UNIVERSITAT POLITÈCNICA DE VALÈNCIA

VRAIN
Valencian Research Institute for Artificial Intelligence

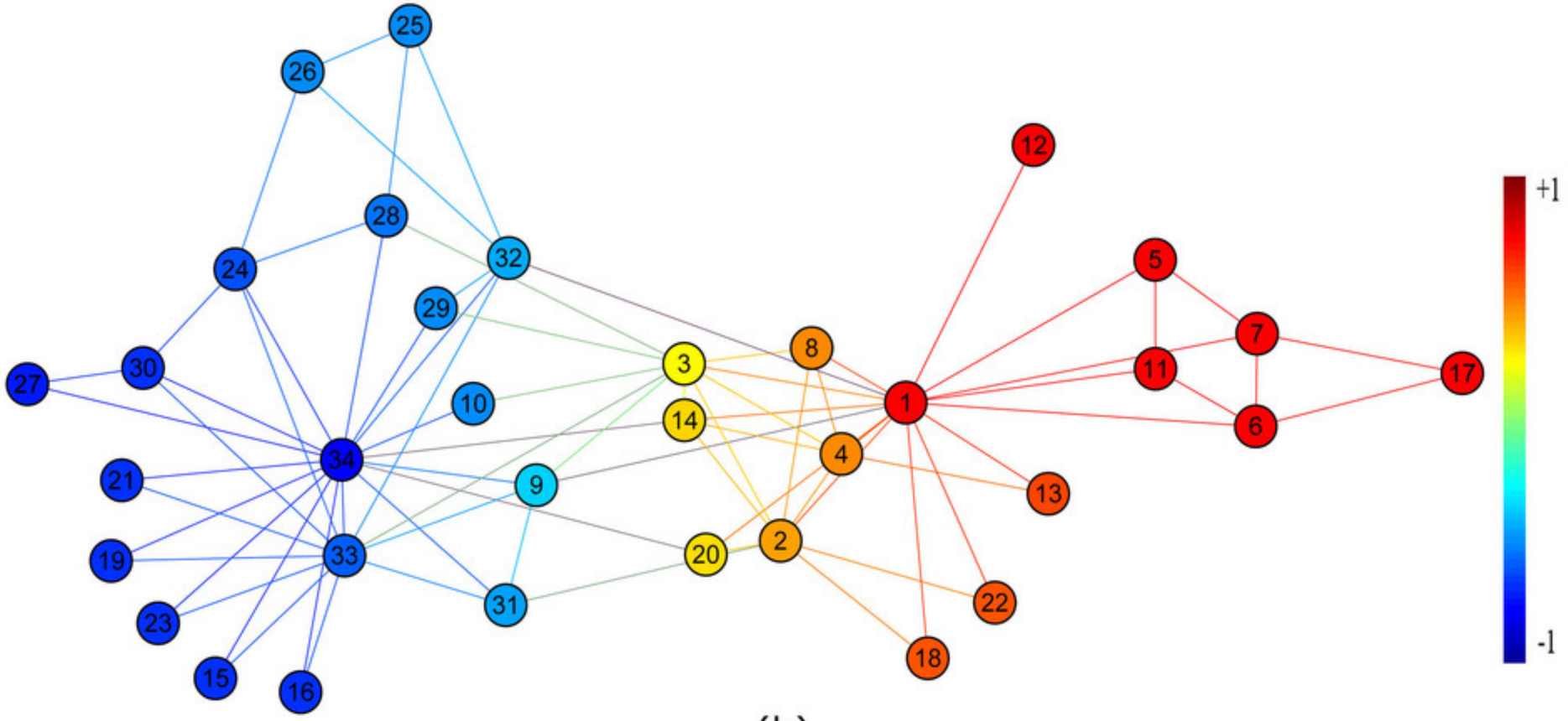
Agent-Based Simulation in Complex Networks

ESSAI 2024. Athens

Miguel Rebollo (@mrebollo)

Session 4. Communities

Zachary's Karate Club



(b)

W.W. Zachary, J. Anthropol. Res. 33:452-473 (1977).

