

# Homophily and triangles

## Social Networks Analysis and Graph Algorithms

Prof. Carlos Castillo — <https://chato.cl/teach>



Universitat  
Pompeu Fabra  
*Barcelona*

# Contents

- *This is all about friends, and friends of friends*
- Homophily
- Clustering coefficient

# Sources

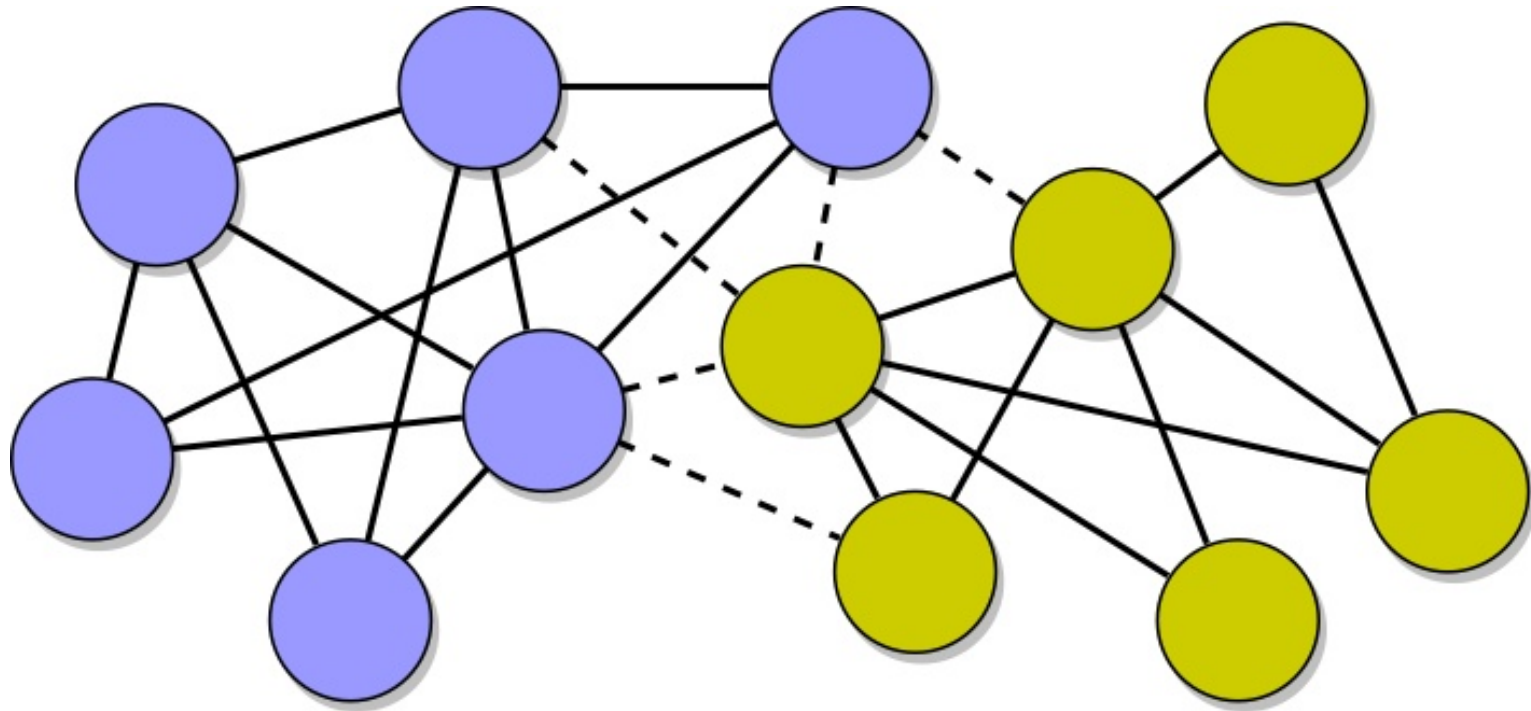
- A. L. Barabási (2016). Network Science – Chapter 02
- F. Menczer, S. Fortunato, C. A. Davis (2020). A First Course in Network Science – Chapter 02
- URLs cited in the footer of specific slides

# Homophily

# Who is a friend? [Homophily]

- In social networks, nodes have characteristics that influence their 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”



