uc3m | Universidad Carlos III de Madrid

Network Theory for the Social Sciences in Python

Methods Workshop: Social Sciences PhD Program (2025/2026)

2. Network models and structural analysis

https://github.com/blas-ko/uc3m_networks_workshop_2025

Blas Kolic
blas.kolic@uc3m.es

Summary

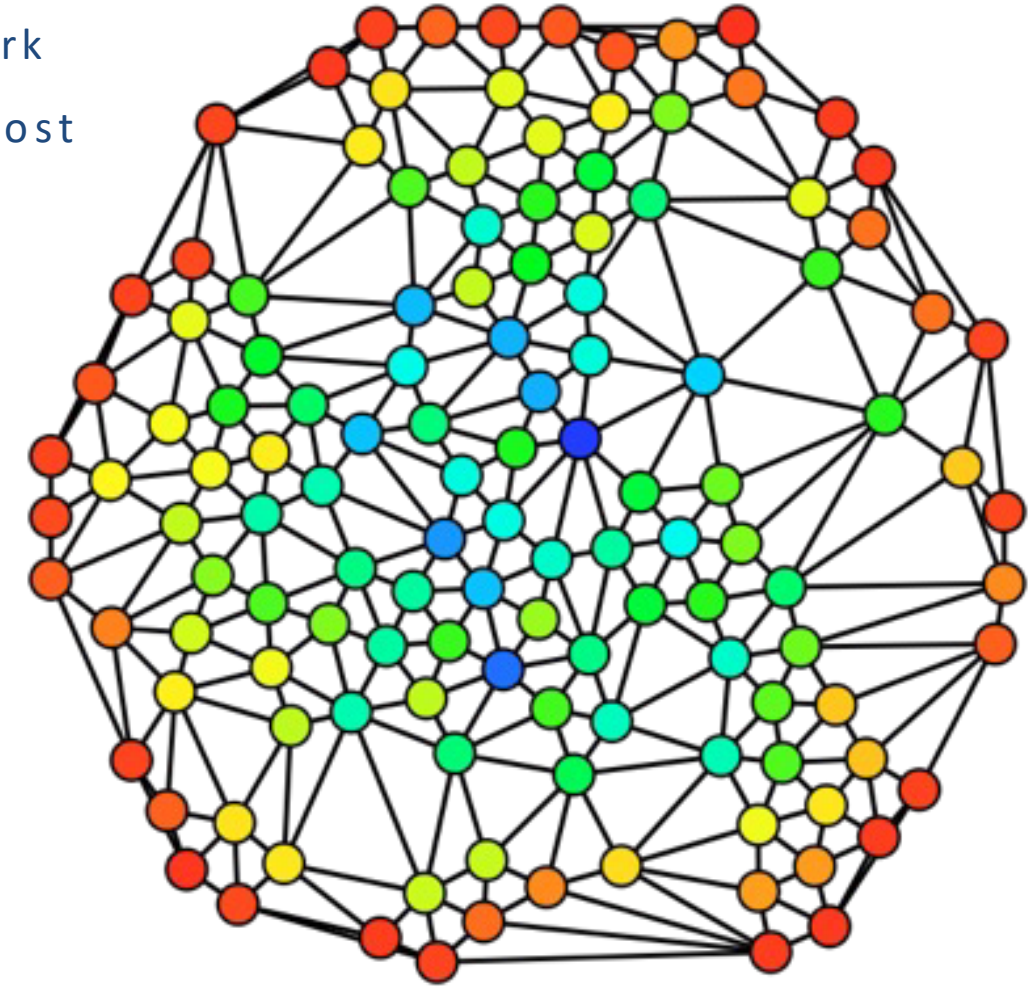
1. Centrality: node importance
    - Degree
    - Closeness
    - Betweenness
    - Eigenvector
2. Network models
    - Erdös-Rényi model (random)
    - Configuration model (degree)
    - Watts-Strogatz model (small world)
3. Community Structure
    - Intuition
    - Modularity as a community score
    - Community detection methods

1 Centrality: node importance

Centrality

What are the most important nodes in a network?

- Most influent people in a social network
- People responsible for spreading most infections in a epidemic
- Most important infrastructures in a transportation network.

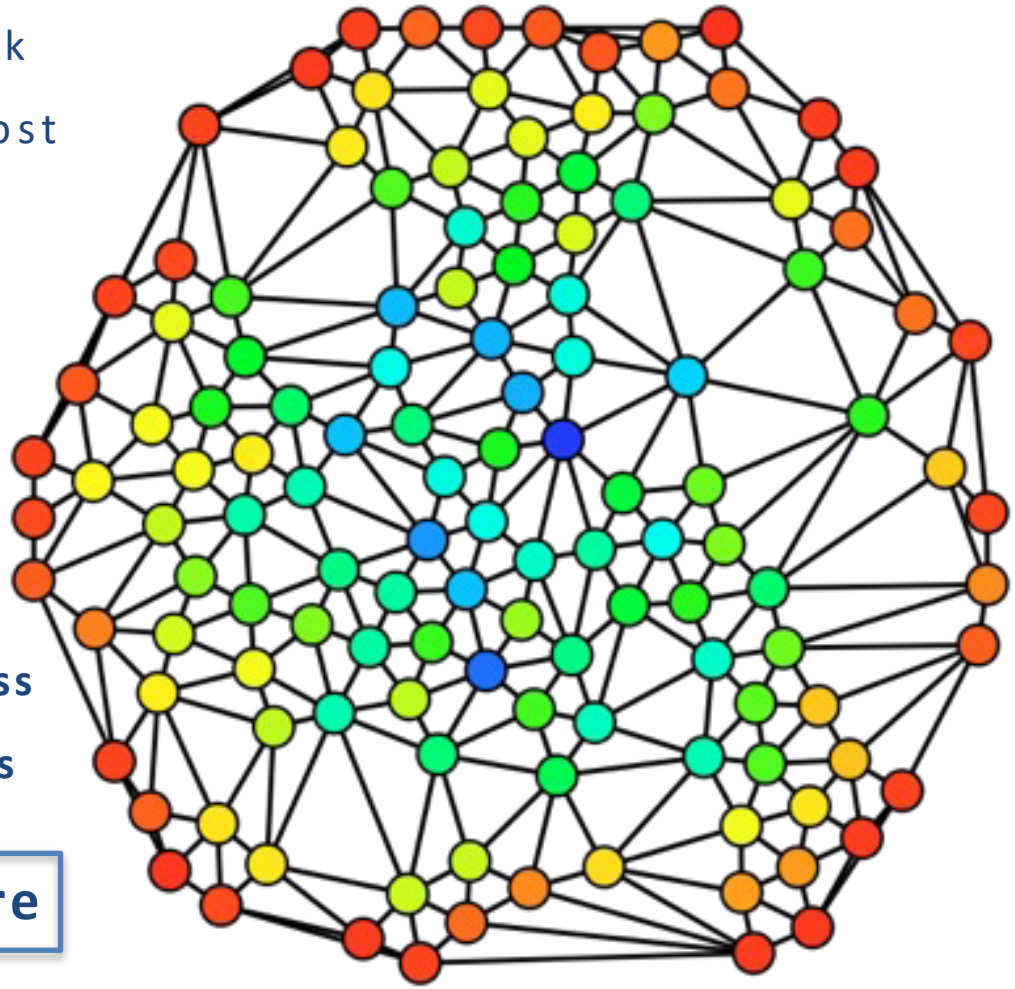


uc3m

Centrality

What are the most important nodes in a network?

- Most influent people in a social network
- People responsible for spreading most infections in a epidemic
- Most important infrastructures in a transportation network.



Importance may mean different things depending on the situation:

- Are you popular? → **degree**
- Are you reachable by most? → **closeness**
- Do you act as a bridge? → **betweenness**

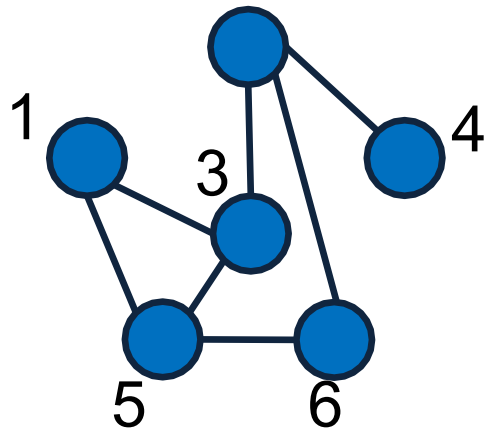
Centrality is a **node-level measure**

Degree centrality

- Number of direct connections

$$c_i = \frac{k_i}{N-1}$$

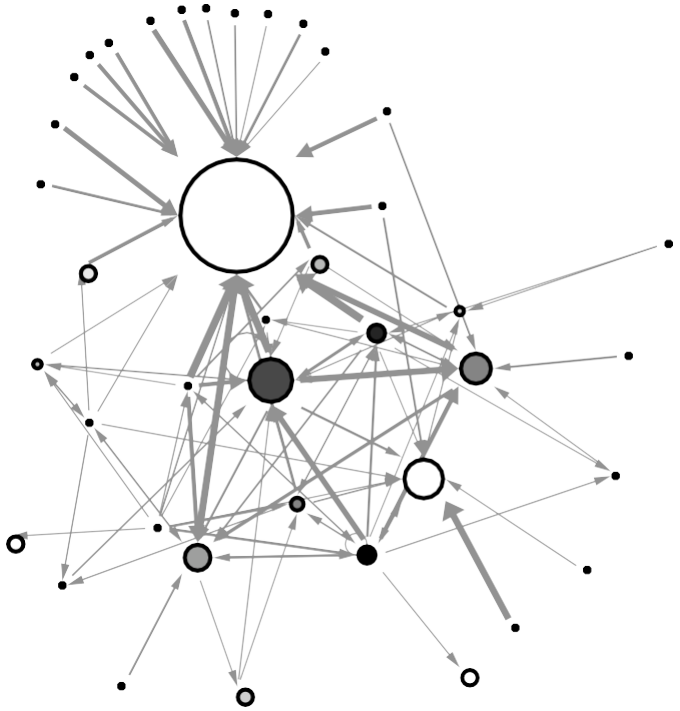
- Computationally cheap
- Local measure: doesn't tell us about global influence
- There can be many ties (nodes with same degree)



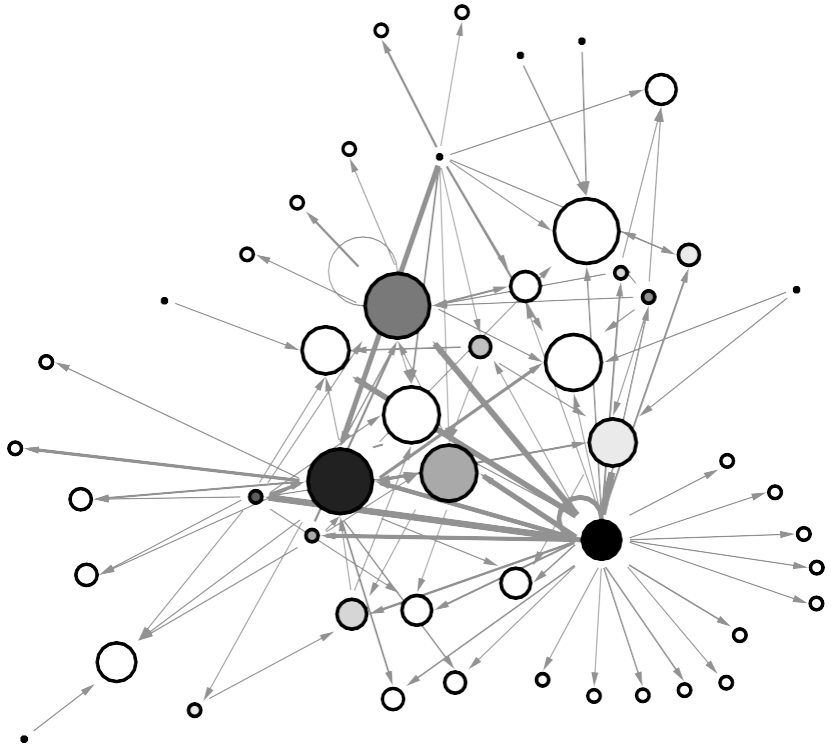
Node	Degree	Ranking
1	2	2
2	3	1
3	3	1
4	1	3
5	3	1
6	2	2