Community definition

All members of a community can be reached from any other member

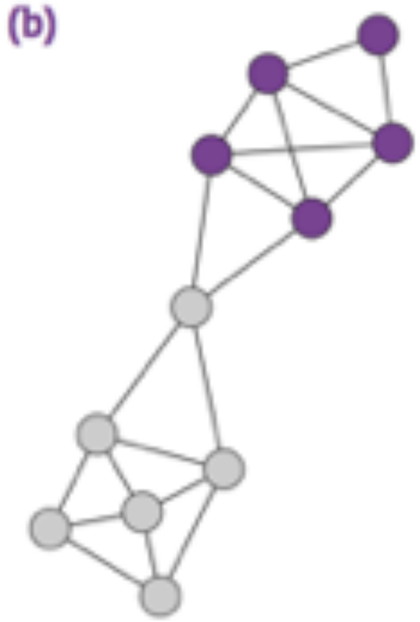
Community belonging is associated with

- connectedness
- density

Types by connection

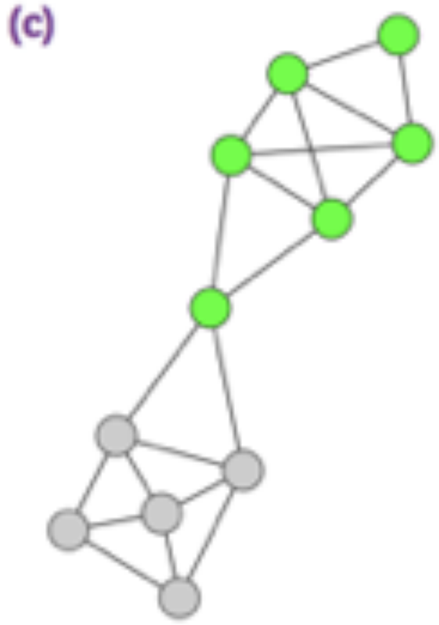
Clique

complete subgraph



Strong

in-links > out-links (each node)



Weak

in-degree > out-degree (total)

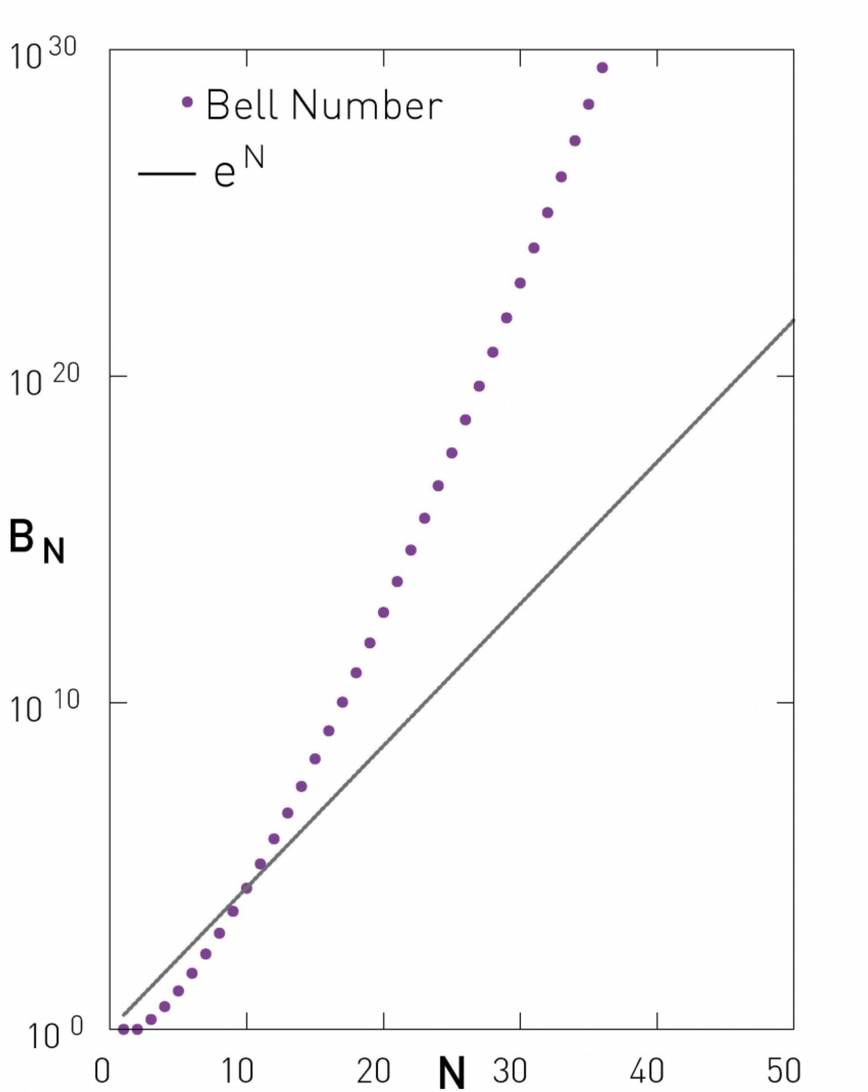
How many communities are there?