# Quantifying homophily

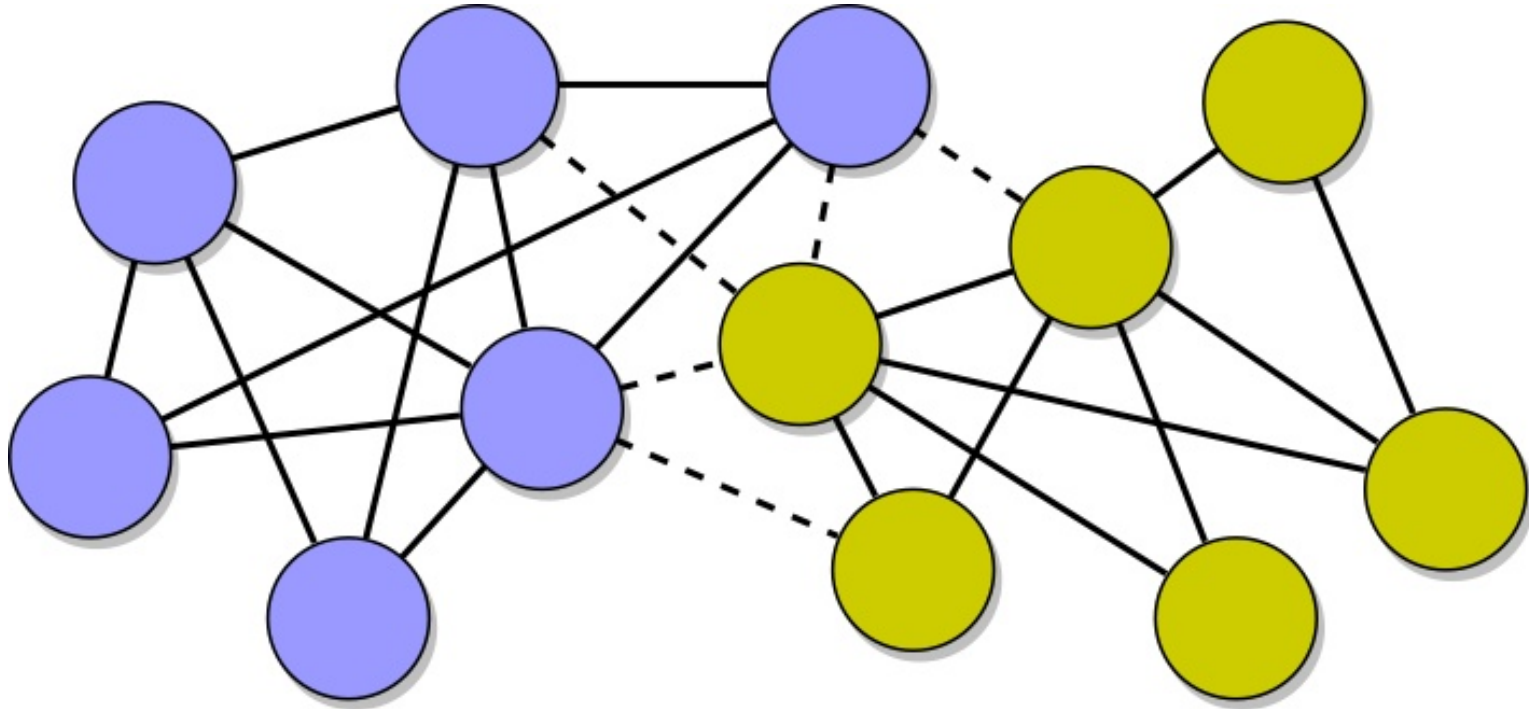
- Let  $G$  be a graph of  $N$  nodes:  $N_a$  “yellow” and  $N_b$  “blue”

$$N = N_a + N_b$$

- 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

$$L = L_{aa} + L_{ab} + L_{bb} \quad L_a = L_{aa} + L_{ab} \quad L_b = L_{bb} + L_{ab}$$

