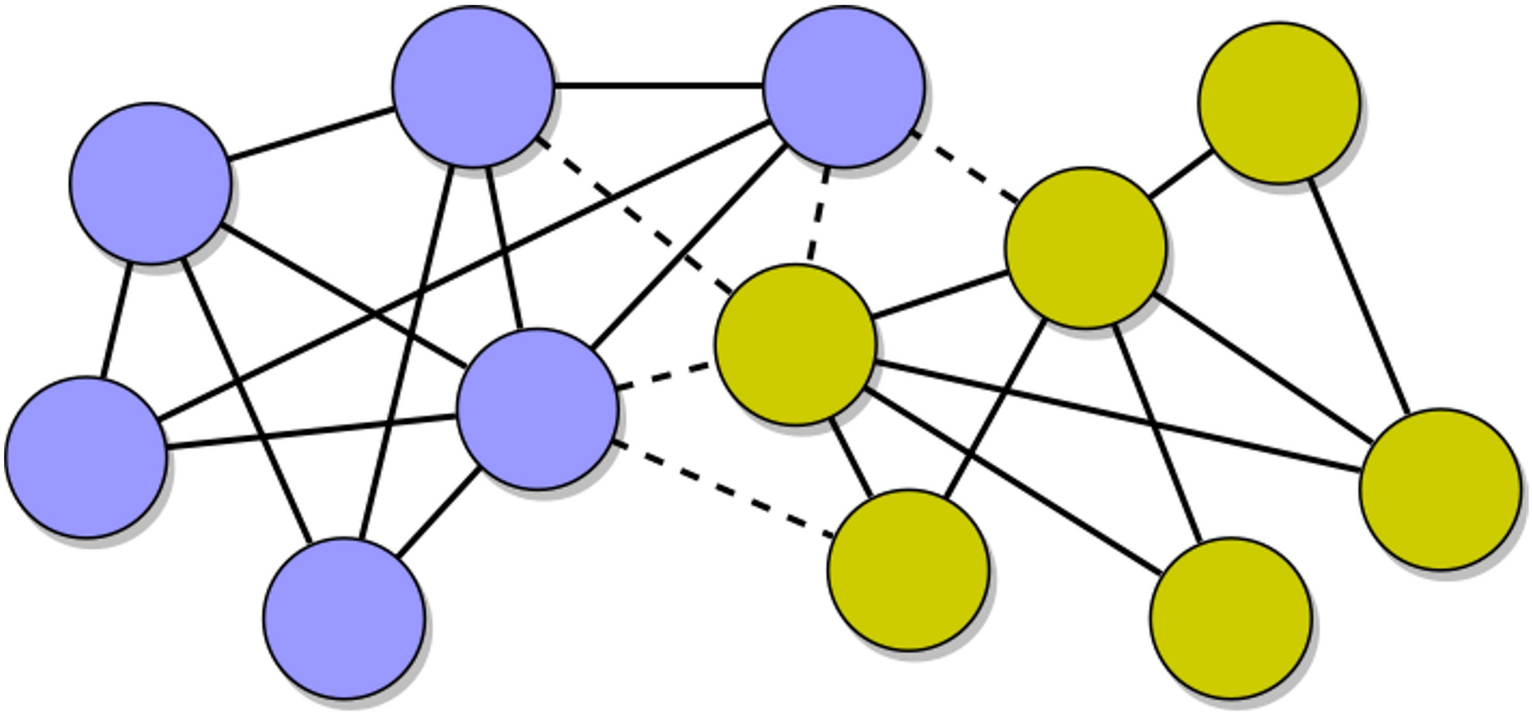
Network	Nodes (N)	Links (L)	Average path length ($\langle \ell \rangle$)	Clustering coefficient (C)
Facebook Northwestern Univ.	10,567	488,337	2.7	0.24
IMDB movies and stars	563,443	921,160	12.1	0
IMDB co-stars	252,999	1,015,187	6.8	0.67
Twitter US politics	18,470	48,365	5.6	0.03
Enron Email	87,273	321,918	3.6	0.12
Wikipedia math	15,220	194,103	3.9	0.31
Internet routers	190,914	607,610	7.0	0.16
US air transportation	546	2,781	3.2	0.49
World air transportation	3,179	18,617	4.0	0.49
Yeast protein interactions	1,870	2,277	6.8	0.07
C. elegans brain	297	2,345	4.0	0.29
Everglades ecological food web	69	916	2.2	0.55

