

Community Structure in Networks

Social Networks Analysis and Graph Algorithms

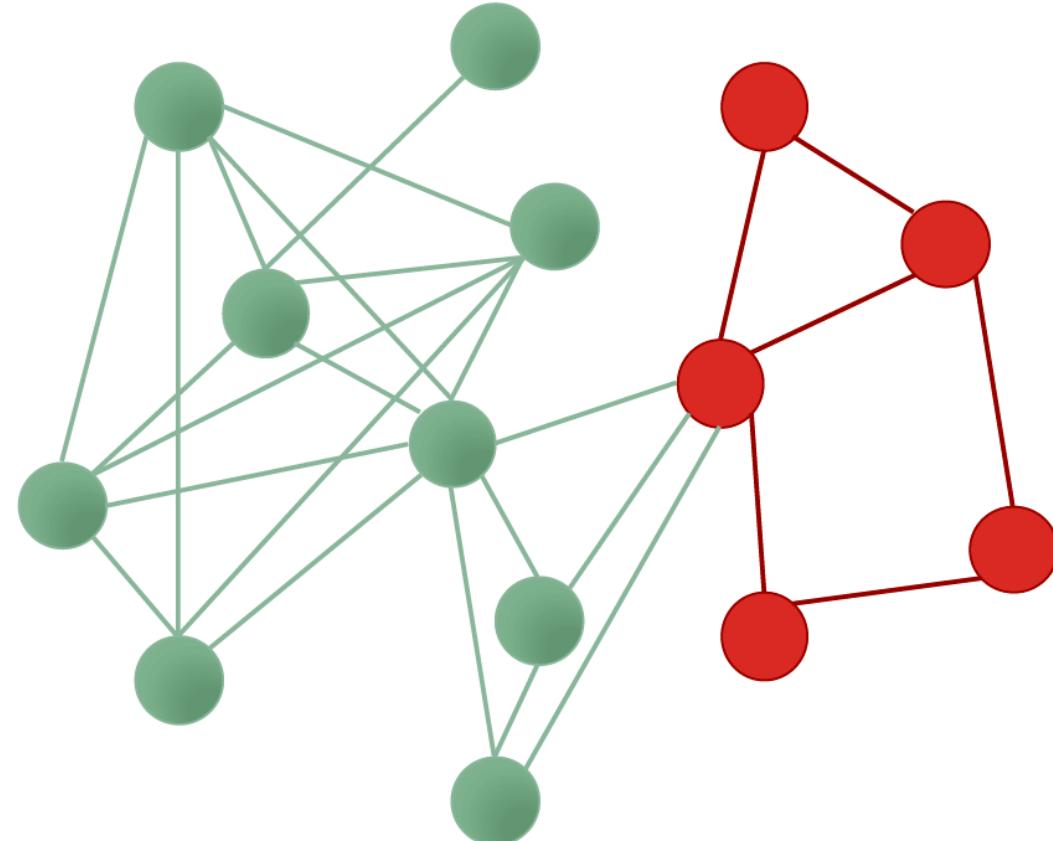
Prof. Carlos Castillo — <https://chato.cl/teach>

Sources

- A. L. Barabási (2016). Network Science – Chapter 09
- D. Easley and J. Kleinberg (2010). Networks, Crowds, and Markets – Chapter 03
- F. Menczer, S. Fortunato, C. A. Davis (2020). A First Course in Network Science – Chapter 06
- URLs cited in the footer of slides

Reminder: sub-graph

A sub-graph is a subset of nodes, and all of the edges among those nodes



Typical community structures

- **One** dense sub-graph
 - embedded somewhere within a larger graph
- **Two** groups (polarization)
 - plus perhaps some ambiguous nodes
- **Multiple** communities

One dense sub-graph

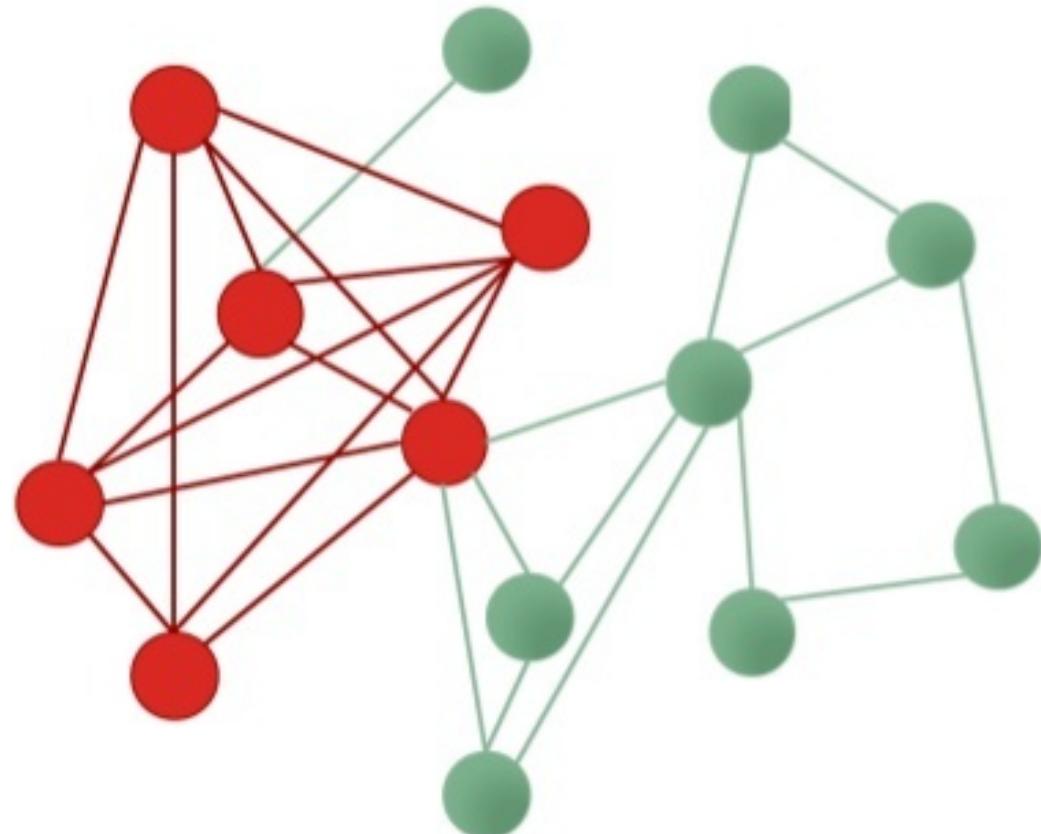
Densest sub-graph

Sub-graph having the maximum **density** according to some measure

There is more than one definition of density!