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





# 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**

$<1 \Rightarrow$  homophily

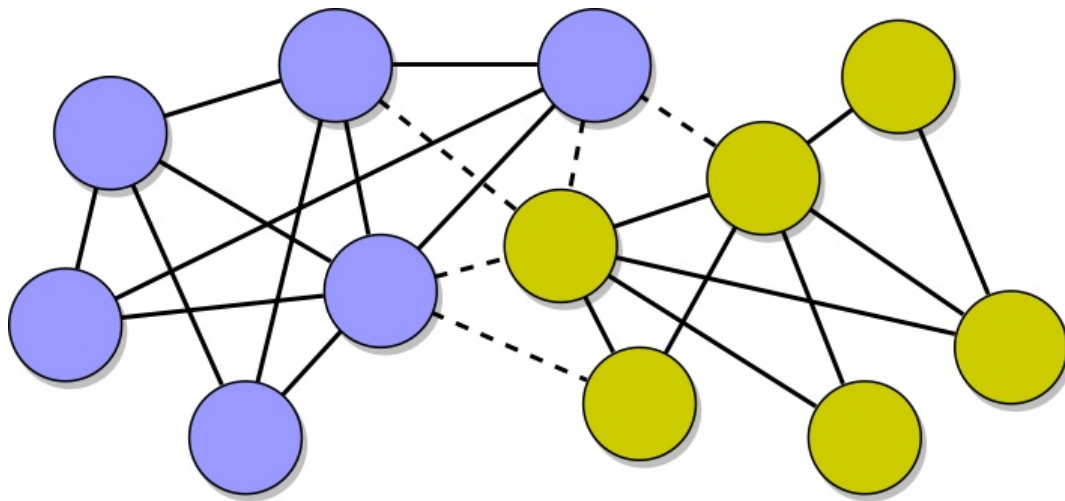
$1 \Rightarrow$  neutral

$>1 \Rightarrow$  heterophily

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

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

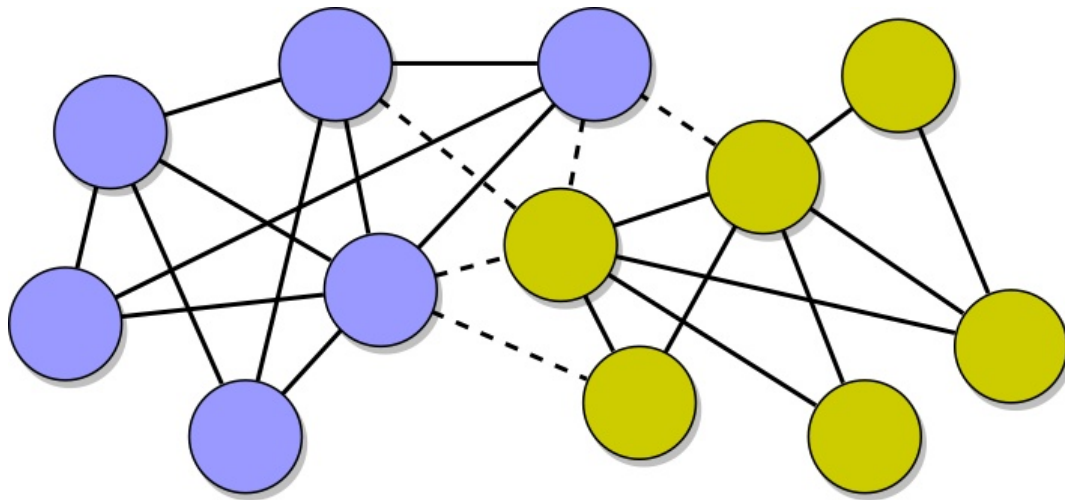
$$\text{Homophily}(a) = \frac{L_{ab}}{L_a \left( \frac{N_b}{N} \right)} = \frac{5}{14 \left( \frac{6}{12} \right)} = \frac{5}{7}$$