Homophily

Who is a friend? [Homophily]

- In social networks, nodes have **features** that influence their connectivity preferences
  - Age, gender identity, ethnicity, sexual preference, location, topics of interest, artistic sensitivities, …
- People tend to befriend those who are like them: that is called **homophily**

“Birds of a feather flock together”

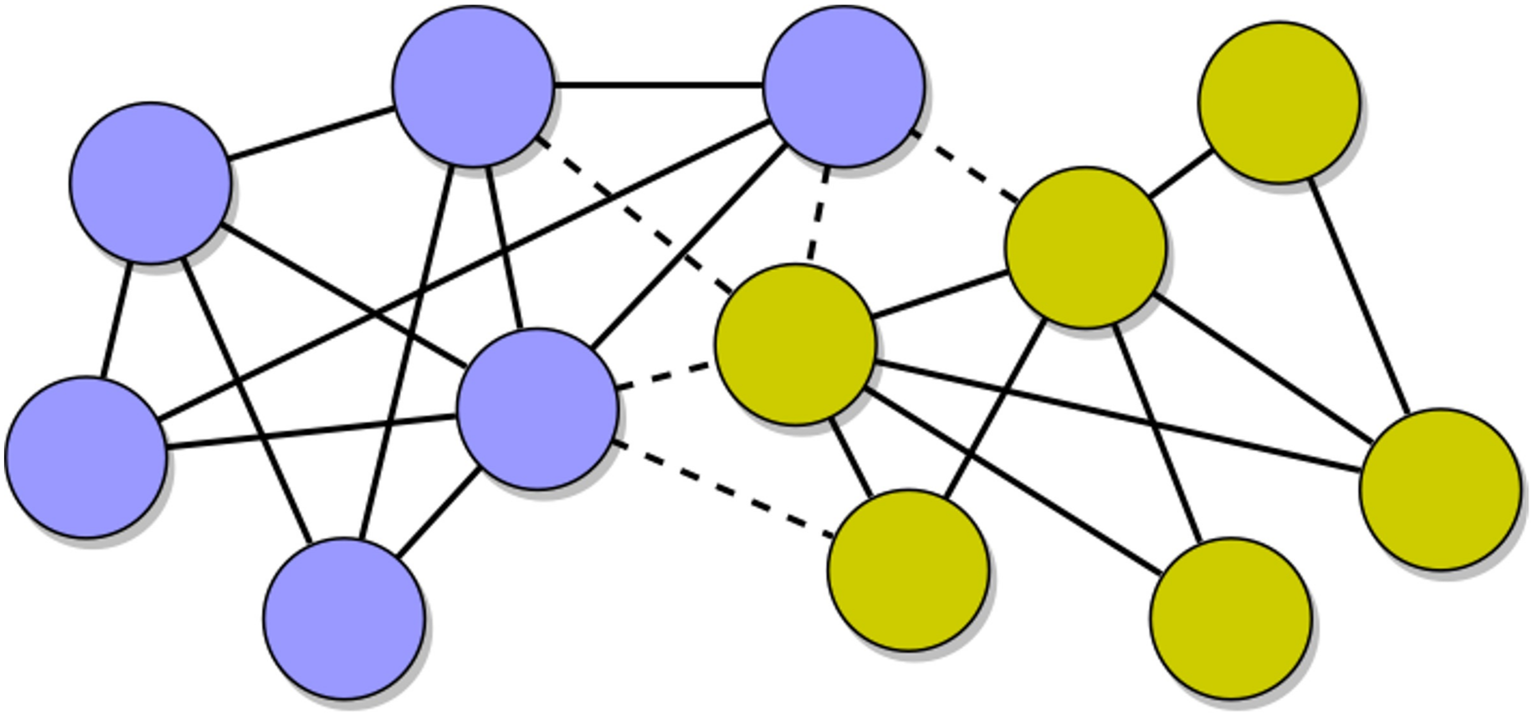


Source: Menczer, Fortunato, David: [A First Course on Networks Science](). Cambridge, 2020.