Degree centrality

- Example: trading networks



Nodes with large degree centrality: a node trades with many

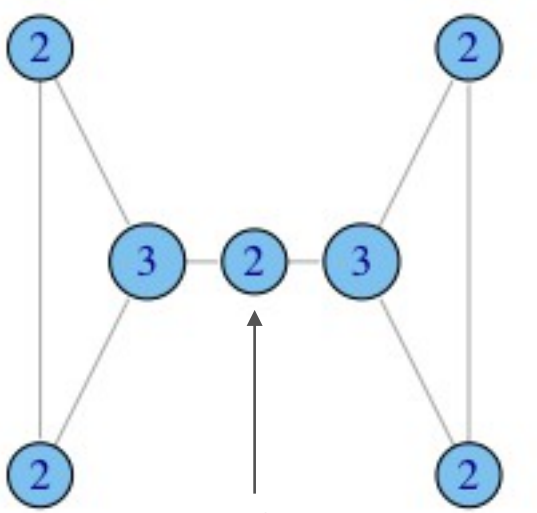


Small centrality nodes: trading is more distributed

uc3m

Degree centrality

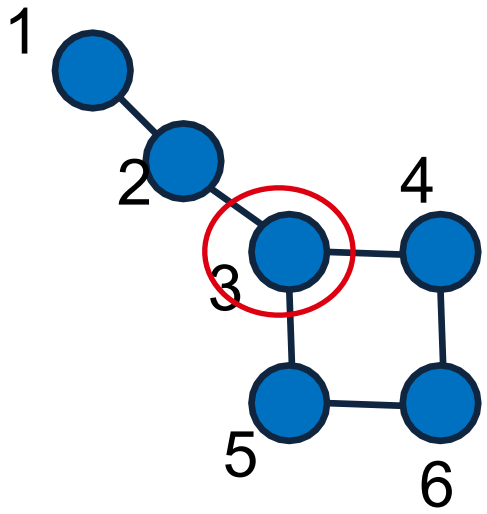
- Problems: it does not capture other types of importance, like:
  - Ability to connect between groups of nodes
  - Probability to receive or transmit information to many nodes



Closeness centrality

- A node is *central* if it is **near** most other nodes.
- Closeness centrality (c_i): The inverse of the sum of **distances** from node i to all other nodes.

$$c_i = \frac{1}{\sum_{j \neq i} d_{ij}}$$



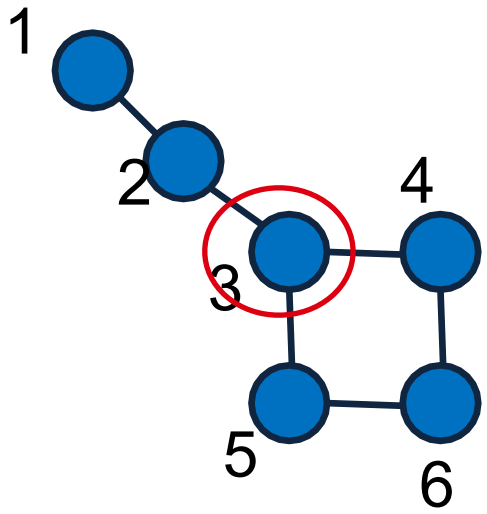
d_{ij}	1	2	3	4	5	6
1		1	2	3	3	4
2	1		1	2	2	3
3	2	1		1	1	2
4	3	2	1		2	1
5	3	2	1	2		1
6	4	3	2	1	1	

distance matrix

Closeness centrality

- A node is *central* if it is **near** most other nodes.
- Closeness centrality (c_i): The inverse of the sum of **distances** from node i to all other nodes.

$$c_i = \frac{1}{\sum_{j \neq i} d_{ij}}$$



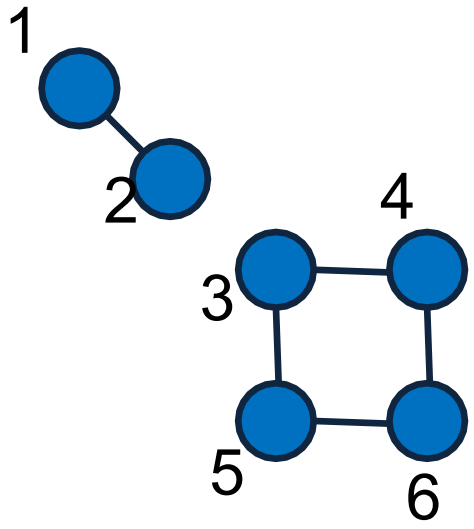
d_{ij}	1	2	3	4	5	6
1		1	2	3	3	4
2	1		1	2	2	3
3	2	1		1	1	2
4	3	2	1		2	1
5	3	2	1	2		1
6	4	3	2	1	1	

distance matrix

Node	Closeness	Rank
1	1/13	4
2	1/9	2
3	1/7	1
4	1/9	2
5	1/9	2
6	1/11	3

Closeness centrality

- Problems:
  - Only applicable to connected networks
  - Computationally expensive



d_{ij}	1	2	3	4	5	6
1		1	Inf.	Inf.	Inf.	Inf.
2	1		Inf.	Inf.	Inf.	Inf.
3	Inf.	Inf.		1	1	2
4	Inf.	Inf.	1		2	1
5	Inf.	Inf.	1	2		1
6	Inf.	Inf.	2	1	1	

Node	Closeness	Rank
1	1/Inf.	?
2	1/Inf.	?
3	1/Inf.	?
4	1/Inf.	?
5	1/Inf.	?
6	1/Inf.	?

Betweenness centrality

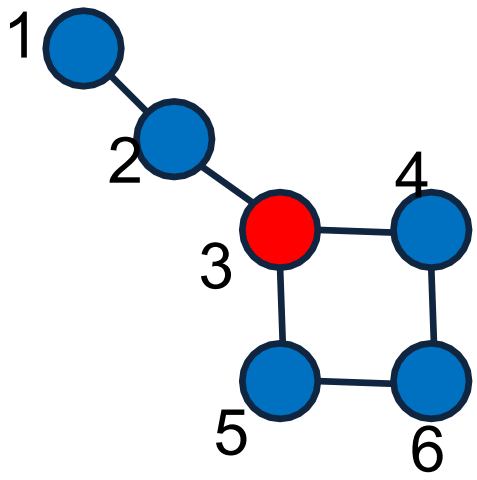
- A node is *central* if it many shortest **paths go through it**.

$$c_i = \sum_{j,k} \frac{g_{jk}(i)}{g_{jk}}$$

where

g_{jk} Shortest paths between j and k

$g_{jk}(i)$ Shortest paths between j and k passing through i



i=3	1	2	3	4	5	6
1		0/1	1/1	1/1	1/1	2/2
2	0/1		1/1	1/1	1/1	2/2
3	1/1	1/1		...	...	...
4	1/1	1/1	...		...	...
5	1/1	1/1	...	...		...
6	2/2	2/2	...	...	...	

shortest path matrix for $i = 3$

Betweenness centrality

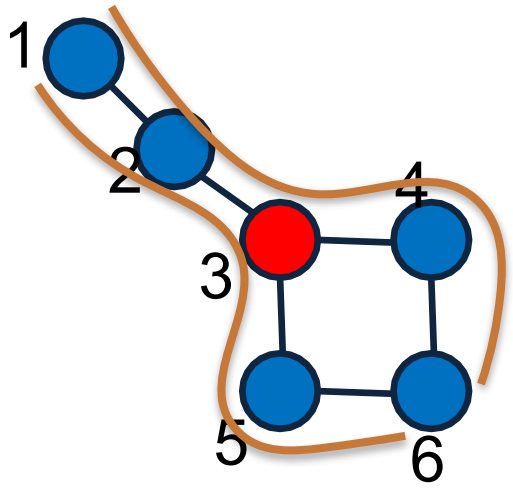
- A node is *central* if it many shortest **paths go through it**.

$$c_i = \sum_{j,k} \frac{g_{jk}(i)}{g_{jk}}$$

where

g_{jk} Shortest paths between j and k

$g_{jk}(i)$ Shortest paths between j and k passing through i



i=3	1	2	3	4	5	6
1		0/1	1/1	1/1	1/1	2/2
2	0/1		1/1	1/1	1/1	2/2
3	1/1	1/1		...	...	...
4	1/1	1/1	...		...	...
5	1/1	1/1	...	...		...
6	2/2	2/2	...	...	...	

shortest path matrix for $i = 3$

Betweenness centrality

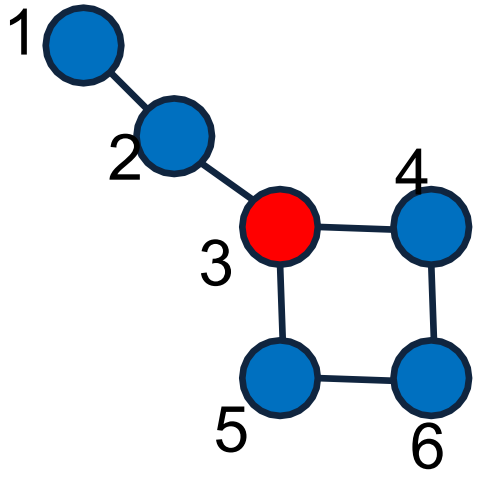
- A node is *central* if it many shortest **paths go through it**.

$$c_i = \sum_{j,k} \frac{g_{jk}(i)}{g_{jk}}$$

where

g_{jk} Shortest paths between j and k

$g_{jk}(i)$ Shortest paths between j and k passing through i



$\sum$

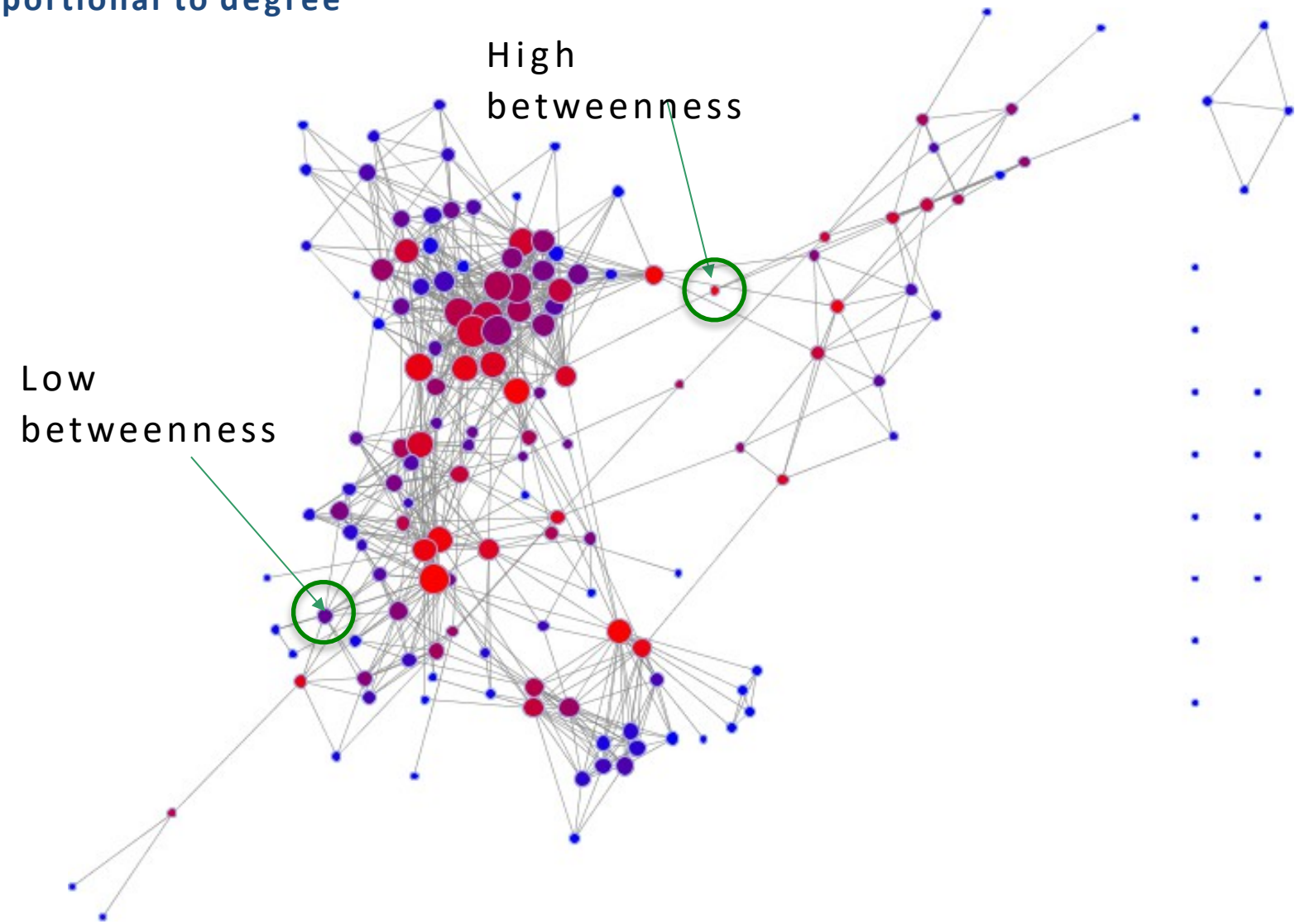
i=3	1	2	3	4	5	6
1		0/1	1/1	1/1	1/1	2/2
2	0/1		1/1	1/1	1/1	2/2
3	1/1	1/1		...	...	...
4	1/1	1/1	...		...	...
5	1/1	1/1	...	...		...
6	2/2	2/2	...	...	...	

shortest path matrix for $i = 3$

$\rightarrow c_3 = 16$

Betweenness centrality

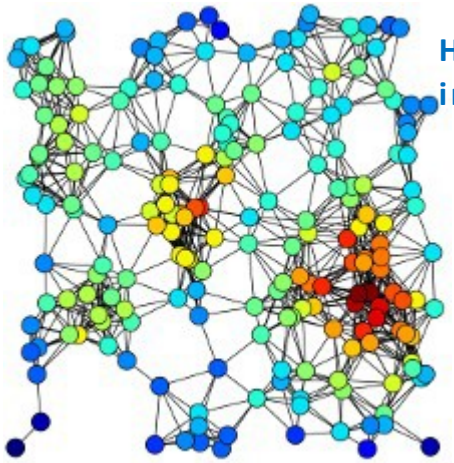
Color: proportional to betweenness
Size: proportional to degree



uc3m

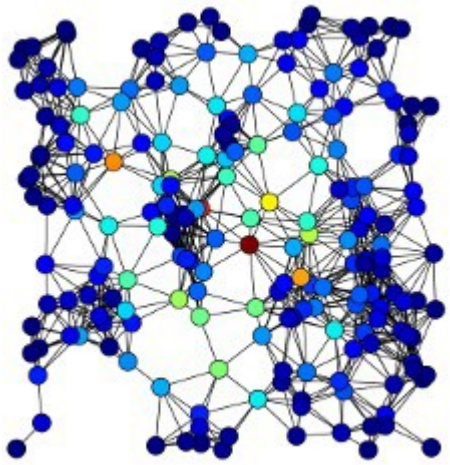
Central nodes according to different metrics

Degree centrality



How much connected in local- scale?