Yellow nodes are homophilic

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

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

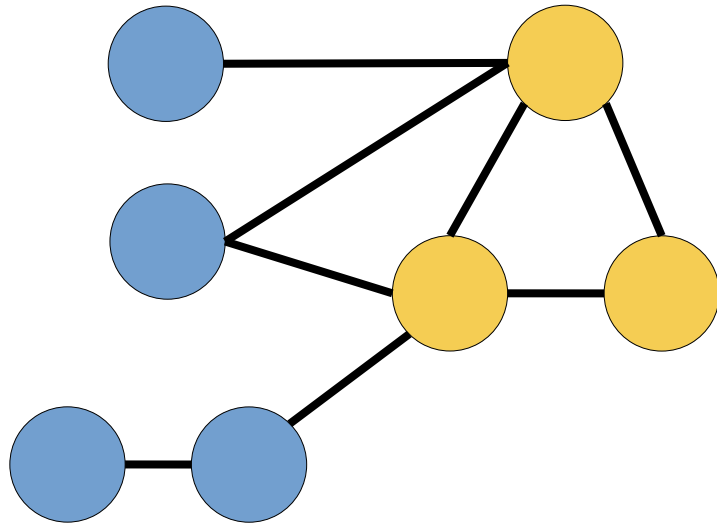


Blue nodes are  
homophilic

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

# Exercise

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



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

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

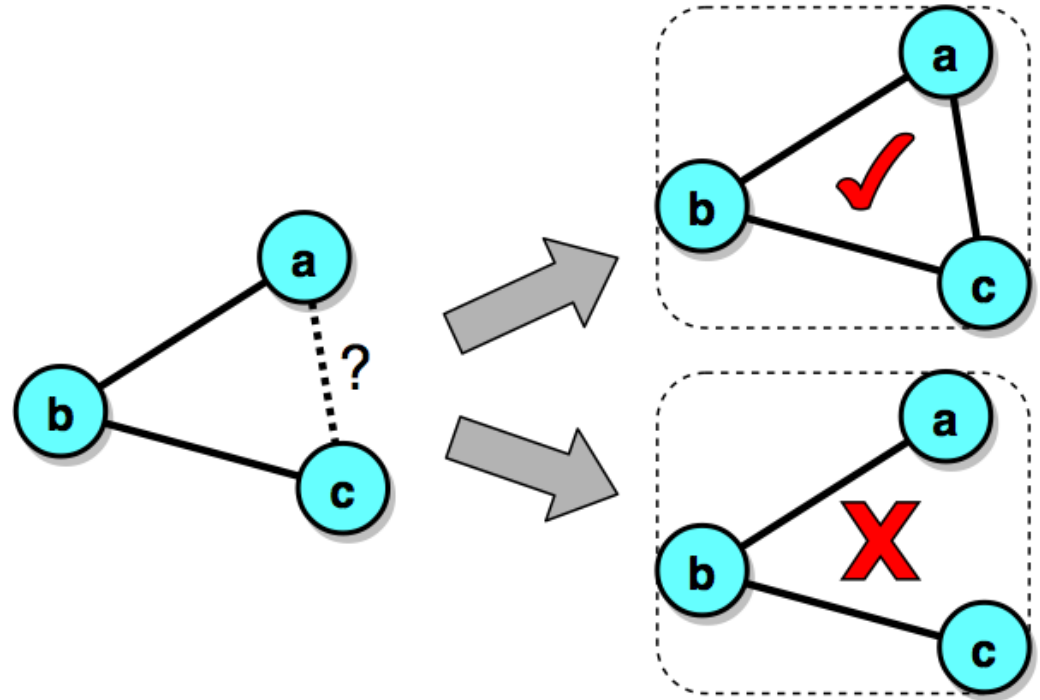
Pin board: <https://upfbarcelona.padlet.org/chato/iig2u83qdzk4y7xc>