Graph partition problem (NP)
(Graph cut)

$$\frac{N!}{N_1!\,N_2!} = \frac{N^{\frac{N+1}{2}}}{N_1^{\frac{N_1+1}{2}}\,N_2^{\frac{N_2+1}{2}}}$$

Problem: N_1 and N_2 unknown
All possibilities = Bell's number

$$B_N = \frac{1}{e}\sum_{j=1}^{\infty}\frac{j^N}{j!}$$



Method 1. Hierarchical Clustering

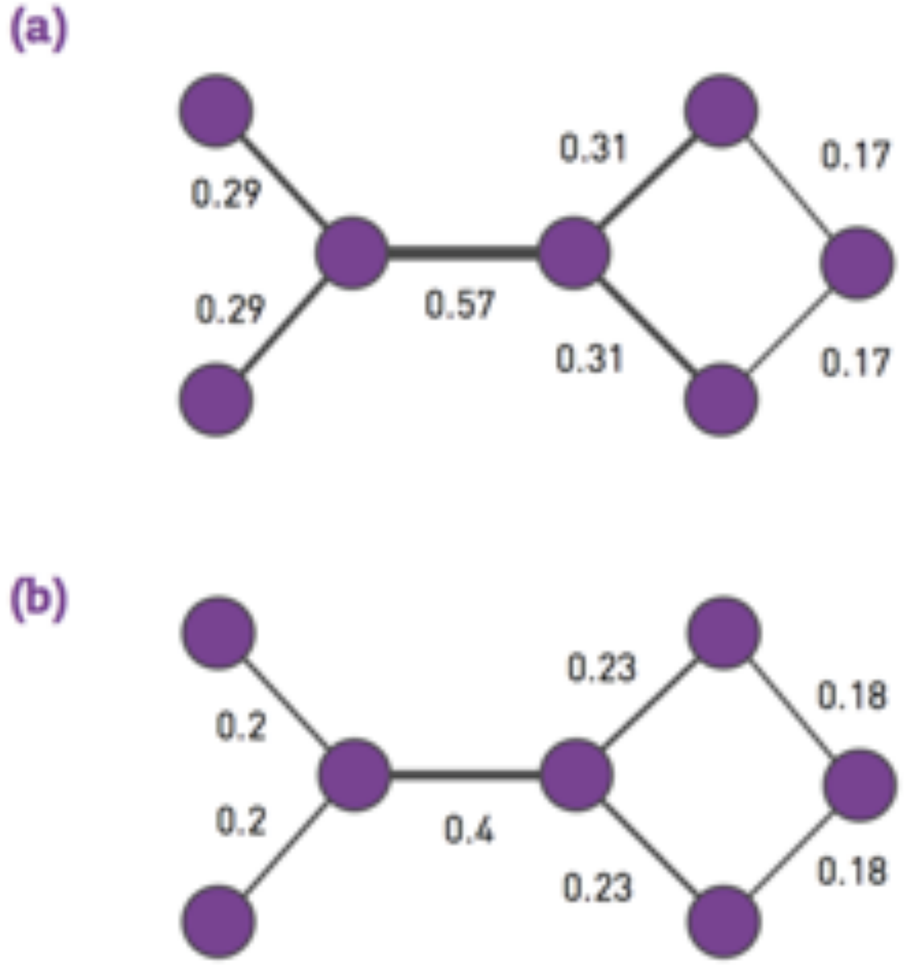
1. Create a similarity matrix
2. Decide group/node similarity
3. Hierarchical clustering to identify group belonging
    1. **Agglomerative**: merge nodes and communities
    2. **Divisive**: split communities removing edges
4. Dendrogram to identify best division