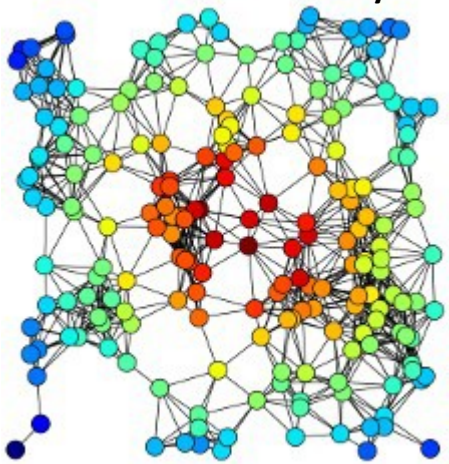
Betweenness centrality



How much connected in global-scale?

How closely connected?

Closeness centrality



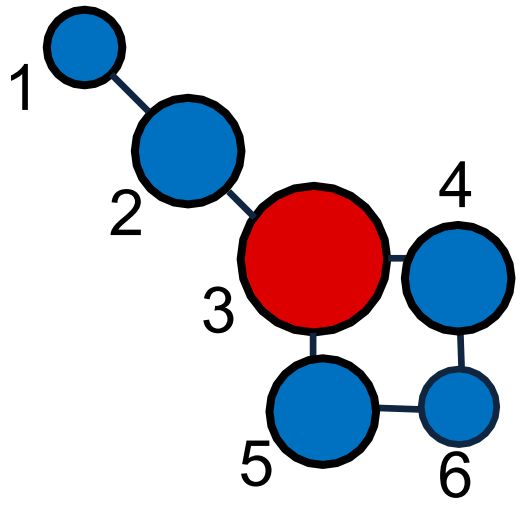
uc3m

Eigenvector centrality

- a node is *central* if its connected to **other central nodes** (recursive definition).

$$c_i = \frac{1}{\lambda} \sum_{j \in \mathcal{N}(i)} c_j$$

$\mathcal{N}(i)$: neighbors of i



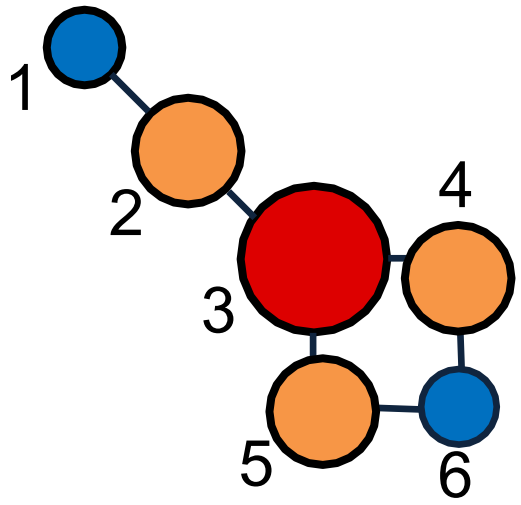
high-degree node

Eigenvector centrality

- a node is *central* if its connected to **other central nodes** (recursive definition).

$$c_i = \frac{1}{\lambda} \sum_{j \in \mathcal{N}(i)} c_j$$

$\mathcal{N}(i)$: neighbors of i



high-degree node

high-degree node connected to them

uc3m

Eigenvector centrality

- a node is *central* if its connected to **other central nodes** (recursive definition).

$$c_i = \frac{1}{\lambda} \sum_{j \in \mathcal{N}(i)} c_j$$

$\mathcal{N}(i)$: neighbors of i

- **Mathematical details**. Note that

$$\mathbf{c} = [c_1, c_2, \ldots, c_N]$$

$$\sum_{j \in \mathcal{N}(i)} c_j = \sum_j A_{ij} c_j \implies \lambda c_i = \sum_j A_{ij} c_j \implies \lambda \mathbf{c} = \mathbf{A}\mathbf{c}$$

"Eigenvalue problem"

Eigenvector centrality

- High eigenvector centrality indicates:
  - Highly connected nodes
  - Nodes with many high-degree connections

B: High in-degree

C: connection from B

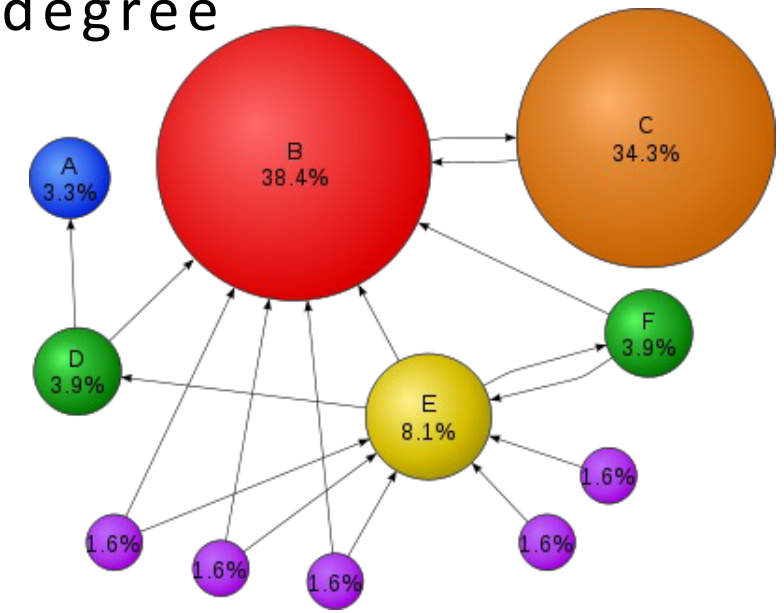


Image from: en.wikipedia.org/wiki/PageRank

Example: centrality predicts new elected pope

Pope Francis died early this year, so the Conclave had to open to elect a new pope.

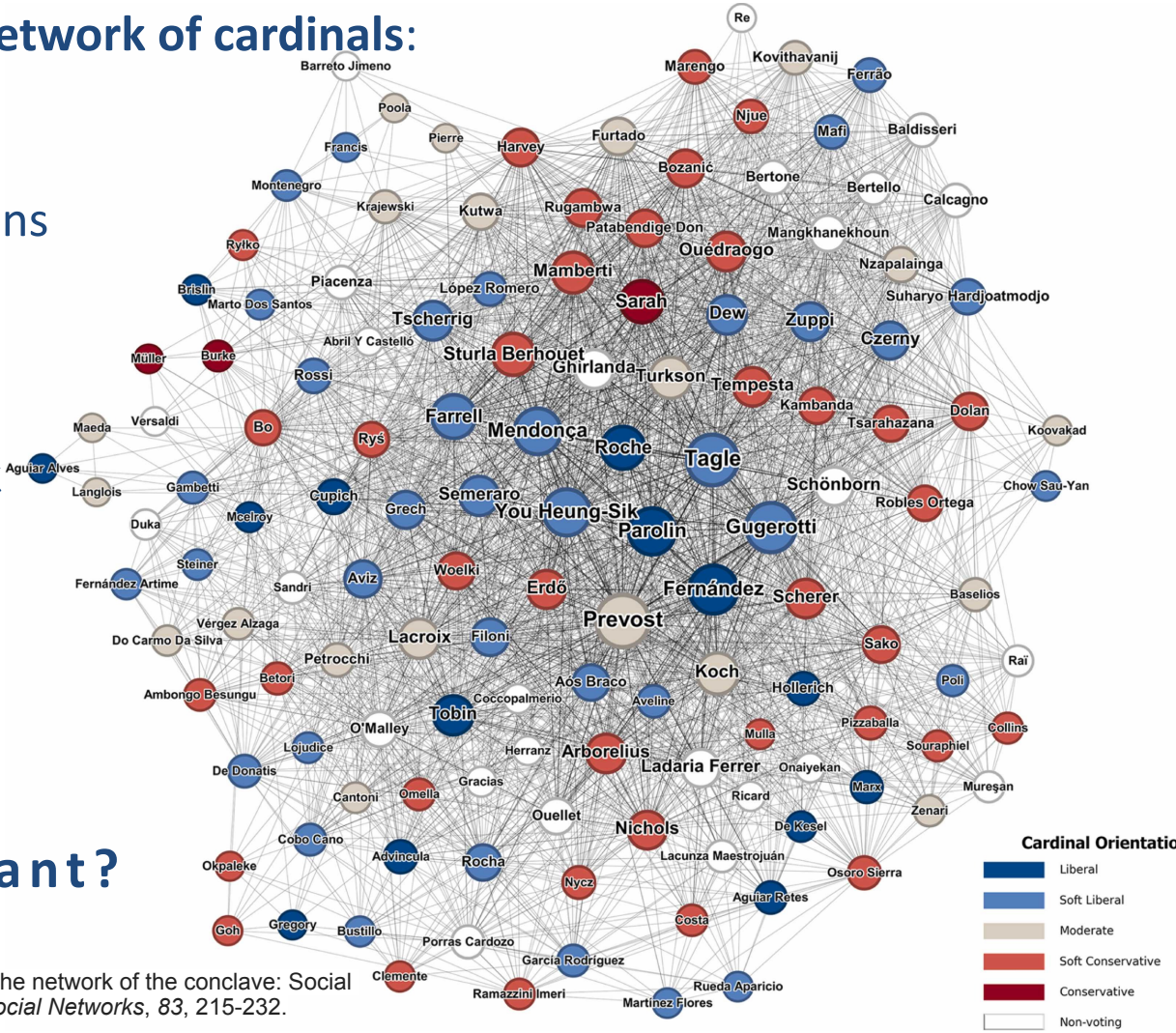
Scholars in Italy created a **social network of cardinals**:

- **Multi-layered**
  1. Consecration relationships
  2. Co-membership at institutions

Before the election, they made a **centrality analysis** of the network

- **Eigenvector** represents “Status"
- **Betweenness** represents “Information Control”

Visually, who is important?



Soda, G., Iorio, A., & Rizzo, L. (2025). In the network of the conclave: Social connections and the making of a pope. *Social Networks*, *83*, 215-232.