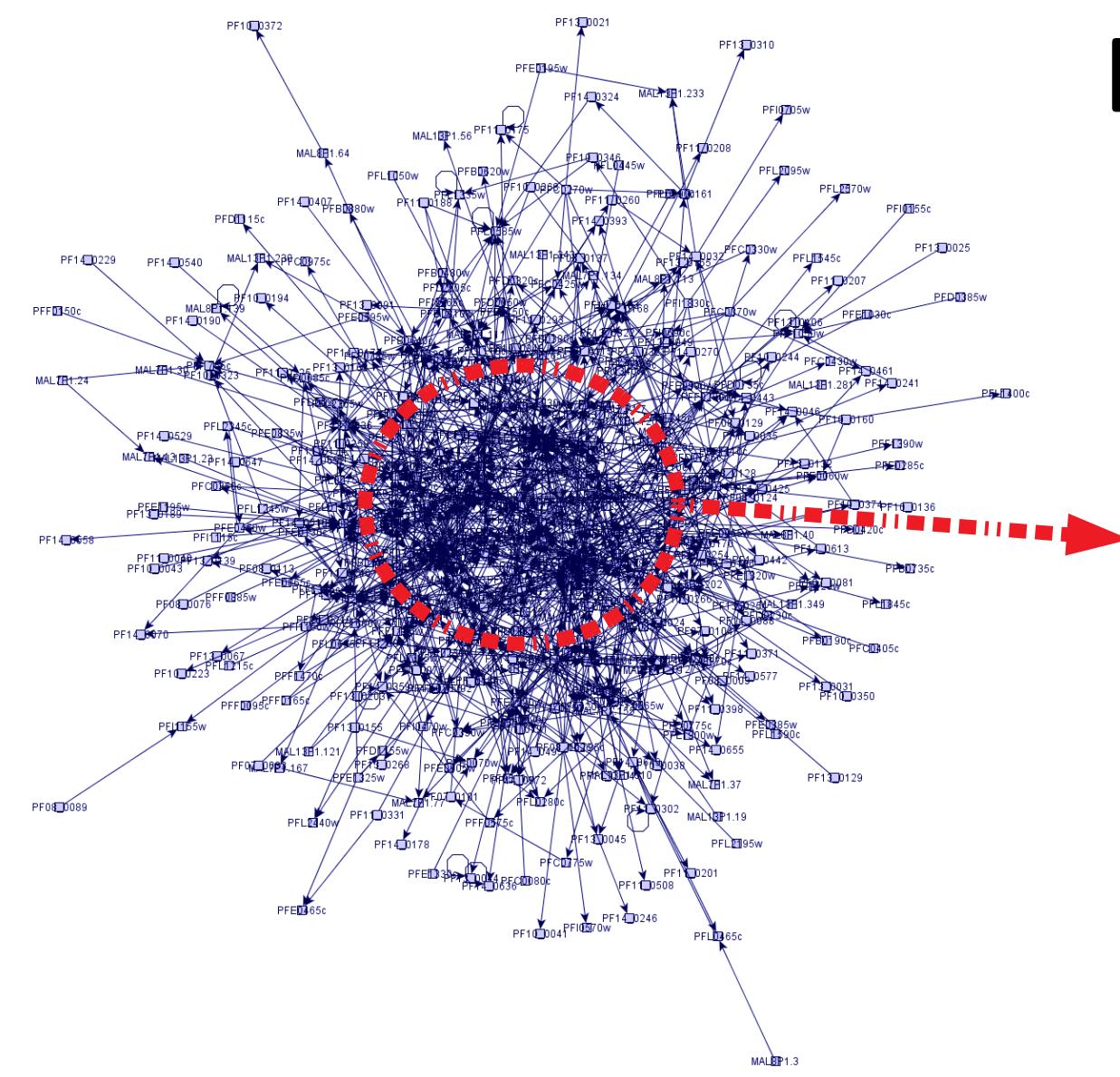
$$\text{Density} = |E|/|V|$$

$$\text{Density} = (2|E|)/(|V|(|V|-1))$$



Many graphs look like “hairballs”

Sometimes, at the center these graphs may have an interesting dense sub-graph



Asthma-related genes



https://www.youtube.com/watch?v=VU_7FHAKMgA

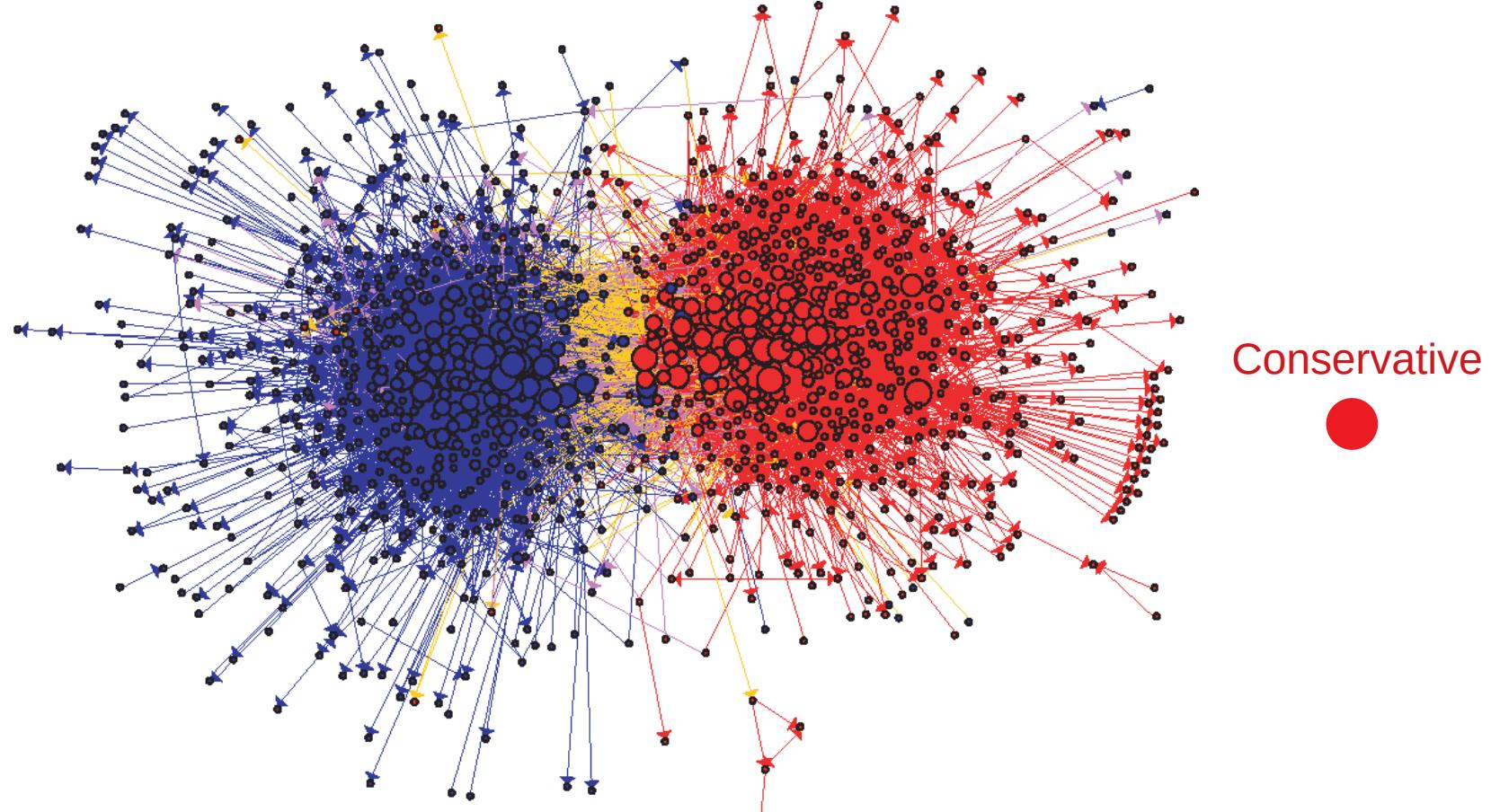
Two groups (polarization)

US Political Blogs (2004)