M. Girvan & M.E.J. Newman, PNAS 99 (2002).

Divisive: Girvan-Newman

1. Define centrality measure
2. Hierarchical clustering
3. Dendrogram

Betweenness



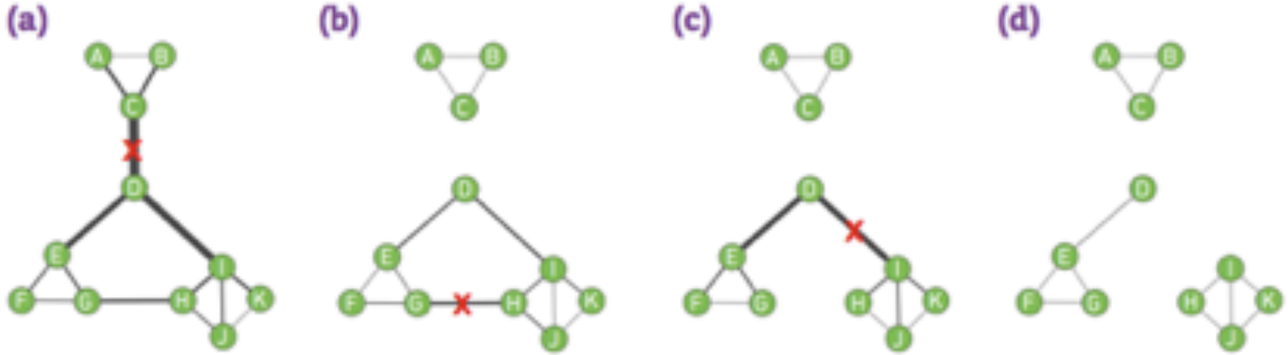
M. Girvan & M.E.J. Newman, PNAS 99 (2002).

Divisive: Girvan-Newman

1. Define centrality measure

2. Hierarchical clustering

3. Dendrogram



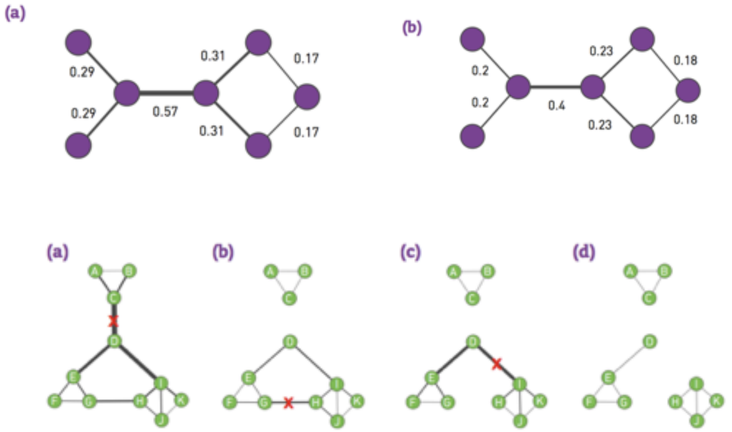
repeat
    1. calculate
    2. remove edge with highest centrality
    3. recalculate

until all links removed

M. Girvan & M.E.J. Newman, PNAS 99 (2002).

Divisive: Girvan-Newman

1. Define centrality measure
2. Hierarchical clustering
3. Dendrogram



M. Girvan & M.E.J. Newman, PNAS 99 (2002).

Method 2. Modularity

Hypothesis:

In a random network, the connection pattern is uniform (lack of community structure)

Definition:

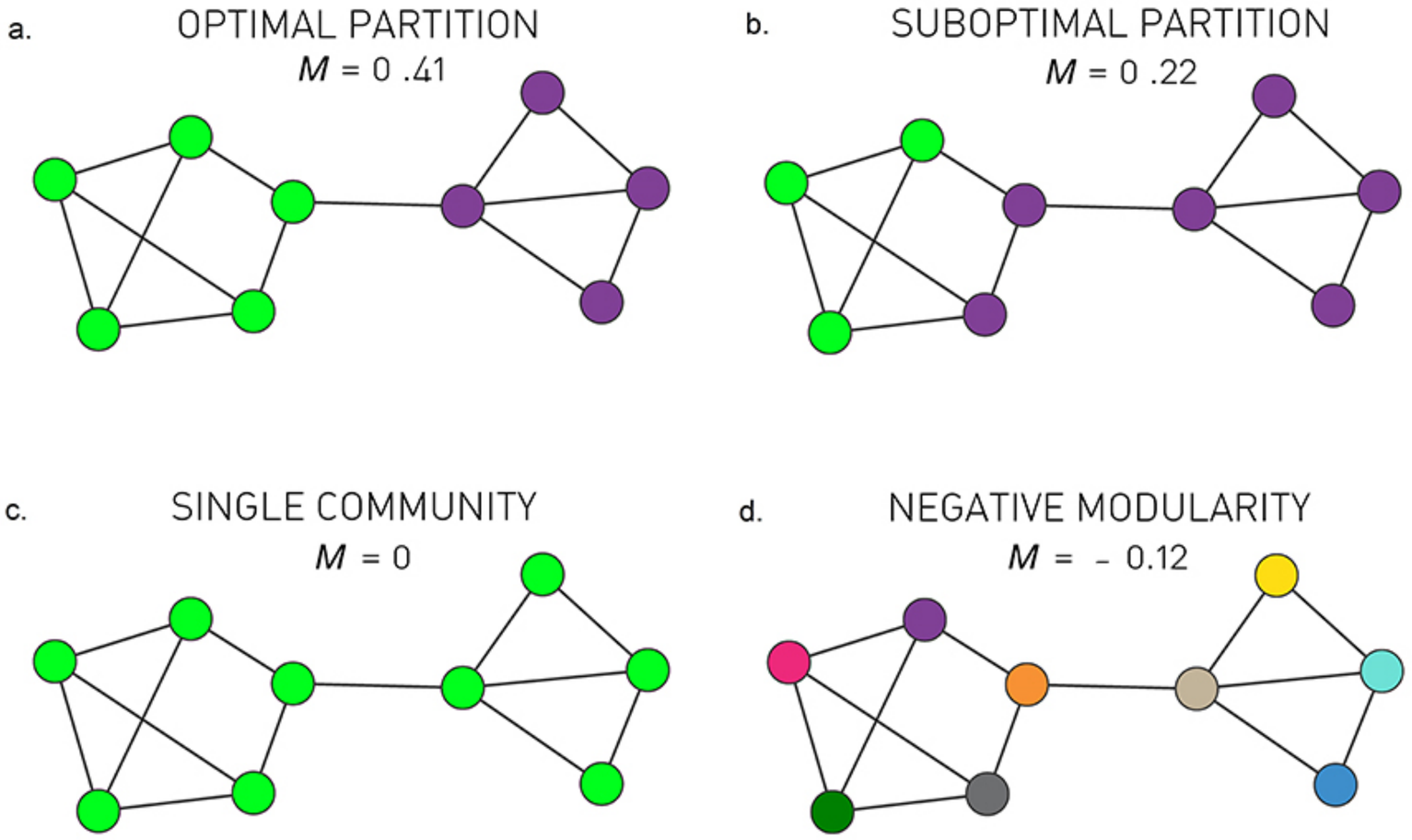
Let be n_c communities with N_c nodes and L_c links each.

If L_c is **larger than expected,** then c can be a community

Method 2. Modularity

$$M_c = \frac{L_c}{L} - \left(\frac{k_c}{2L}\right)^2$$

L: links of the network
L_c: links in c
k_c: total degree of c



Method 2. Modularity

Greedy algorithm

1. each node in one community
2. inspect all community pairs formed with 1 link and calculate new modularity ΔM
3. select the community division with highest ΔM
4. repeat until all nodes are in the same community
5. Select the partition with maximal M