Example: centrality predicts new elected pope

https://www.unibocconi.it/en/news/network-conclave

Top 5 by Status

1. Robert Prevost *(moderate, US)*
2. Lazzaro You Heung-sik *(soft liberal, South Korea)*
3. Arthur Roche *(liberal, UK)*
4. Jean-Marc Aveline *(soft liberal, France)*
5. Claudio Gugerotti *(soft liberal, Italy)*

Top 5 for Information Control

1. Anders Arborelius *(soft conservative, Sweden)*
2. Pietro Parolin *(liberal, Italy)*
3. Víctor Fernández *(liberal, Argentina)*
4. Gérald Lacroix *(moderate, Canada)*
5. Joseph Tobin *(liberal, USA)*

- They find that **Prevost** is the cardinal with the highest Status
- They create a structural ranking of all cardinals according to their centrality
- Different centralities give different rankings

2 Network models

Network models

Often, **real networks** often exhibit common properties:

- Degree distributions (power-law → hubs)
- High clustering (lots of mutual connections)
- Small world (small network diameter)
- Degree assortativity (rich gets richer)
- Sparsity (most connections don't happen)

We would like to understand if:

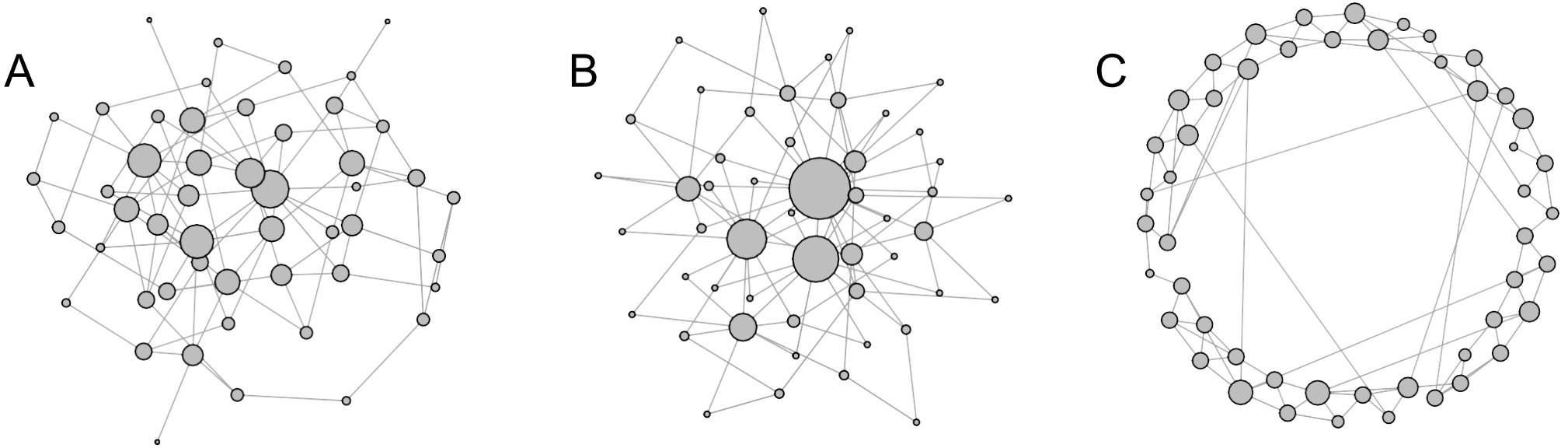
- These properties are **statistically significant**
- They can be explained by **simple mechanisms**

Network models generate networks with controlled properties.

Network models

We are going to consider several statistic models of networks (random networks):

- The **Erdös-Rényi** model (total randomness; controlled sparsity) (A)
- The **Configuration** model (controlled degree distribution) (B)
- The **Watts-Strogatz** model (small-world with high clustering) (C)

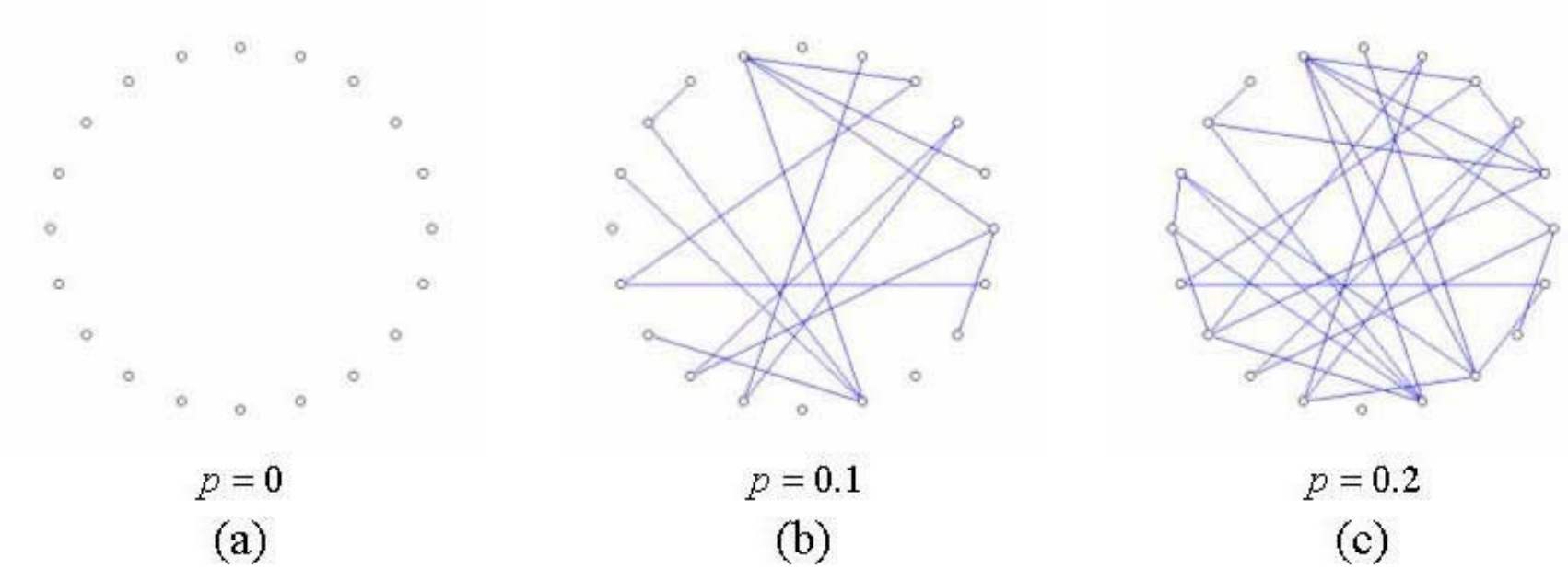


Erdös-Rényi (ER) model

Proposed by mathematicians Erdös and Rényi in 1959

Definition

- Start with **N** nodes and no edges.
- For each pair of nodes, add an edge with probability **p**.
- Each run produces a **different random network**.



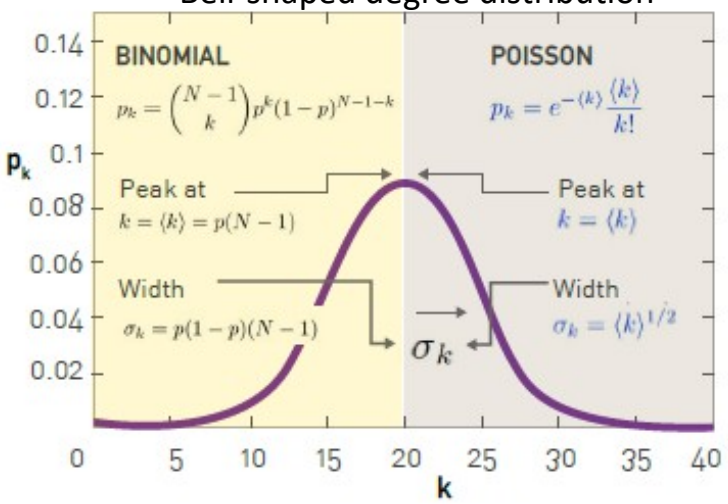
Erdös-Rényi model

Some properties

- Average degree: $\langle k \rangle = p(N-1)$
- Average number of links: $\langle L \rangle = \frac{1}{2}\langle k \rangle N = pL_{max}$
- Clustering coefficient: $C = \frac{\langle k \rangle}{N} \approx p$
- Distances: $\langle d \rangle \approx \frac{\ln N}{\ln \langle k \rangle}$
- Degree distribution: binomial
  - If $\langle k \rangle \ll N$ (sparse) $\rightarrow$ Poisson

$$P(k) \approx e^{\langle k \rangle} \frac{\langle k \rangle^{\langle k \rangle}}{\langle k \rangle !}$$

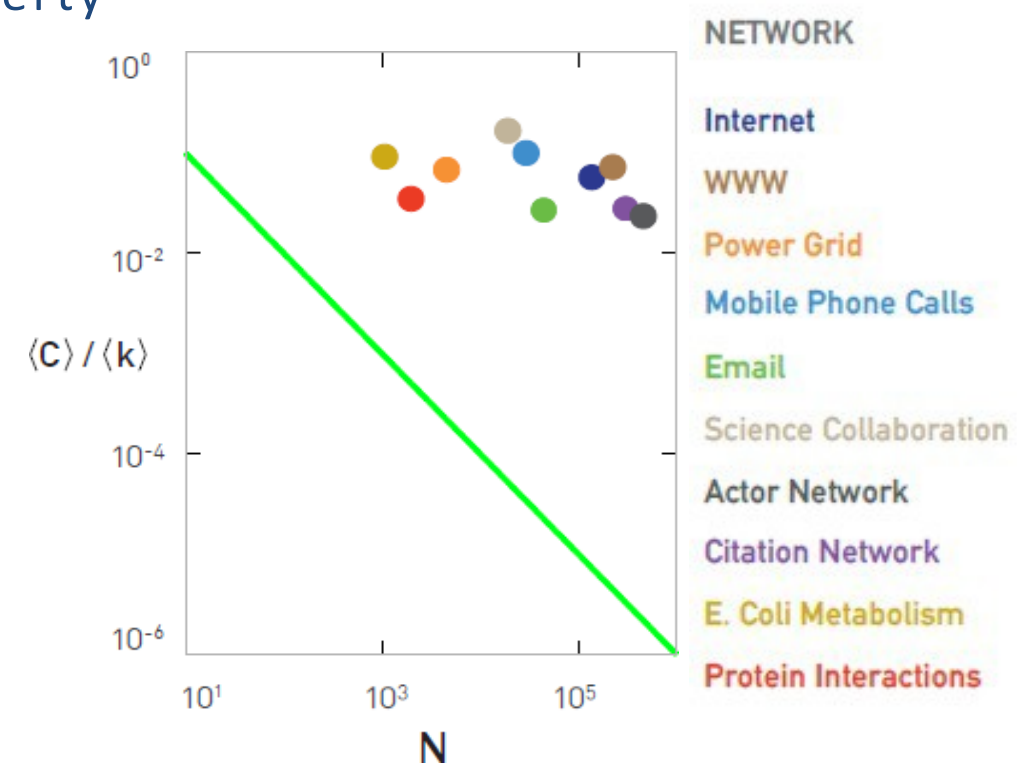
Bell-shaped degree distribution



Erdös-Rényi model

Is it a good model for real networks?

- They have the **small world** property
- Degree distribution:
  - **ER**: exponential
  - **Real**: heavy tailed
- Clustering:
  - **ER**: small ($\sim p$)
  - **Real**: high



Configuration model

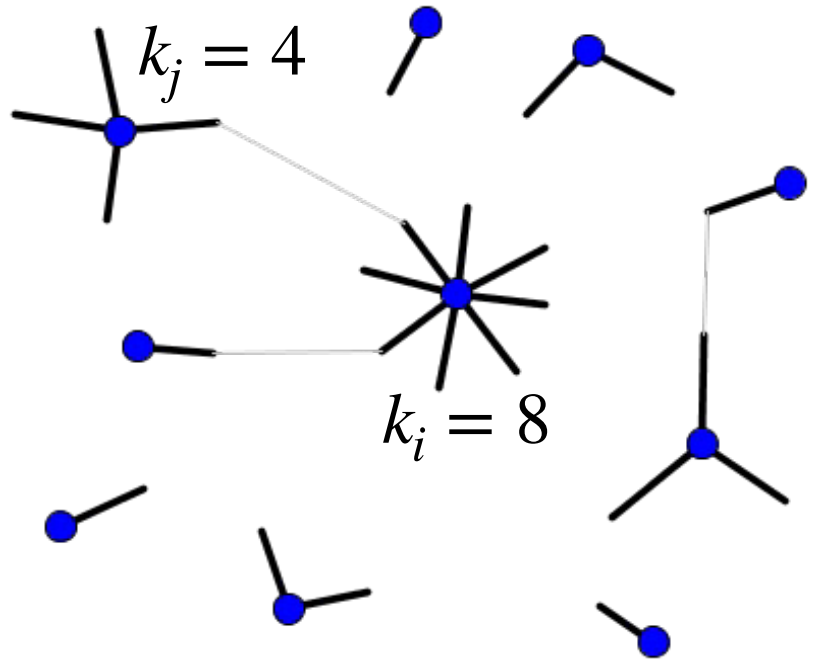
Proposed by mathematician Béla Bobollás in 1980.