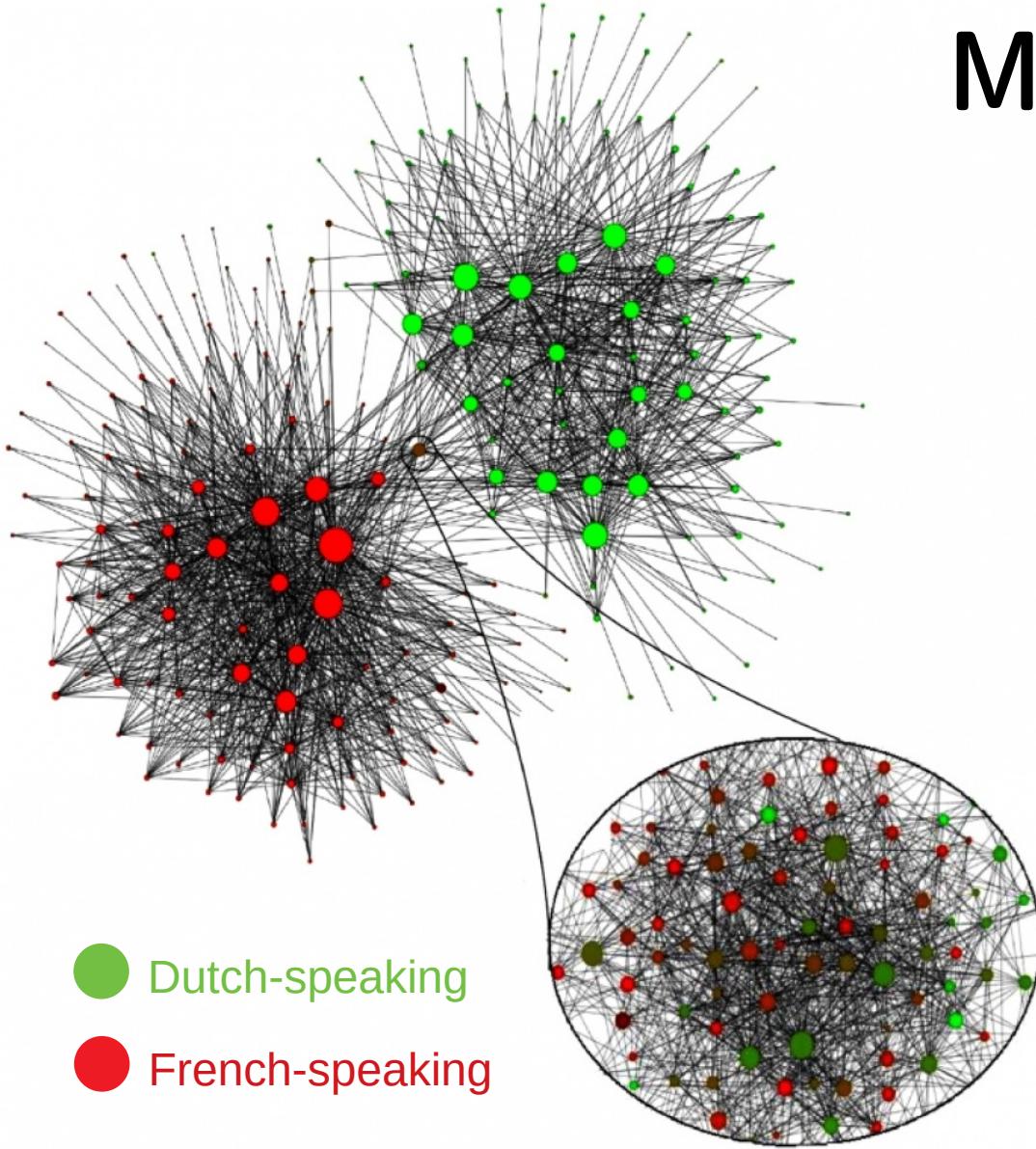
Liberal



Conservative



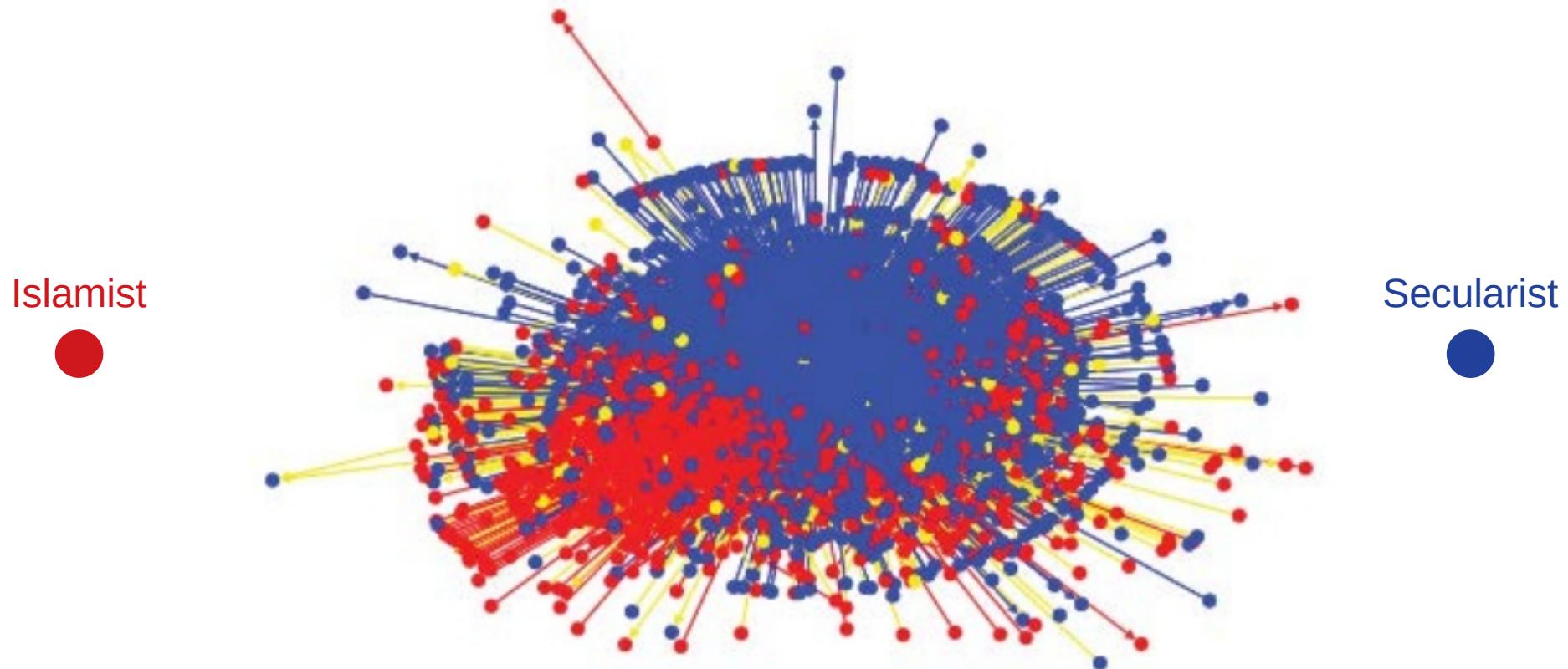
Mobile phone users in Belgium (2008)



Each node is a community of 100 mobile users or more that tend to call each other

V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *J. Stat. Mech.*, 2008.

Egyptian Twitter Users (2013)

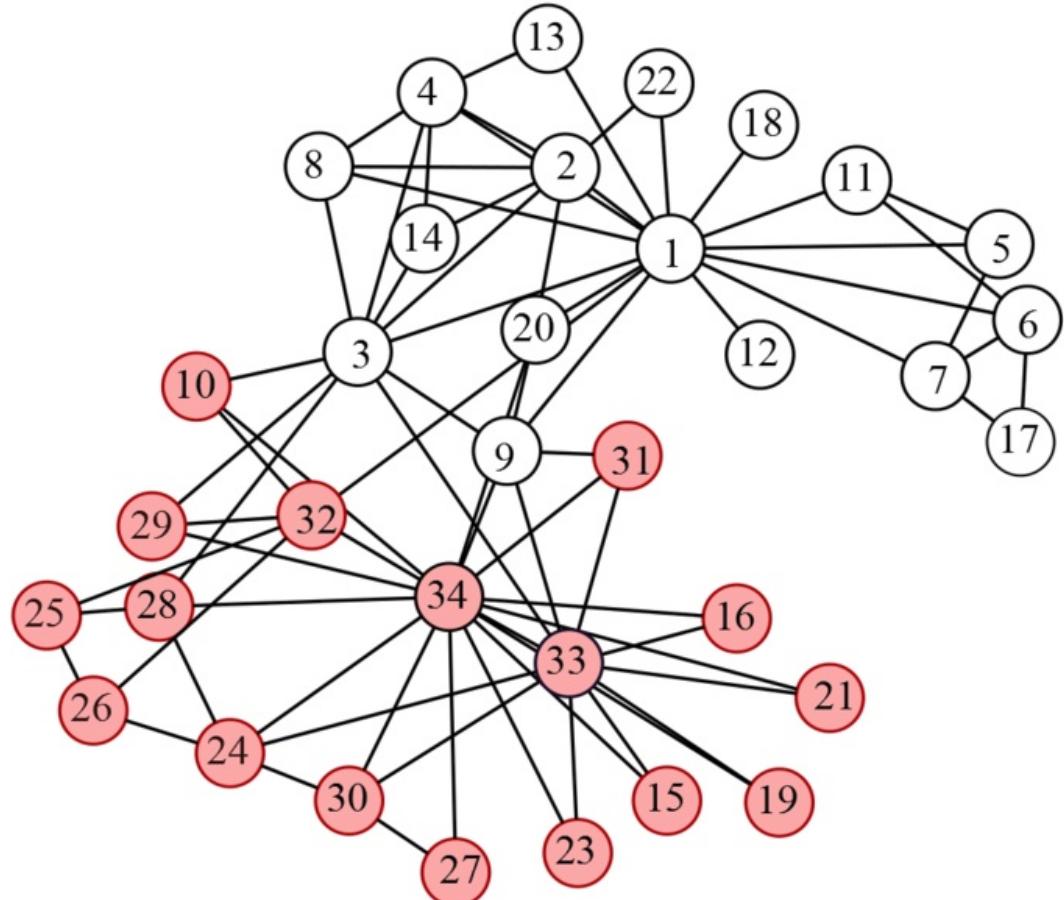


Political Books



Wayne Zachary's PhD Thesis (1972)