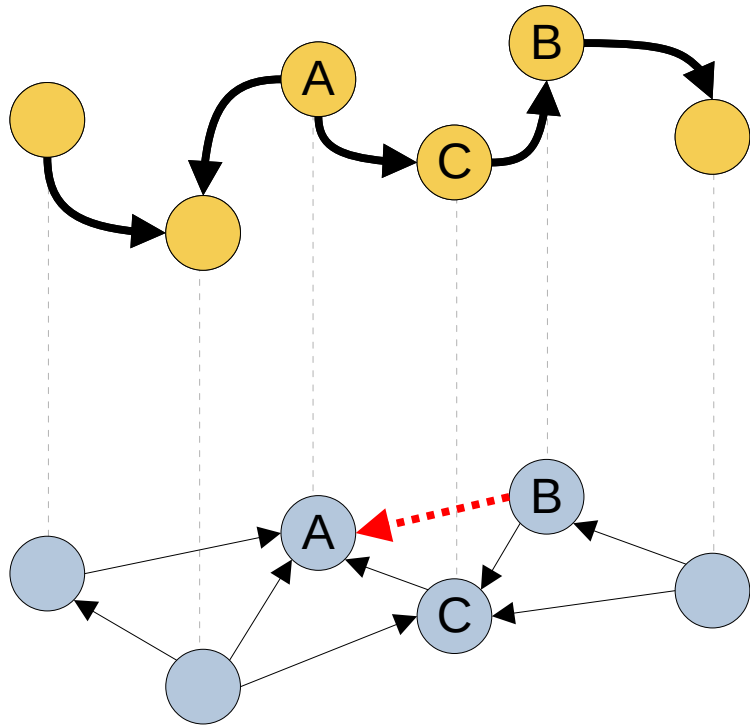
# Clustering coefficient

# Who is a friend? [Triangle closure]

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



# Tendency to form triangles

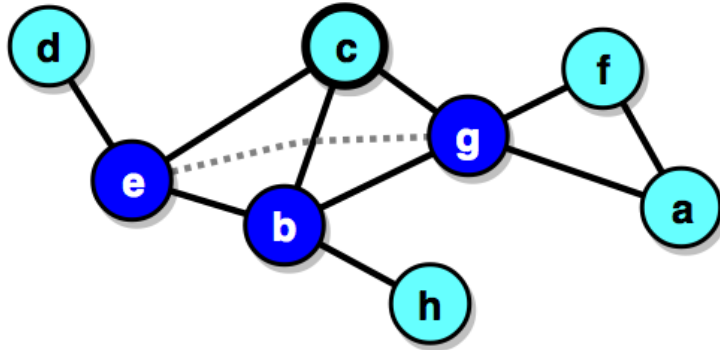


The dynamics on the network, i.e., information diffusion, affect the dynamics of the network, i.e., the creation of links

- *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



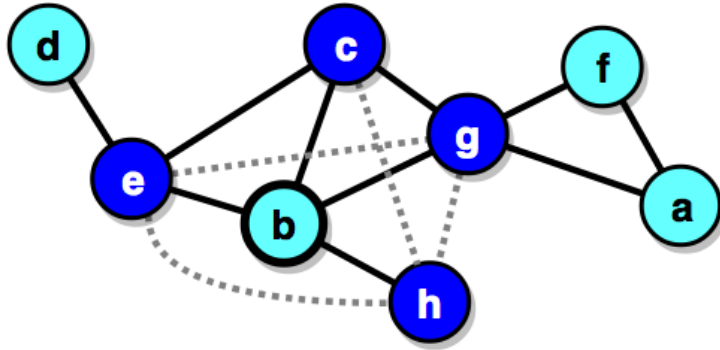
# Example 1



Node c has 3 neighbors: a, b, g

They form two triangles out of the possible 3 (the missing one is drawn with a dotted line)

## 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)

# Remember

The maximum number of links between  $k$  nodes is

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

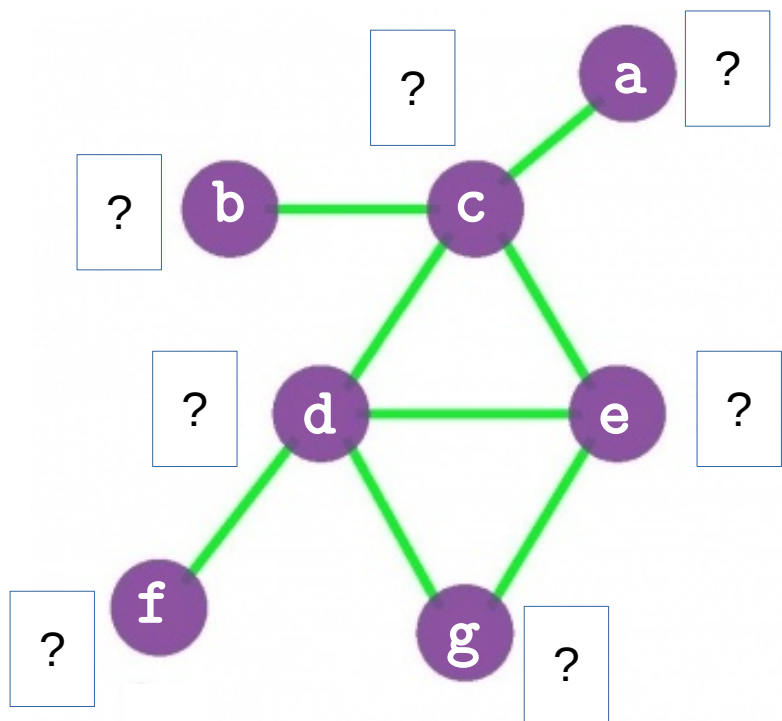
# Local clustering coefficient

- 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

# Summary

# Things to remember

- How to quantify if a group is
  - Homophilic
  - Heterophilic
  - Neutral
- Local and global clustering coefficient



# 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