Definition

- Start with **N** nodes each with prescribed degree k_i.
- Connect nodes at random while **preserving their degrees**.
- High-degree nodes are more likely to connect:

$$p_{ij} \approx \frac{k_i k_j}{2L}$$

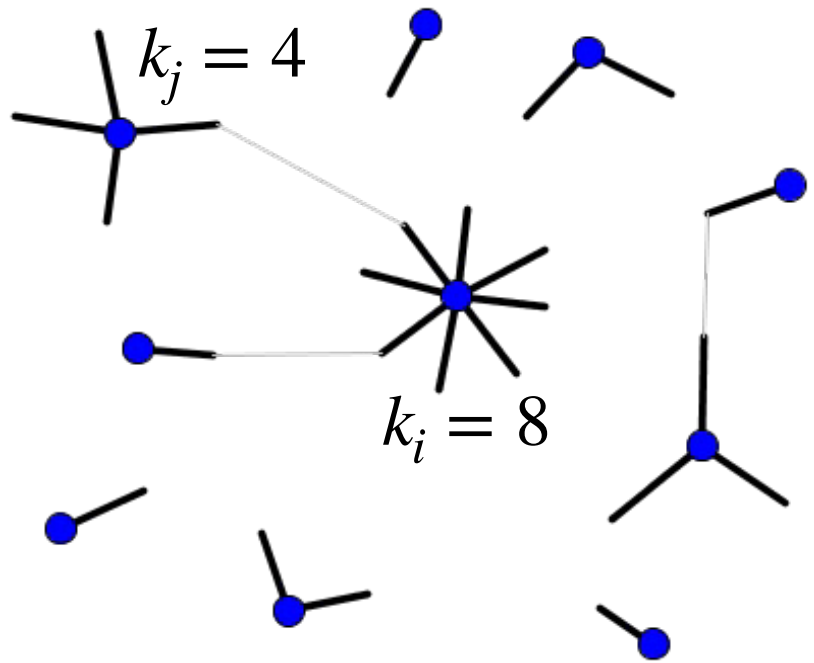
- Produces a random network that **keeps the degree sequence** but randomises everything else.



Configuration model

Some properties

- Keeps **degree distribution** $P(k)$ intact.
- Average degree **intact**.
- Number of links also **intact**.
- Clustering coefficient: $C \to 0$ as N grows.
- Assortativity: typically $r \approx 0$
- Average distance: $\langle d \rangle \approx \dfrac{\ln N}{\ln \langle k \rangle}$

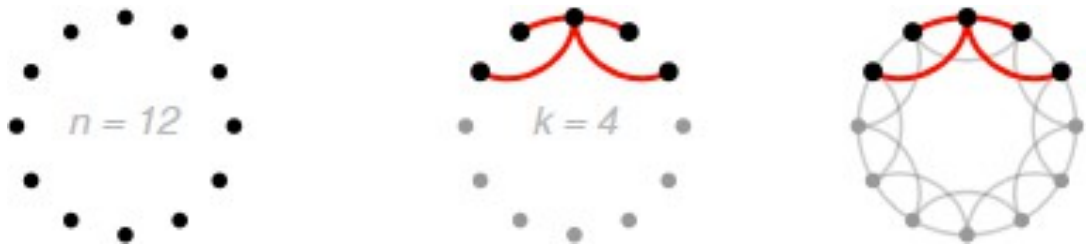


Watts-Strogatz model

Previous models show the **small-world** property but very **low clustering**. Can we have both? Watts-Strogatz model of small-world (1998).

Definition

- Combines a **regular network** (same degrees) with a **random network**
- Start with **N** nodes each connected with its **k** nearest neighbours (high clustering)



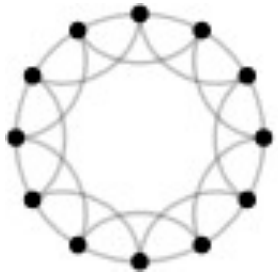
- With probability **p**, rewire an edge at random (this creates shortcuts → small world)

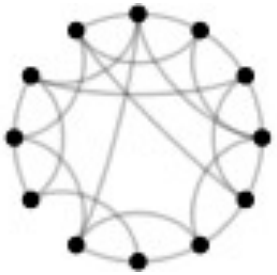
uc3m

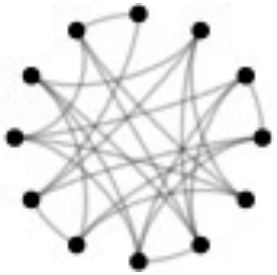
Watts-Strogatz model

Some properties

- $p = 0 \rightarrow$ regular network (no shortcuts, high clustering)
- $p = 1 \rightarrow$ Erdos-Renyi network (all shortcuts, no clustering)

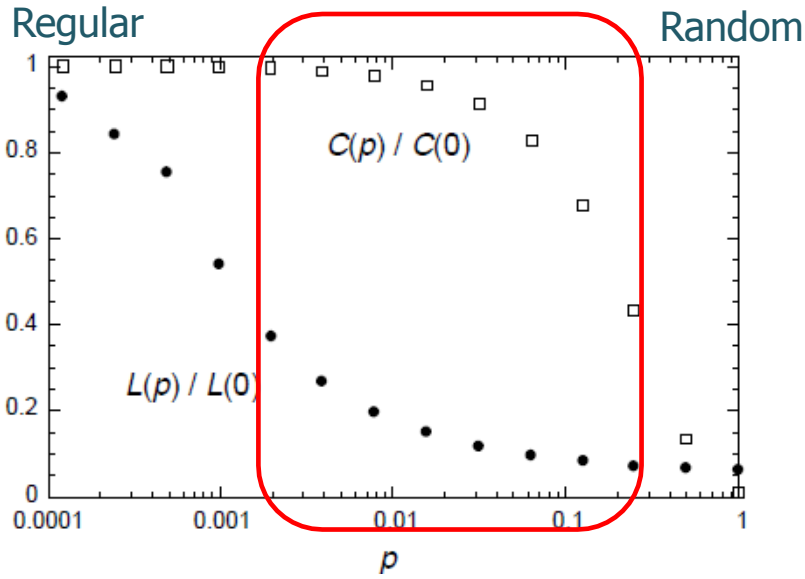






- Intermediate p:
  high clustering & small world

- However, **degree distribution decays exponentially**



uc3m

3 Community structure

Graphs at different scales

So far, we have explored:

- **Global properties** of a network:
  - Number of nodes/links
  - Clustering coefficient
  - Assortativity coefficient
  - Density
- **Local properties** of a network:
  - Node centralities

Graphs at different scales

So far, we have explored:

- **Global properties** of a network:
  - Number of nodes/links
  - Clustering coefficient
  - Assortativity coefficient
  - Density
- **Local properties** of a network:
  - Node centralities

What about meso-level properties?

- Community structure
- Cliques
- Cores
- Motifs

Graphs at different scales

So far, we have explored:

- **Global properties** of a network:
  - Number of nodes/links
  - Clustering coefficient
  - Assortativity coefficient
  - Density
- **Local properties** of a network:
  - Node centralities

What about meso-level properties?

- **Community structure**
- Cliques
- Cores
- Motifs

Swiss politics support network

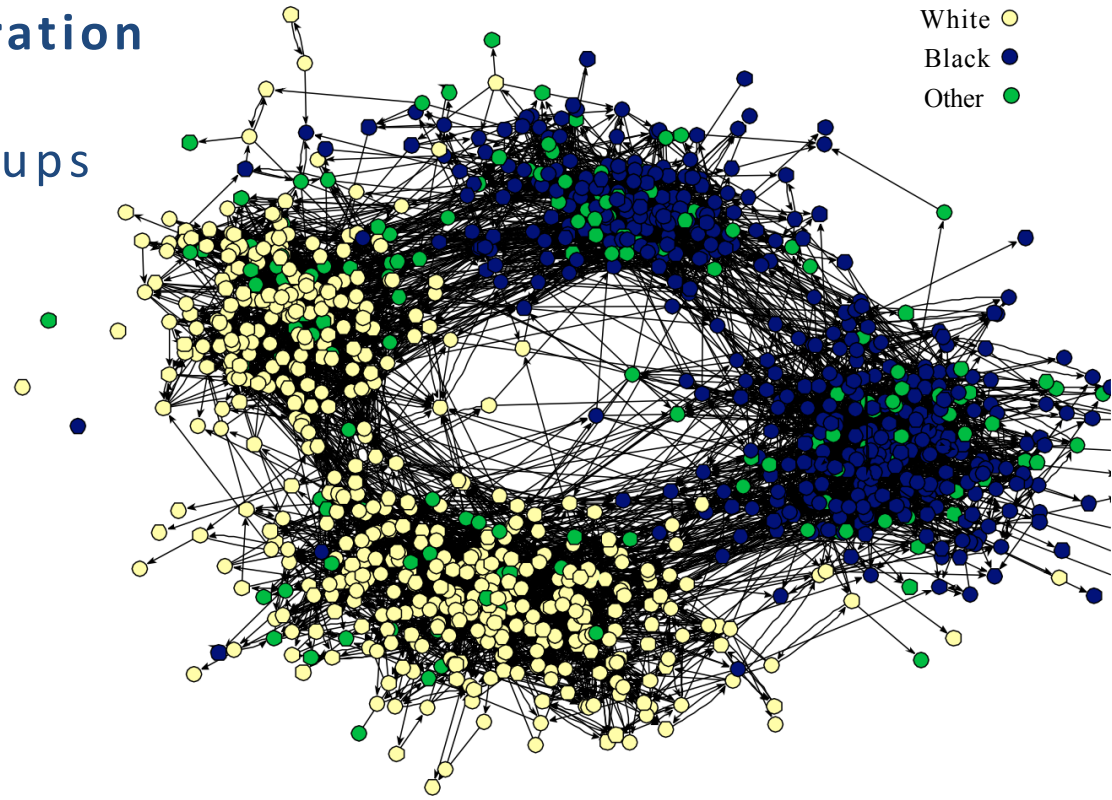
Garcia, David, et al. "Ideological and temporal components of network polarization in online political participatory media." Policy & internet 7.1 (2015): 46-79.

uc3m

Communities

At a **meso-scale**, networks tend to form **groups**

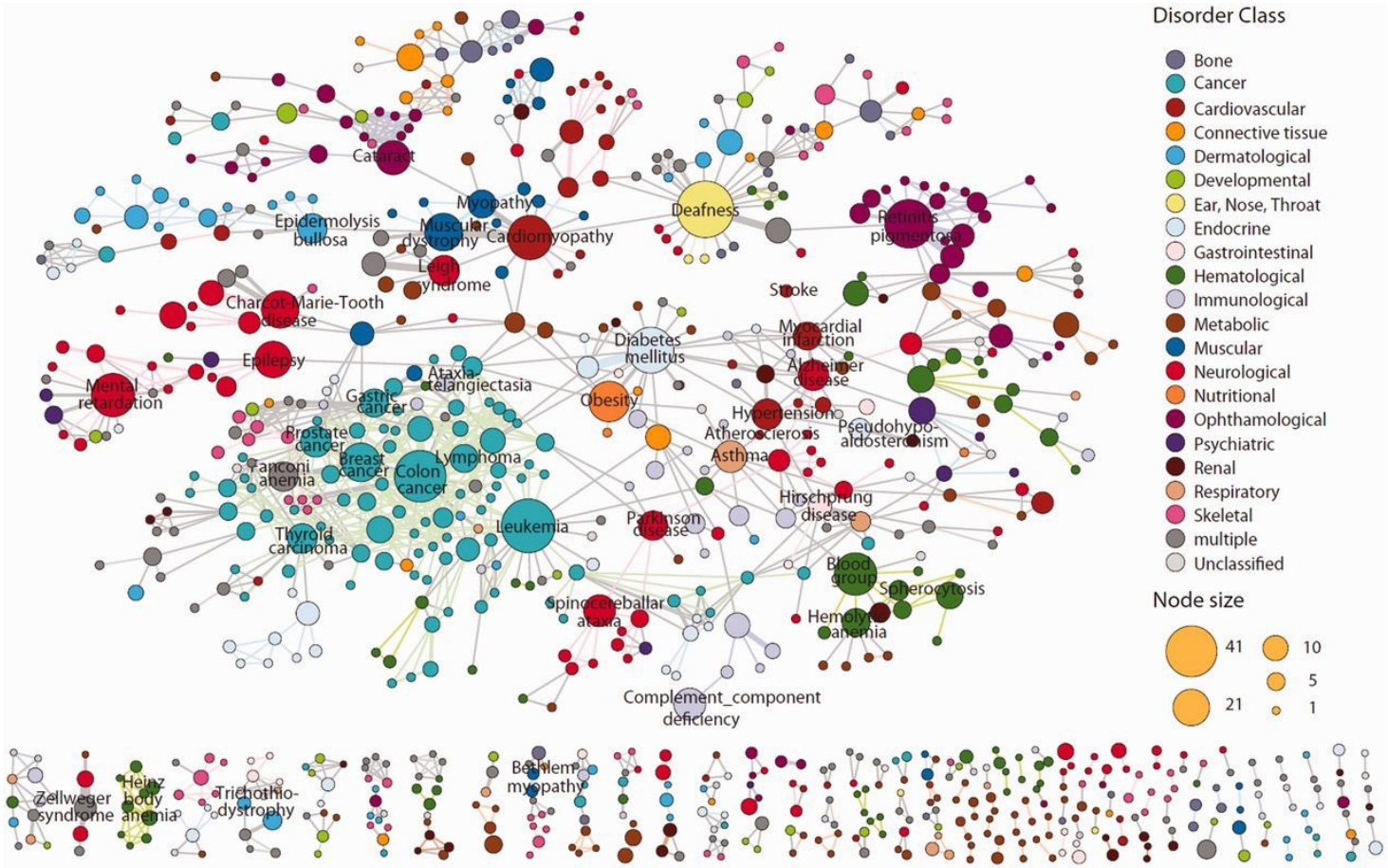
- The distribution of edges may depend on social, spatial, or functional characteristics.
- Edges have a **high concentration within** groups and a **low concentrations between** groups