- Studied 34 members of a karate club
- Found 78 links between members who regularly interacted outside the club
- The club splitted in two during the study
- 1=sensei, 34=president



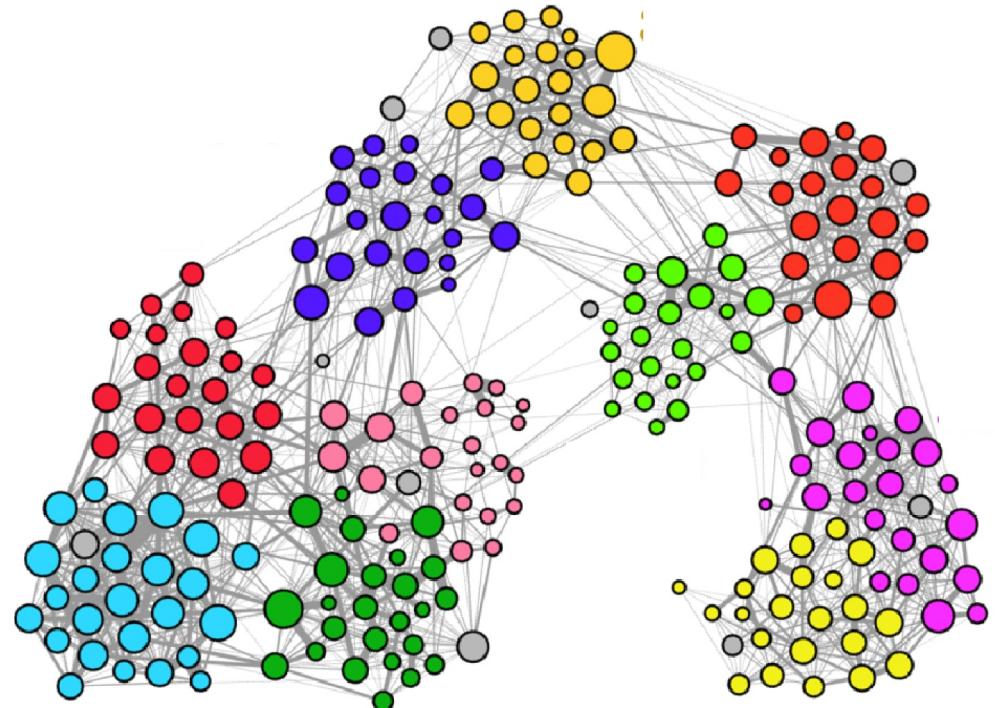
Multiple communities

Primary school contacts

Links connect students who spent more than two minutes face to face

Students wore RF-ID badges hanging on their chest, which have a range of about 1.0-1.5 meters

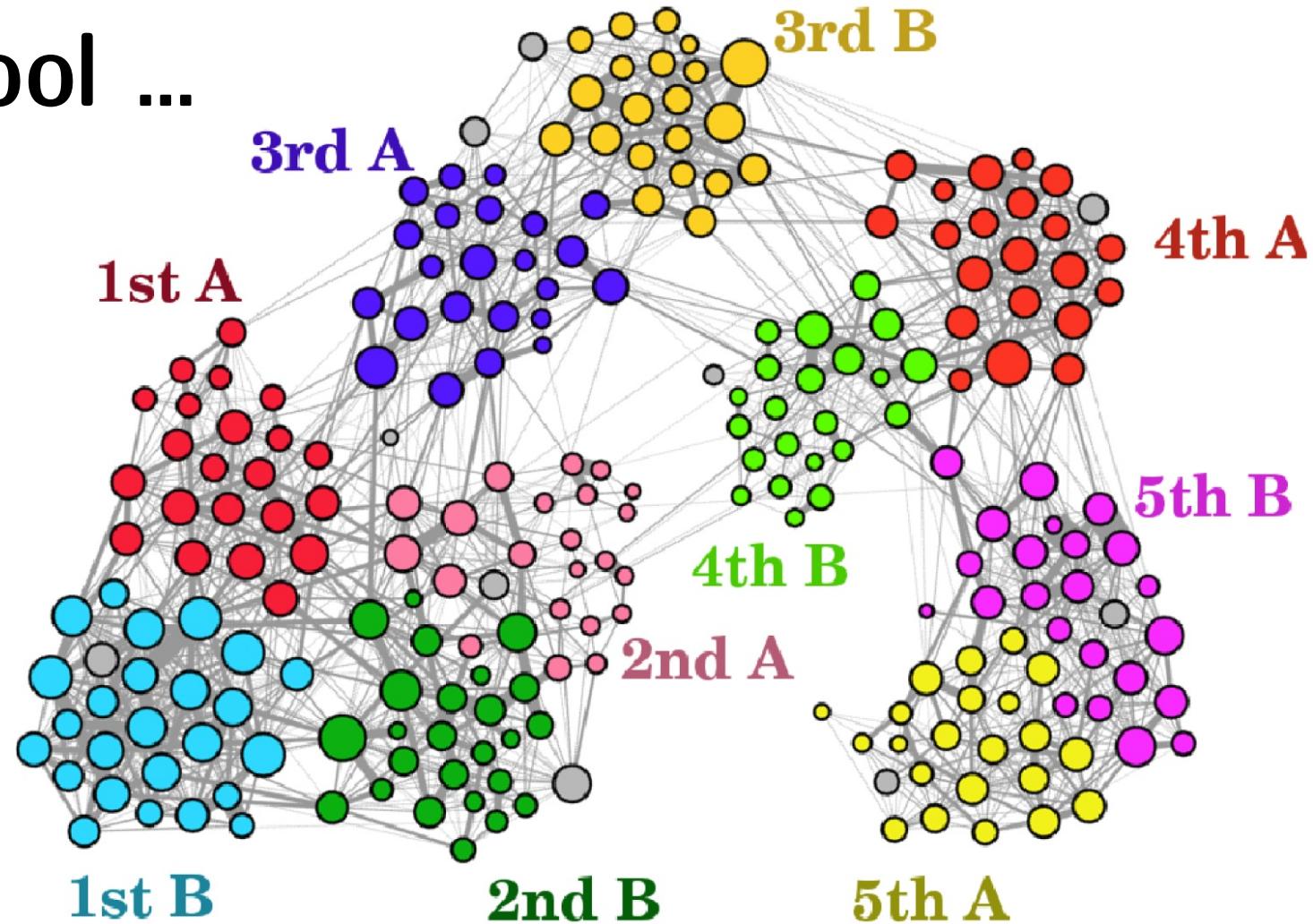
What do you think the colors represent in this visualization?



Primary school ...

Colors represent classes. Teachers are shown in gray color.
Node sizes are number of connections.