Quantifying homophily

- Let G be a graph of N nodes: N_a “yellow” and N_b “blue”
  - $N = N_a + N_b$
- Let G have L undirected links (including self loops), of which L_{aa} connect yellow to yellow, L_{ab} connect yellow to blue, and L_{bb} connect blue to blue
  - $L = L_{aa} + L_{ab} + L_{bb}$ $L_a = L_{aa} + L_{ab}$ $L_b = L_{bb} + L_{ab}$

$N_a = 6, N_b = 6, L_a = 14, L_b = 16, L_{ab} = 5,$



Source: Menczer, Fortunato, David: A First Course on Networks Science. Cambridge, 2020.

Expected links across groups

If yellow nodes have L_a links placed at random (incl. self loops), how many should go to a blue node?

$$L_a \left(\frac{N_b}{N} \right)$$

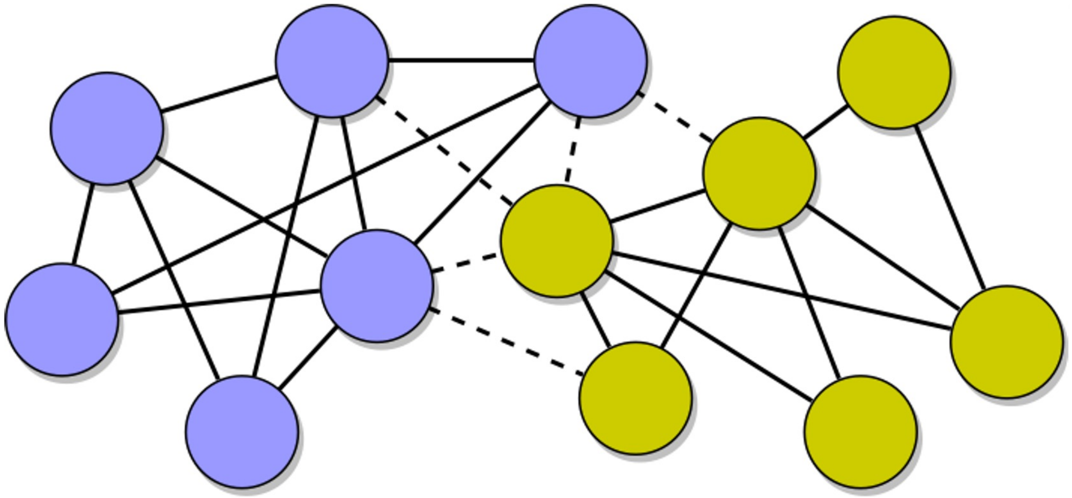
Quantifying homophily of a group

- We compare observed against the **expected number of links crossing to the other group**
- <1 ⇒ homophily 1 ⇒ neutral >1 ⇒ heterophily

$$\text{Homophily}(a) = \frac{L_{ab}}{L_a\left(\frac{N_b}{N}\right)} \qquad \text{Homophily}(b) = \frac{L_{ab}}{L_b\left(\frac{N_a}{N}\right)}$$