Friendship relationships in a high-school. Moody, (2001)

Communities

- Example: diseasome co-gene network

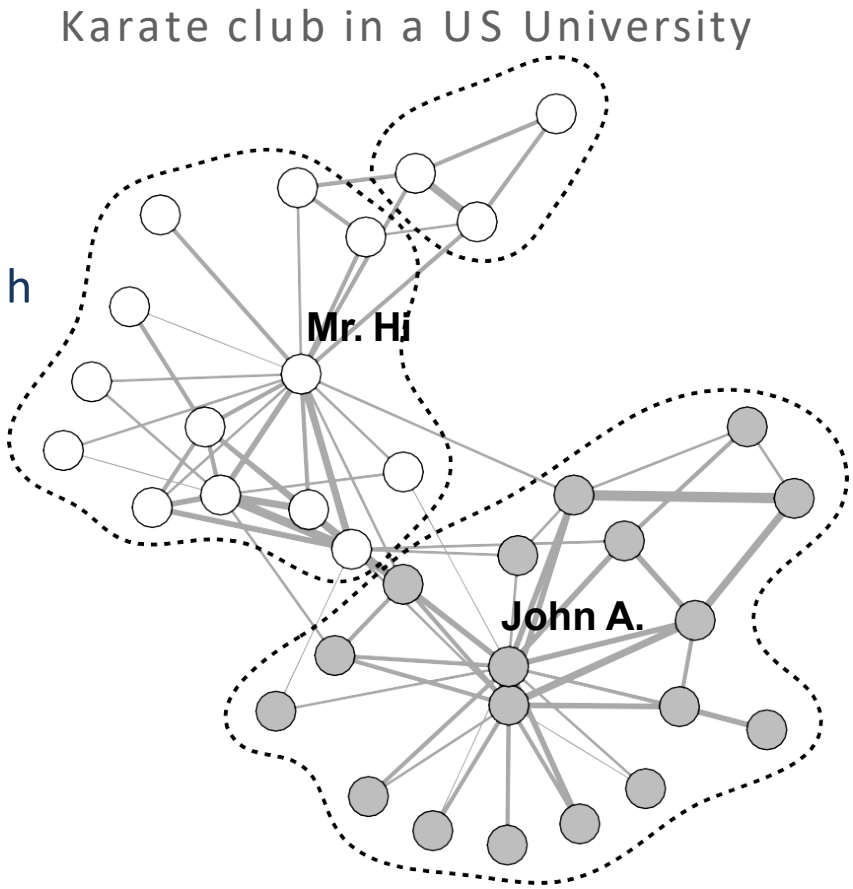


Diseasome: interaction network between diseases.
Two diseases are connected if they share a genetic component

Communities

What is a community?

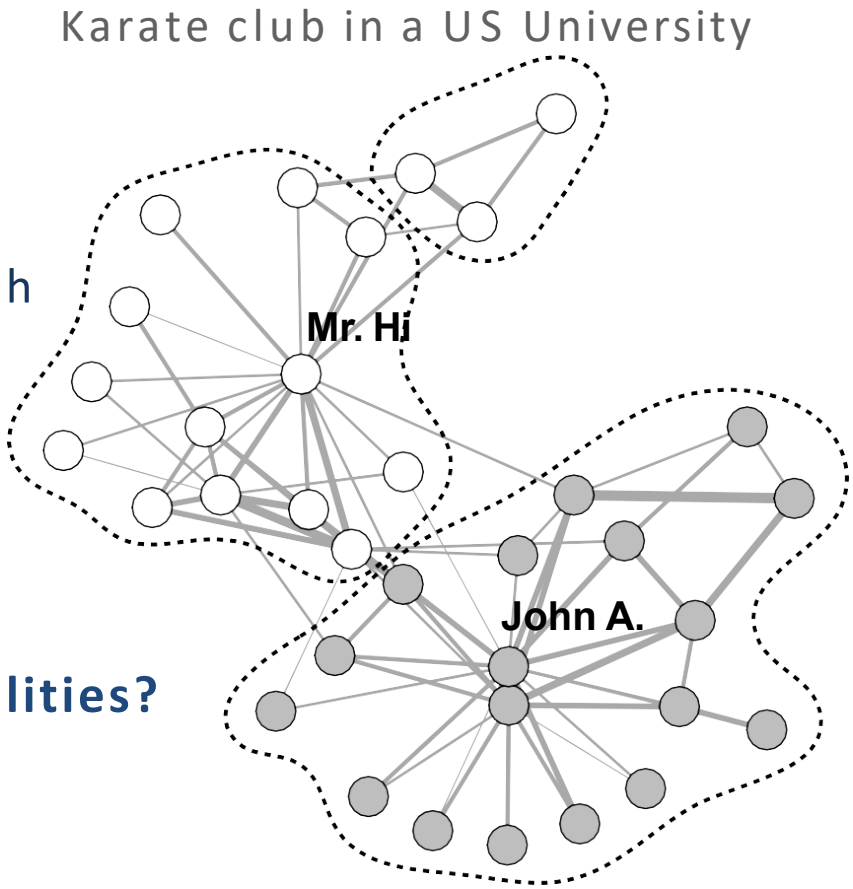
- Densely connected subgraphs
- Based on connectedness and density hypothesis
  - **Connectedness**: members in a community must be *reachable* through any other member.
  - **Density**: nodes within a community should have a *higher connection probability* than between communities
- Can we discover latent communities?



Communities

What is a community?

- Densely connected subgraphs
- Based on connectedness and density hypothesis
  - **Connectedness**: members in a community must be *reachable* through any other member.
  - **Density**: nodes within a community should have a *higher connection probability* than between communities
- Can we discover latent communities?



How can we determine **connection probabilities?**

$$p_{ij} = \frac{k_i k_j}{2L}$$

Configuration model

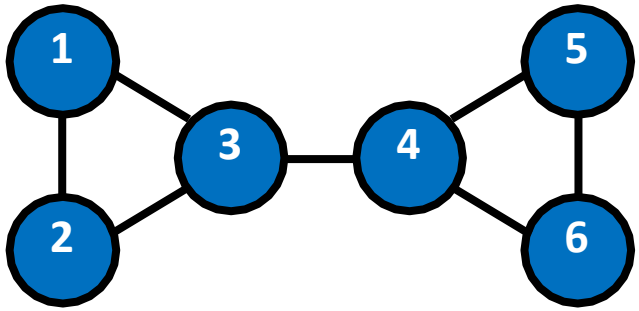
uc3m

Communities

Compare **empirical** vs **expected** structure

$$p_{ij} = \frac{k_i k_j}{2L}$$

configuration model

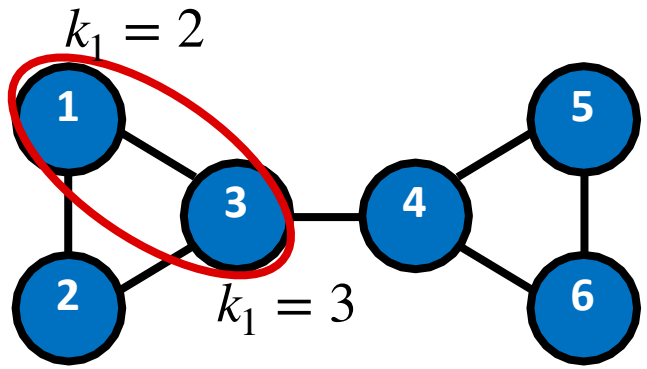


$L = 7$ edges

$$A_{ij} - p_{ij}$$

Communities

Compare **empirical** vs **expected** structure



$L = 7$ edges

$$A_{ij} - p_{ij}$$

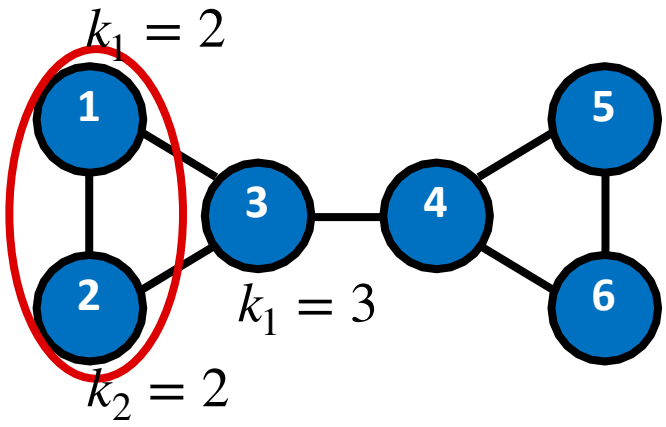
$$p_{ij} = \frac{k_i k_j}{2L}$$

configuration model

1 - 6/14 = 0.57

Communities

Compare **empirical** vs **expected** structure



$$A_{ij} - p_{ij}$$

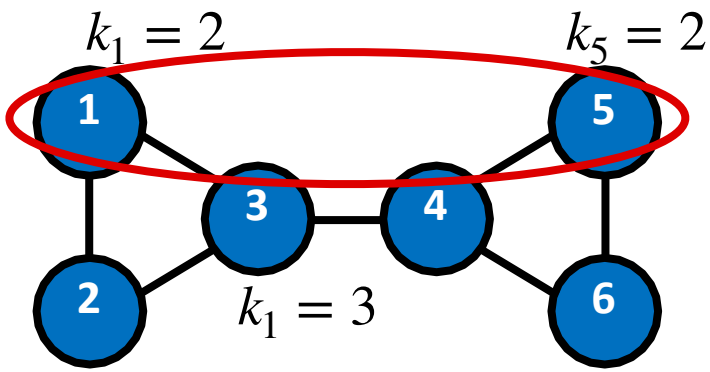
$$p_{ij} = \frac{k_i k_j}{2L}$$

configuration model

1 - 4/14 = 0.72

Communities

Compare **empirical** vs **expected** structure



$L = 7$ edges

$$A_{ij} - p_{ij}$$

$$p_{ij} = \frac{k_i k_j}{2L}$$

configuration model

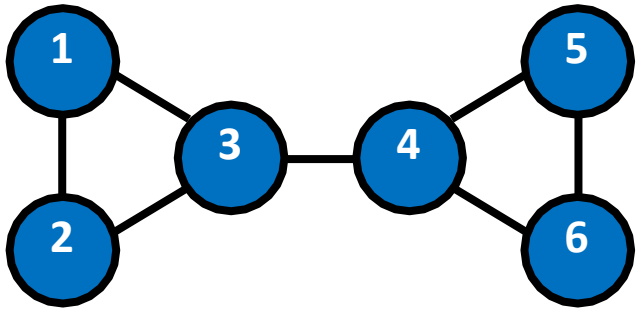
0 - 4/14 = -0.28

Communities

Compare **empirical** vs **expected** structure

$$p_{ij} = \frac{k_i k_j}{2L}$$

configuration model



$L = 7$ edges

$$A_{ij} - p_{ij}$$

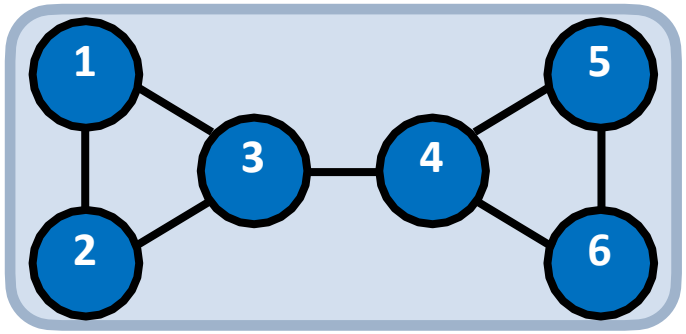
This way, we can aggregate all comparisons

Communities

Compare **empirical** vs **expected** structure

$$p_{ij} = \frac{k_i k_j}{2L}$$

configuration model



$$A_{ij} - p_{ij}$$

This way, we can aggregate all comparisons

$L = 7$ edges

Problem: the overall sum is 0.

$$\sum_{ij} (A_{ij} - p_{ij}) = \sum_{ij} A_{ij} - \frac{1}{2L} \sum_{i} k_i \sum_{j} k_j$$

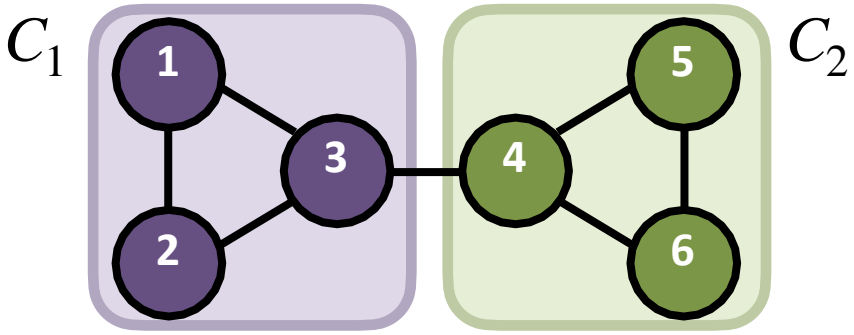
$$= 2L - \frac{1}{2L}(2L)^2 = 2L - 2L = 0$$

Communities

Compare **empirical** vs **expected** structure

$$p_{ij} = \frac{k_i k_j}{2L}$$

configuration model



$L = 7$ edges

However, the sum **over sections** is not 0.