Note: these communities are hierarchical (more on this later)



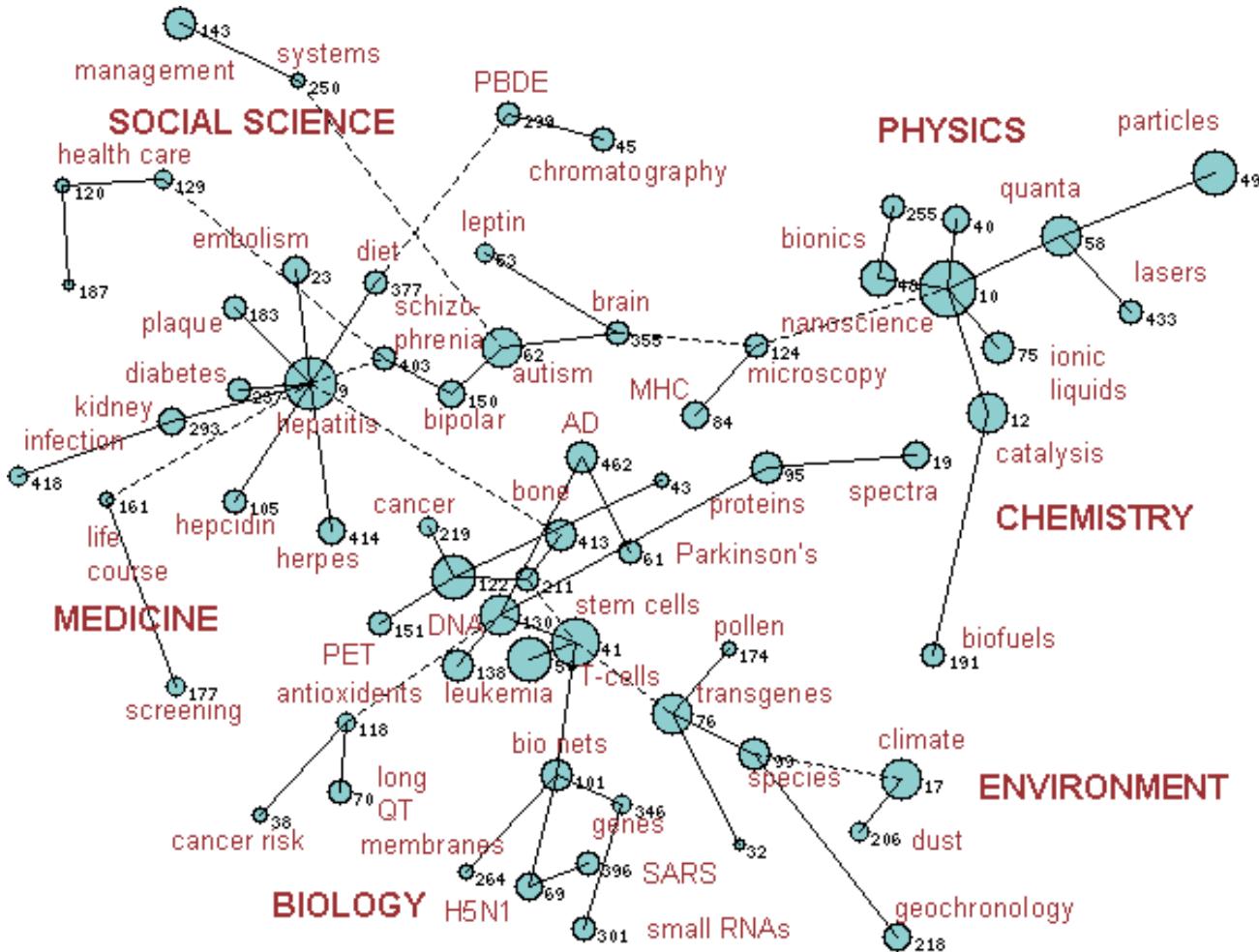
Stehlé, J., et al. (2011).

High-resolution measurements of face-to-face contact patterns in a primary school.

PloS one, 6(8), e23176.

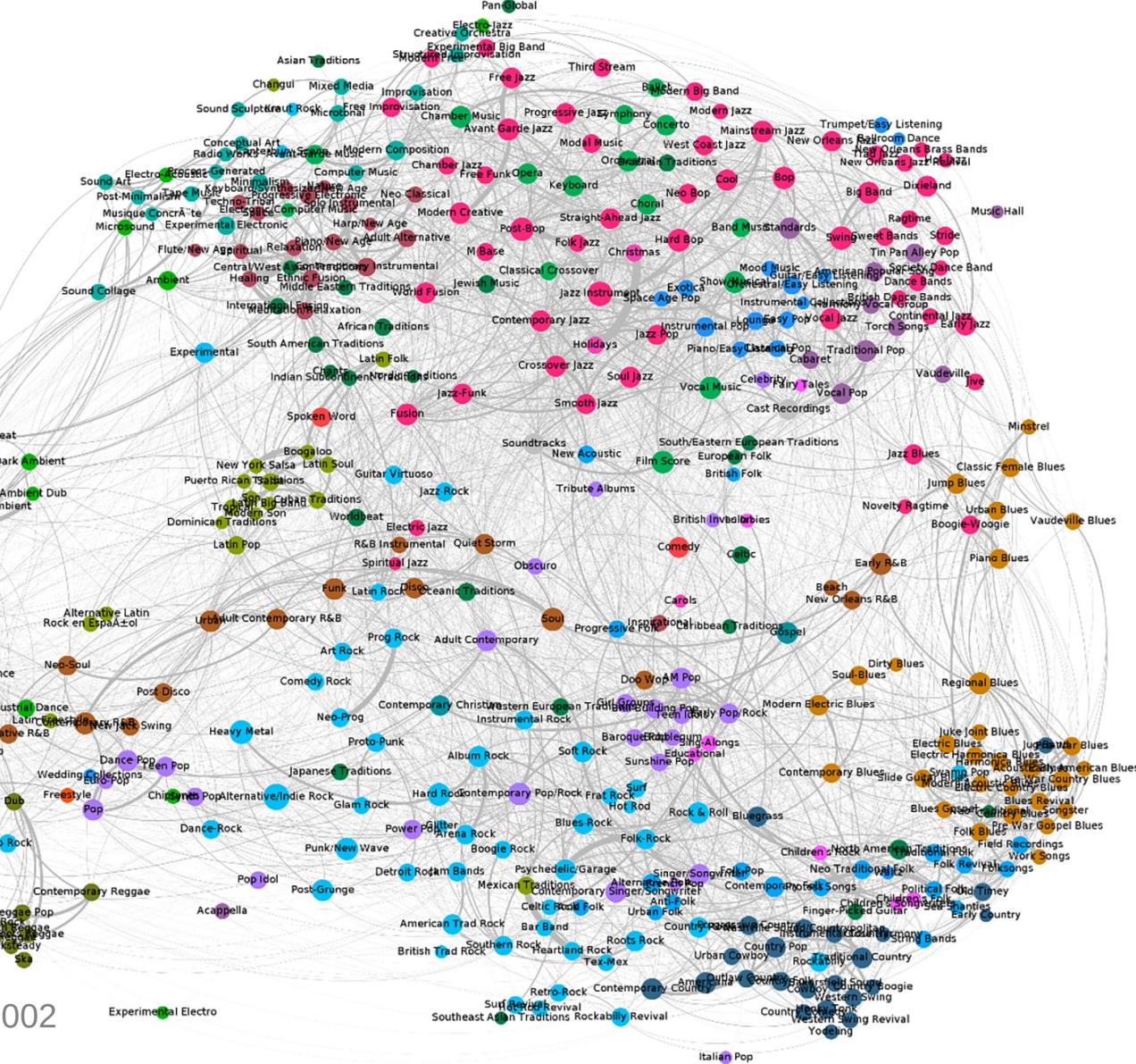
Science

Two topics T_1 , T_2 , are connected if there is at least one paper that cites:
a paper u in T_1 and
a paper v in T_2 .