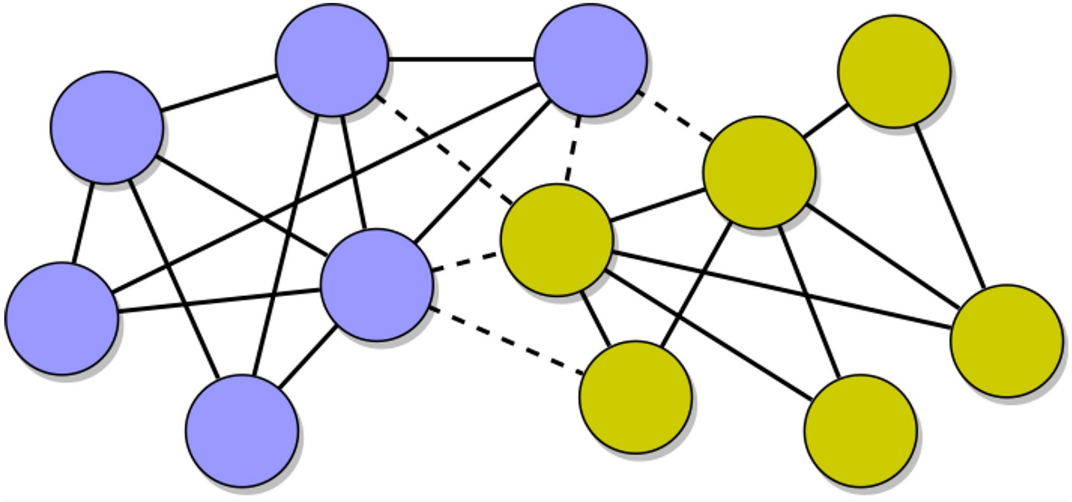
$$\text{Homophily}(a) = \frac{L_{ab}}{L_a\left(\frac{N_b}{N}\right)} = \frac{5}{14\left(\frac{6}{12}\right)} = \frac{5}{7}$$

Yellow nodes are homophilic



$N_a = 6$, $N_b = 6$, $L_a = 14$, $L_b = 16$, $L_{ab} = 5$,

$$\text{Homophily}(b) = \frac{L_{ab}}{L_b\left(\frac{N_a}{N}\right)} = \frac{5}{16\left(\frac{6}{12}\right)} = \frac{5}{8}$$

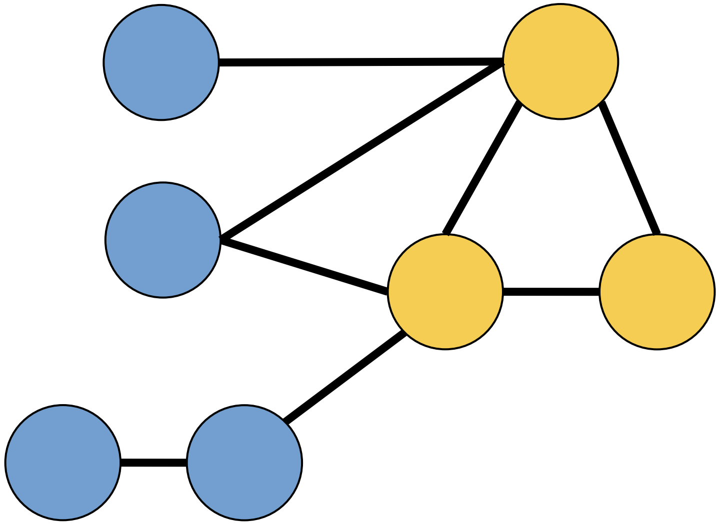


Blue nodes are homophilic

$N_a = 6$, $N_b = 6$, $L_a = 14$, $L_b = 16$, $L_{ab} = 5$,

Exercise

Compute homophily of both groups and indicate if each group is homophilic, heterophilic, or neutral



$$\text{Homophily}(a) = \frac{L_{ab}}{L_a\left(\frac{N_b}{N}\right)}$$

$$\text{Homophily}(b) = \frac{L_{ab}}{L_b\left(\frac{N_a}{N}\right)}$$

Pin board: https://upfbarcelona.padlet.org/chato/iig2u83qdzk4y7xc

Social influence

Homophily:
Similar nodes become connected

Features → Connections

The opposite mechanism may also happen!

Social influence:
Connected nodes become more similar

Connections → Features

Social network structure (ie, who our friends are) can determine our thinking!

Echo-chambers: like-minded people tightly connected

- No diversity of opinions
- Confirmation bias
- Reinforcement of prejudices

Summary

Things to remember

- How to quantify degree correlations:
  - Positive: assortative networks
  - Negative: disassortative networks
  - Neutral
- Friendship paradox
- Local and global clustering coefficient
- Homophily

Sources

- A. L. Barabási (2016). Network Science – Chapter 02
- F. Menczer, S. Fortunato, C. A. Davis (2020). A First Course in Network Science – Chapter 02
- URLs cited in the footer of specific slides

Practice on your own

- Determine if the set {C, D, G} is homophilic or heterophilic
- Calculate local clustering coefficient of each node in this graph