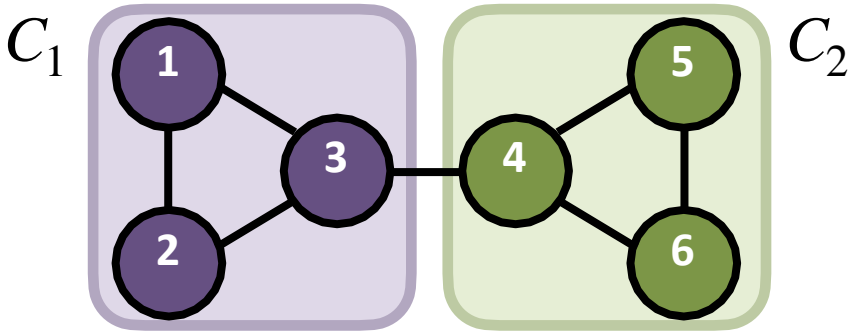
$$\sum_{ij \in C_1} (A_{ij} - p_{ij}) = 2.01 \quad + \quad \sum_{ij \in C_2} (A_{ij} - p_{ij}) = 2.01$$

Communities

Compare **empirical** vs **expected** structure

$$p_{ij} = \frac{k_i k_j}{2L}$$

configuration model



$L = 7$ edges

Network partition

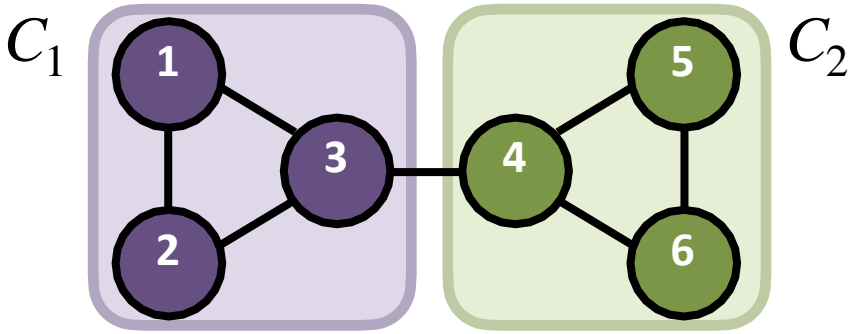
$$\mathscr{C} = [\underbrace{\{1,2,3\}}_{C_1}, \underbrace{\{4,5,6\}}_{C_2}]$$

However, the sum **over sections** is not 0.

$$\sum_{ij \in C_1} (A_{ij} - p_{ij}) = 2.01 \quad + \quad \sum_{ij \in C_2} (A_{ij} - p_{ij}) = 2.01$$

Communities: Modularity

Compare **empirical** vs **expected** structure



$L = 7$ edges

Network partition

$$\mathscr{C} = [\underbrace{\{1,2,3\}}_{C_1}, \underbrace{\{4,5,6\}}_{C_2}]$$

Modularity: clustering score of partition $\mathscr{C}$.

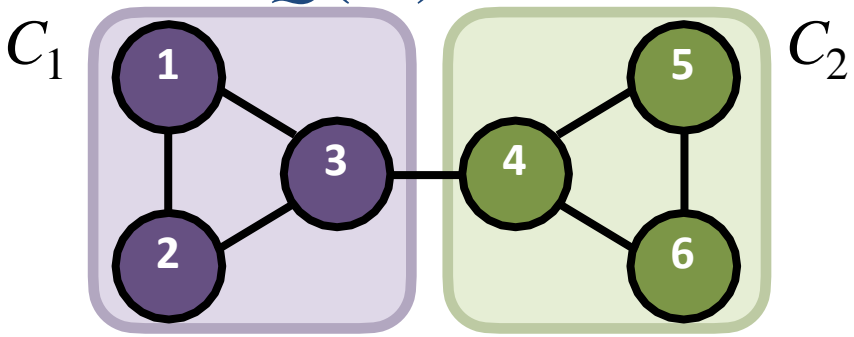
$$Q(\mathscr{C}) = \frac{1}{2L} \sum_{C \in \mathscr{C}} \sum_{ij \in C} \left(A_{ij} - \frac{k_i k_j}{2L} \right) \in [-0.5, 1]$$

normalization

Communities: Modularity

Compare **empirical** vs **expected** structure

$$Q(\mathscr{C}) = 0.35$$



$L = 7$ edges

Network partition

$$\mathscr{C} = [\underbrace{\{1,2,3\}}_{C_1}, \underbrace{\{4,5,6\}}_{C_2}]$$

Modularity: clustering score of partition $\mathscr{C}$.

$$Q(\mathscr{C}) = \frac{1}{2L} \sum_{C \in \mathscr{C}} \sum_{ij \in C} \left(A_{ij} - \frac{k_i k_j}{2L} \right) \in [-0.5, 1]$$

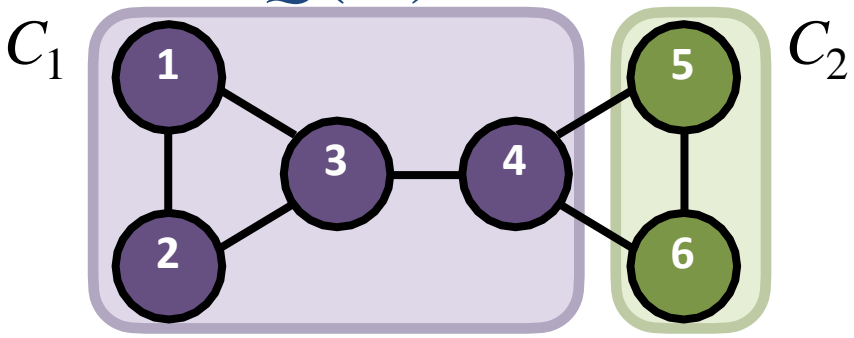
normalization

uc3m

Communities: Modularity

Compare **empirical** vs **expected** structure

$$Q(\mathscr{C}) = 0.122$$



$L = 7$ edges

Network partition

$$\mathscr{C} = [\underbrace{\{1,2,3,4\}}_{C_1}, \underbrace{\{5,6\}}_{C_2}]$$

Modularity: clustering score of partition $\mathscr{C}$.

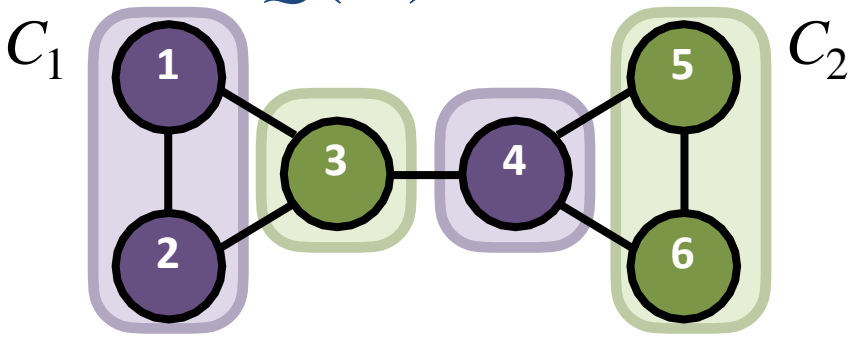
$$Q(\mathscr{C}) = \frac{1}{2L} \sum_{C \in \mathscr{C}} \sum_{ij \in C} \left(A_{ij} - \frac{k_i k_j}{2L} \right) \in [-0.5, 1]$$

normalization

Communities: Modularity

Compare **empirical** vs **expected** structure

$$Q(\mathscr{C}) = -0.21$$



$L = 7$ edges

Network partition

$$\mathscr{C} = [\underbrace{\{1,2,4\}}_{C_1}, \underbrace{\{3,5,6\}}_{C_2}]$$

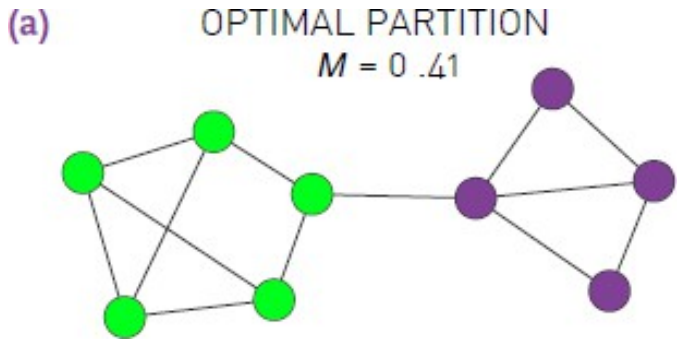
Modularity: clustering score of partition $\mathscr{C}$.

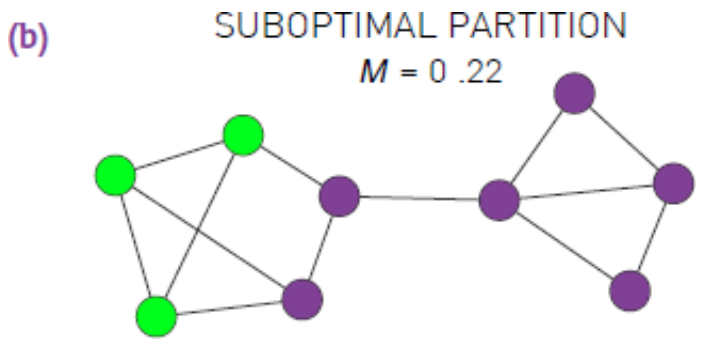
$$Q(\mathscr{C}) = \frac{1}{2L} \sum_{C \in \mathscr{C}} \sum_{ij \in C} \left(A_{ij} - \frac{k_i k_j}{2L} \right) \in [-0.5, 1]$$

normalization

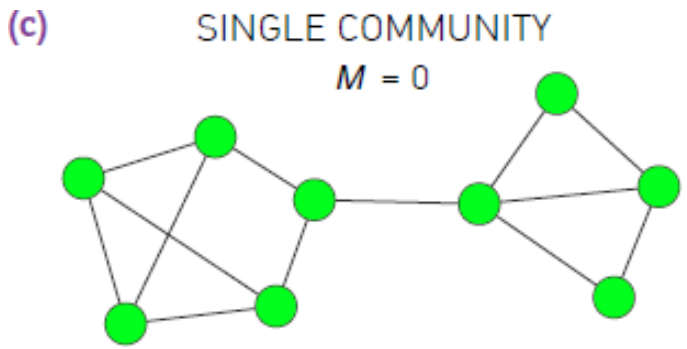
Communities: Modularity

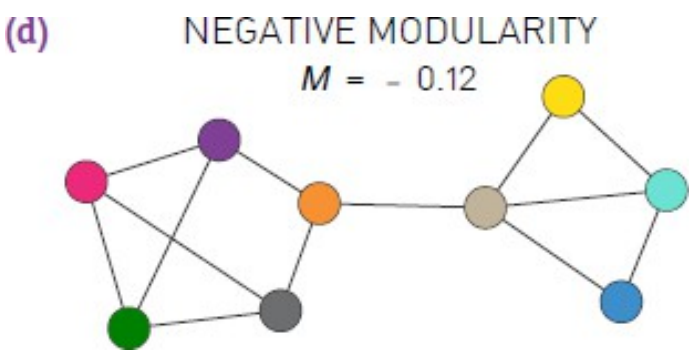
- Higher modularity implies better partition (Fig. (a) and (b))





- Zero and Negative modularity





Communities: Modularity

Detecting communities →

Optimal partition $\mathscr{C}$ that maximises modularity Q

Communities: Modularity

Detecting communities →

Optimal partition $\mathscr{C}$ that maximises modularity Q

NAME	NATURE	COMP.
Ravasz	Hierarchical Agglomerative	$O(N^2)$
Girvan-Newman	Hierarchical Divisive	$O(N^3)$
Greedy Modularity	Modularity Optimization	$O(N^2)$
Greedy Modularity (Optimized)	Modularity Optimization	$O(N\log^2 N)$
Louvain	Modularity Optimization	O(L)
Infomap	Flow Optimization	O(NlogN)
Clique Percolation (CFinder)	Overlapping Communities	Exp(N)
Link Clustering	Hierarchical Agglomerative; Overlapping Communities	$O(N^2)$

Communities: Modularity

Detecting communities →

Optimal partition $\mathscr{C}$ that maximises modularity Q

NAME	NATURE	COMP.
Ravasz	Hierarchical Agglomerative	$O(N^2)$
Girvan-Newman	Hierarchical Divisive	$O(N^3)$
Greedy Modularity	Modularity Optimization	$O(N^2)$
Greedy Modularity (Optimized)	Modularity Optimization	$O(N\log^2 N)$
Louvain	Modularity Optimization	O(L)
Infomap	Flow Optimization	O(NlogN)
Clique Percolation (CFinder)	Overlapping Communities	Exp(N)
Link Clustering	Hierarchical Agglomerative; Overlapping Communities	$O(N^2)$

Scale well with big networks

Other community detection algorithms

Other algorithms:

1. **Modularity optimization algorithms**
   - Most known is the **Louvain method**, based on **local optimization of modularity**.
     - **Louvain:** Blondel, Vincent D., et al. "Fast unfolding of communities in large networks." *Journal of statistical mechanics: theory and experiment* 2008.10 (2008): P10008. (Most common method nowadays!)