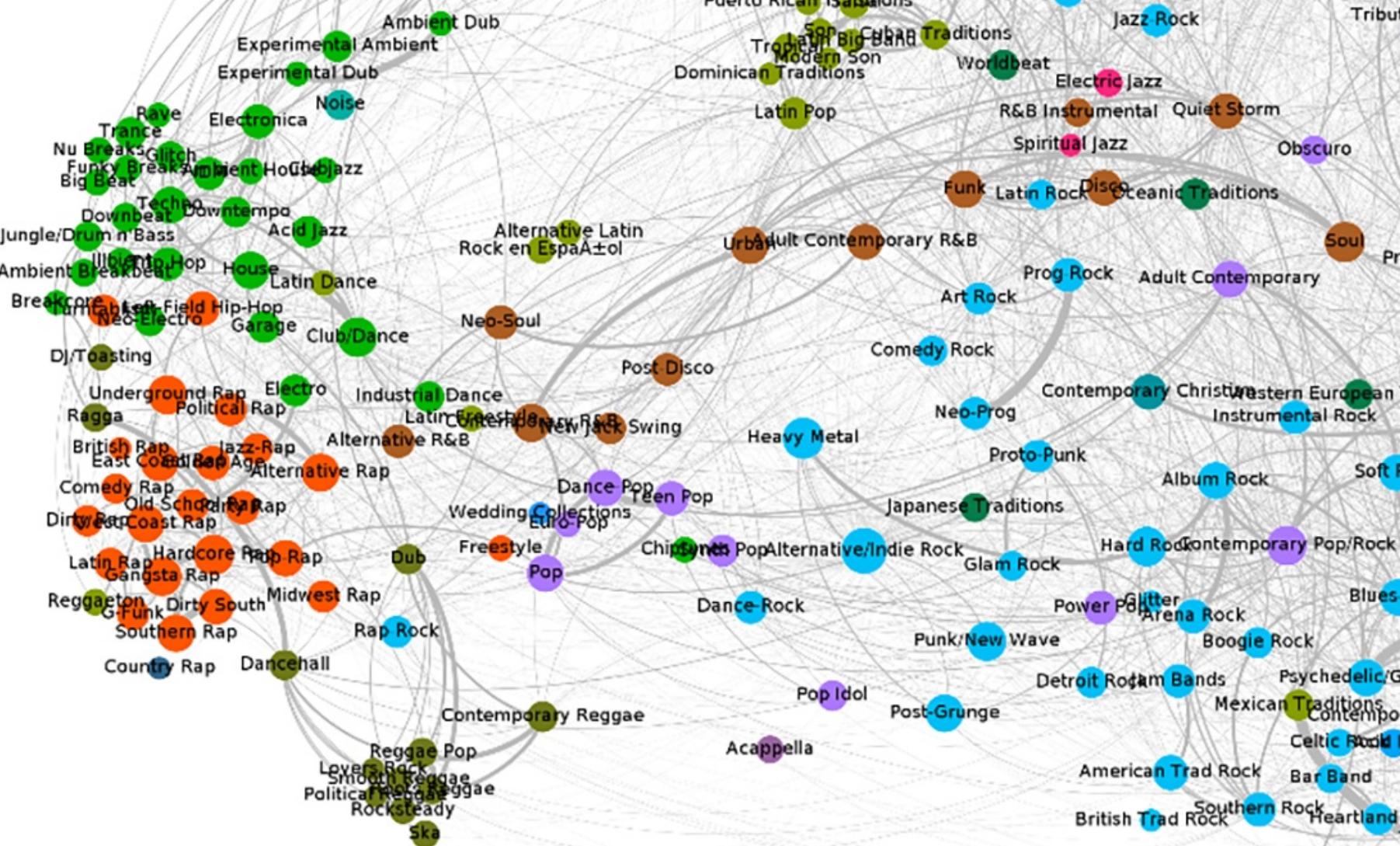


Music

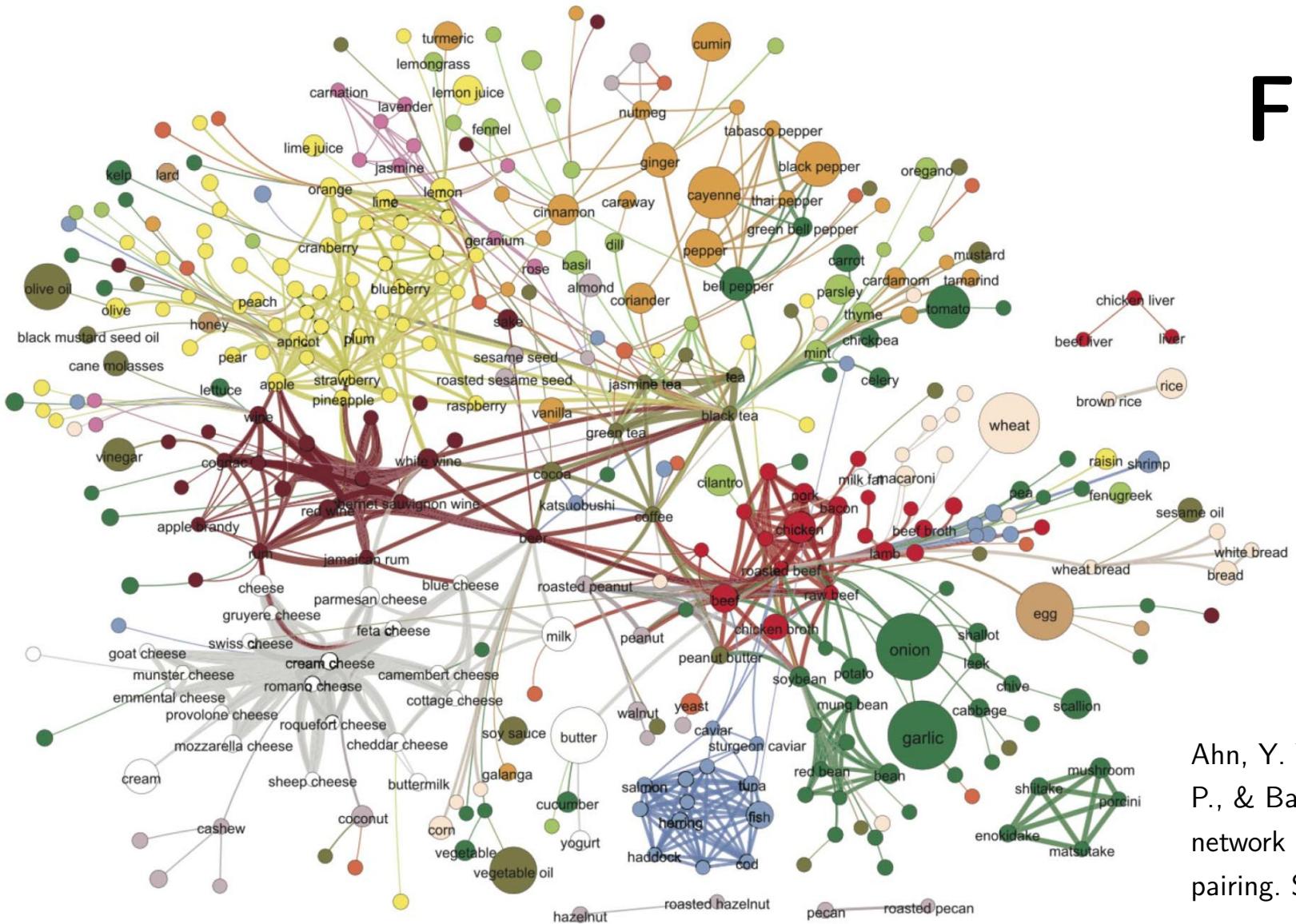
Two Genres, G1,
G2, are connected
if there is a
musician producing
tracks in both
genres; width of
link is number of
musicians



Music (cont.)

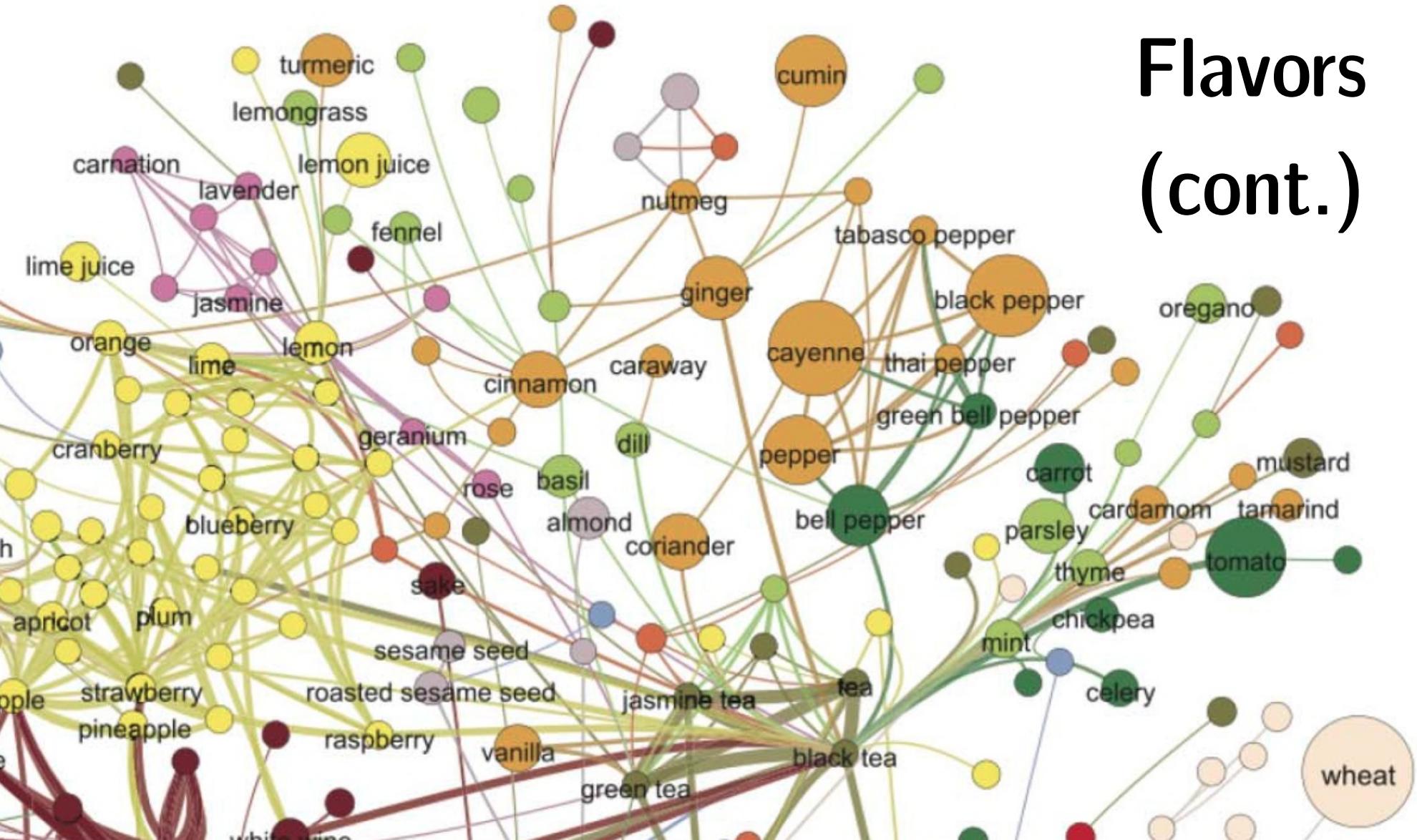


Flavors

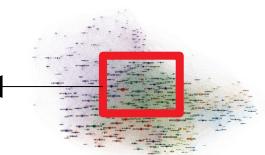
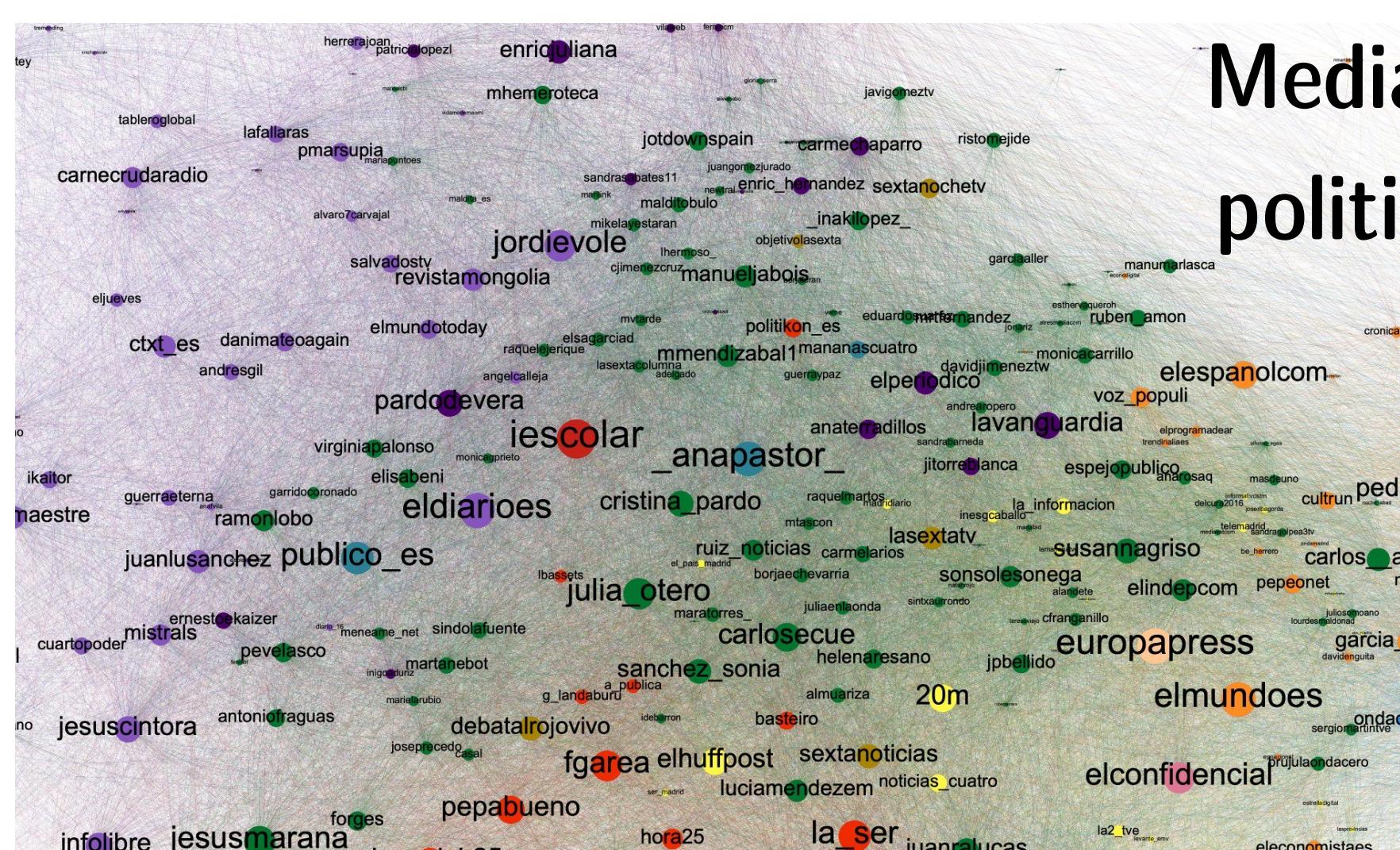


Ahn, Y. Y., Ahnert, S. E., Bagrow, J. P., & Barabási, A. L. (2011). Flavor network and the principles of food pairing. *Scientific reports*, 1, 196.

Flavors (cont.)