2. **Random walk algorithms**
   - Idea: a random Walker tends to stay within a community
   - Examples
     - **Infomap**: Maps of information flow reveal community structure in complex networks, PNAS 105, 1118 (2008)
     - **Walktrap**: Pascal Pons, Matthieu Latapy: Computing communities in large networks using random walks,

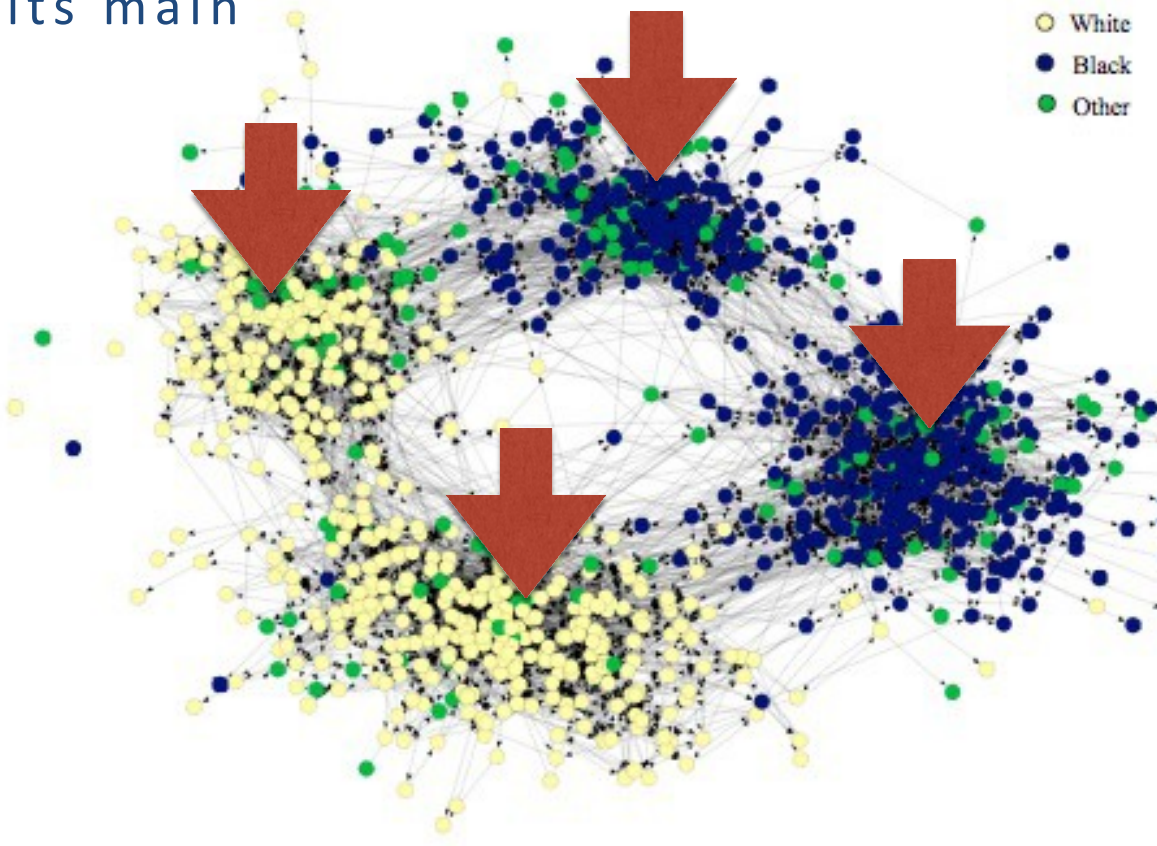
3. **Clique-based algorithms**
   - Idea: cliques tend to be within communities and do overlap
   - Examples:
     - **Clique percolation method**: a community is an overlapping set of cliques. G. Palla, I. Derényi, I. Farkas and T. Vicsek (2005). "Uncovering the overlapping community structure of complex networks in nature and society". Nature 435 (7043): 814–818

Communities

Example: **integration policies**

What are the segregation patterns?

- Find the communities
- In each community, find its main actors (centrality) and characteristics
- Policy design based on **meso-structure.**



uc3m

5 Summary

Summary of session 2

- **Centrality**
  - Provides several notions of importance for nodes in a network.
  - Reveals and ranks structural functions.
- **Network models**
  - Helps understanding mechanisms of network formation.
  - Used for making statistical tests of empirical network features.
- **Communities** are a peso-scale characteristic of networks
  - Meso-scale structure within a network.
  - Intuition: dense connections inside groups of nodes; sparse connections between them.
  - *Modularity* measures how strong those groups are.

EXTRA

Community algorithms

Since finding the optimal partition is NP hard, there are some algorithms to approximate it.

Most algorithms to calculate communities are based on hierarchical clustering

- **Agglomerative**: merging nodes with high similarity into communities
- **Divisive**: splitting communities into better partitions

  Where to stop?
- *When modularity is maximal*

Community algorithms

Agglomeration example: Ravasz algorithm

1. Define distance between nodes using "structural equivalence"

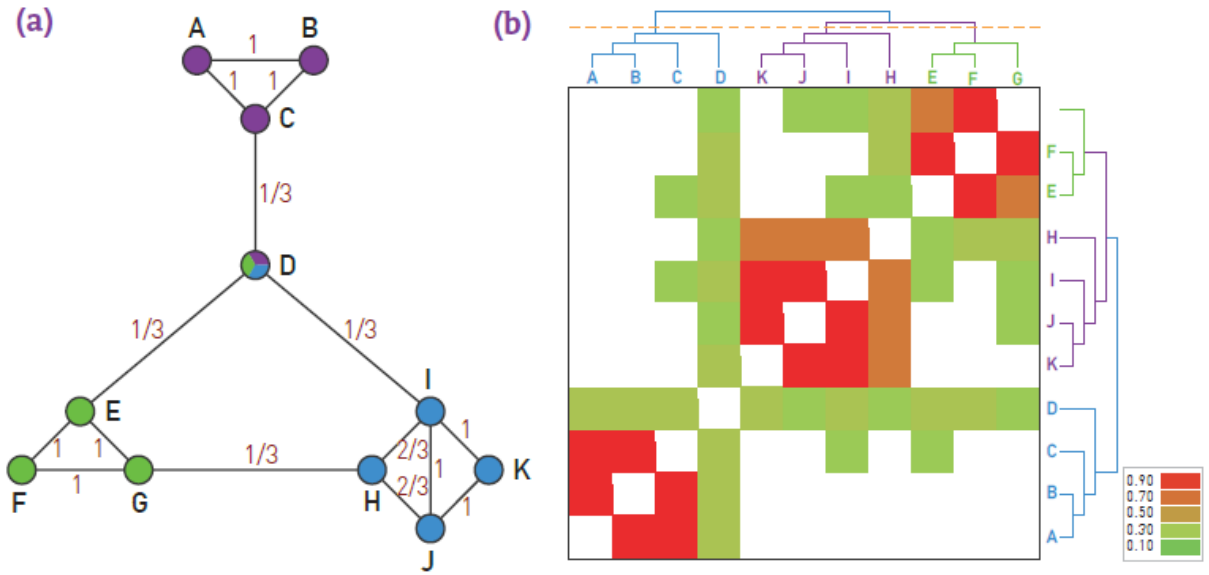
$$w_{ij} = \frac{|\mathcal{N}_i \cap \mathcal{N}_j|}{|\mathcal{N}_i \cup \mathcal{N}_j|}$$

High if i and j have many common neighbors

where $\mathcal{N}_i$ is the set of neighbors of i

2. Apply hierarchical clustering on the matrix $W = [w_{ij}]$.

3. Using the dendrogram, stop when modularity is maximal



Community algorithms

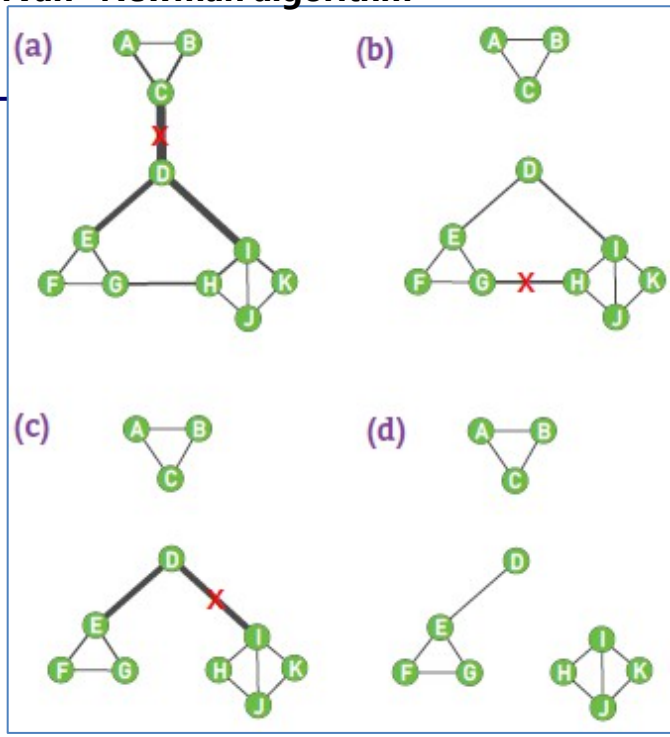
Other agglomerative algorithms:

- **Hierarchical clustering:**
  S. Wasserman, K. Faust, Social Networks Analysis, Cambridge University Press, Cambridge, 1994.

- **Fast greeedy modularity optimization:**
  Aaron Clauset, M. E. J. Newman, and Cristopher Moore: Finding community structure in very large networks, Phys. Rev. E 70, 066111 (2004)

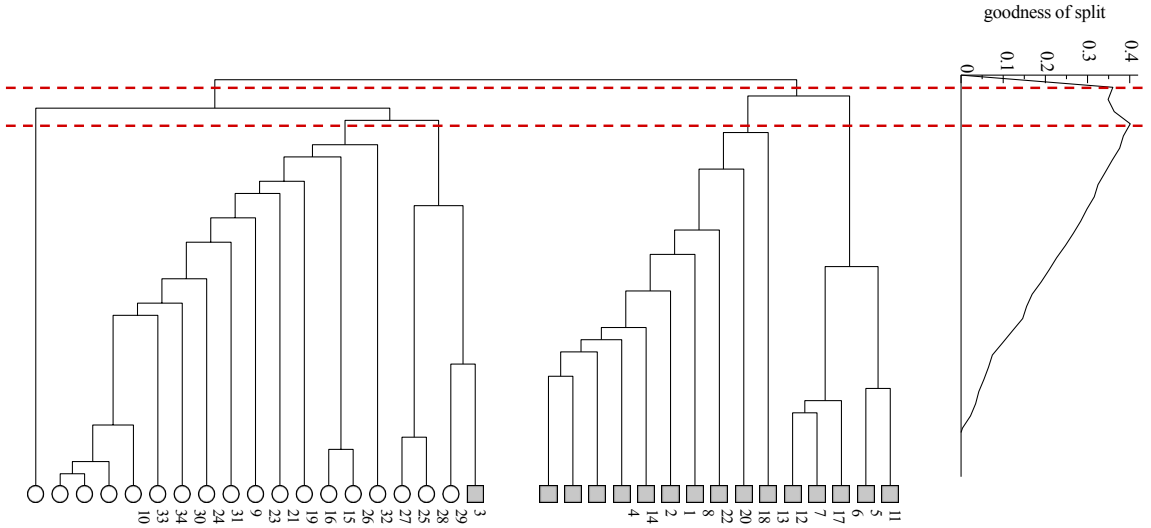
Community algorithms

Girvan –Newman algorithm



Divisive example: The Girvan-Newman algorithm

1. Start with all the nodes in one community.
2. Remove those links that are most likely to be between communities (low similarity or low betwenness)
3. Stop when modularity is maximal



uc3m