Media and politicians

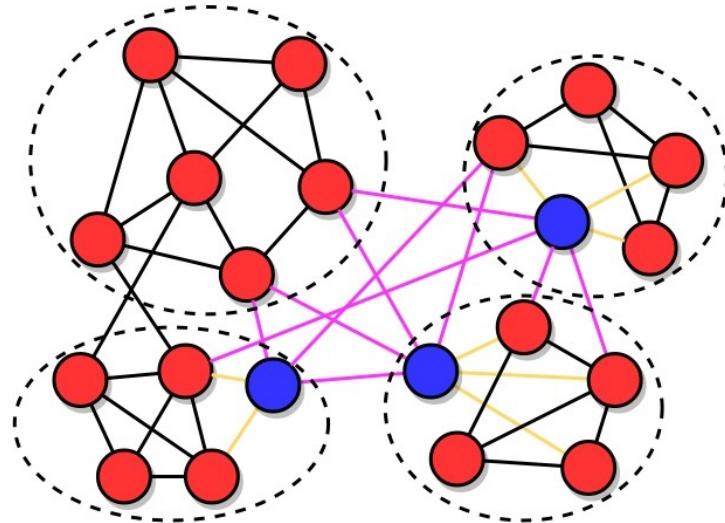


<https://twitter.com/jbo/status/1120444347772821504/photo/1>

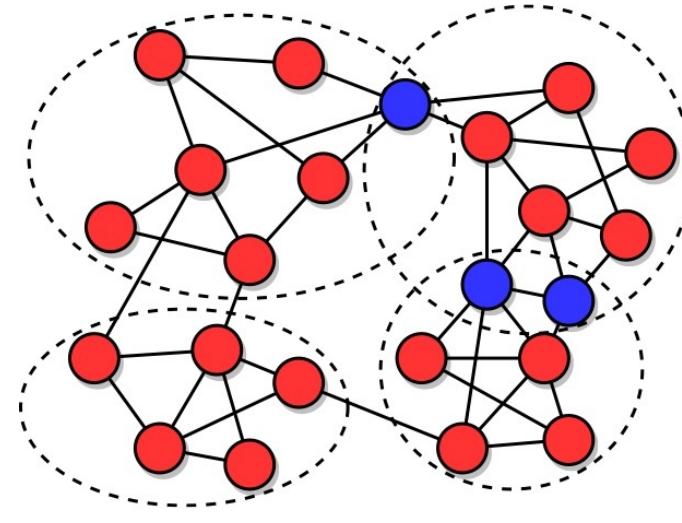
Partitions vs Overlapping communities

Hierarchical communities

Partition vs Overlapping communities



Partition, or *hard* clusters

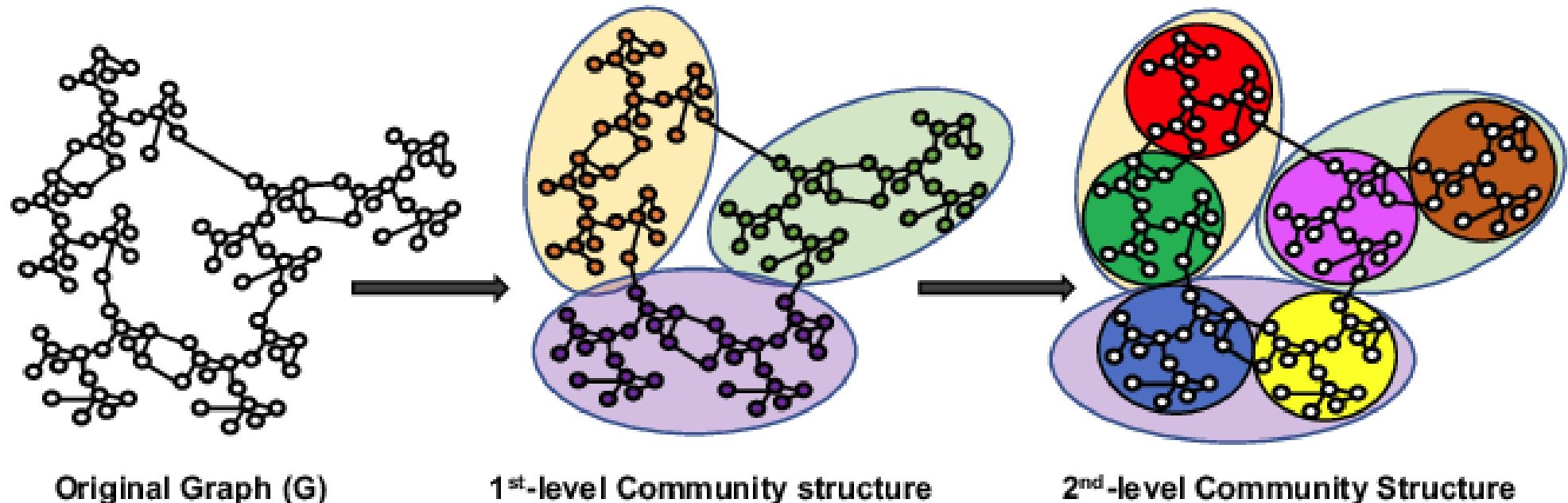


Overlapping communities, or *soft* clusters

What's special about blue nodes?

Blue nodes are in more than one community

Hierarchical communities



Original Graph (G)

1st-level Community structure

2nd-level Community Structure

How to generate a graph having
community structure?

Exercise

Invent a variant of the ER model that generates graphs having **two communities**, but not necessarily disconnected



Standard ER Model:

- Generate N nodes
- For each of the $N(N-1)/2$ pairs of nodes:
 - Perform a Bernoulli trial with probability p
 - If the trial succeeds, connect those nodes

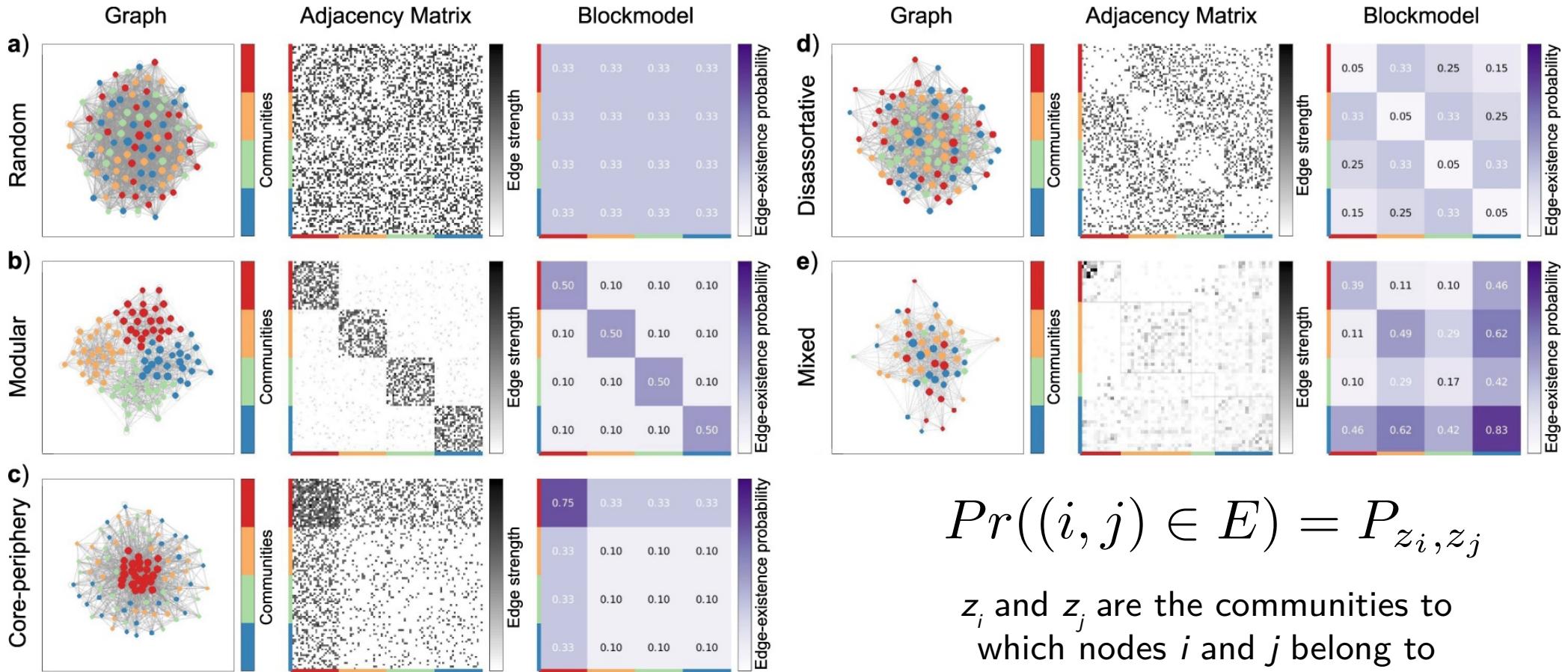
Pin board: <https://upfbarcelona.padlet.org/chato/tt20-community-structure-yhk7nozup8xbiejb>



How to generate community structure?

- The **stochastic block model** generates graphs with community structure
- Can be described as a variation of the ER model in which:
 - There are m groups
 - Link probability scalar p becomes an $m \times m$ matrix P that contains in position (i,j) the probability of a link between a node in group i and a node in group j

Examples of stochastic block model



Summary

Things to remember

- Many networks have community structure
- Sometimes it's:
 - One dense sub-graph
 - Two communities (polarization)
 - Multiple communities
- Partitions vs overlapping communities
- Hierarchical communities