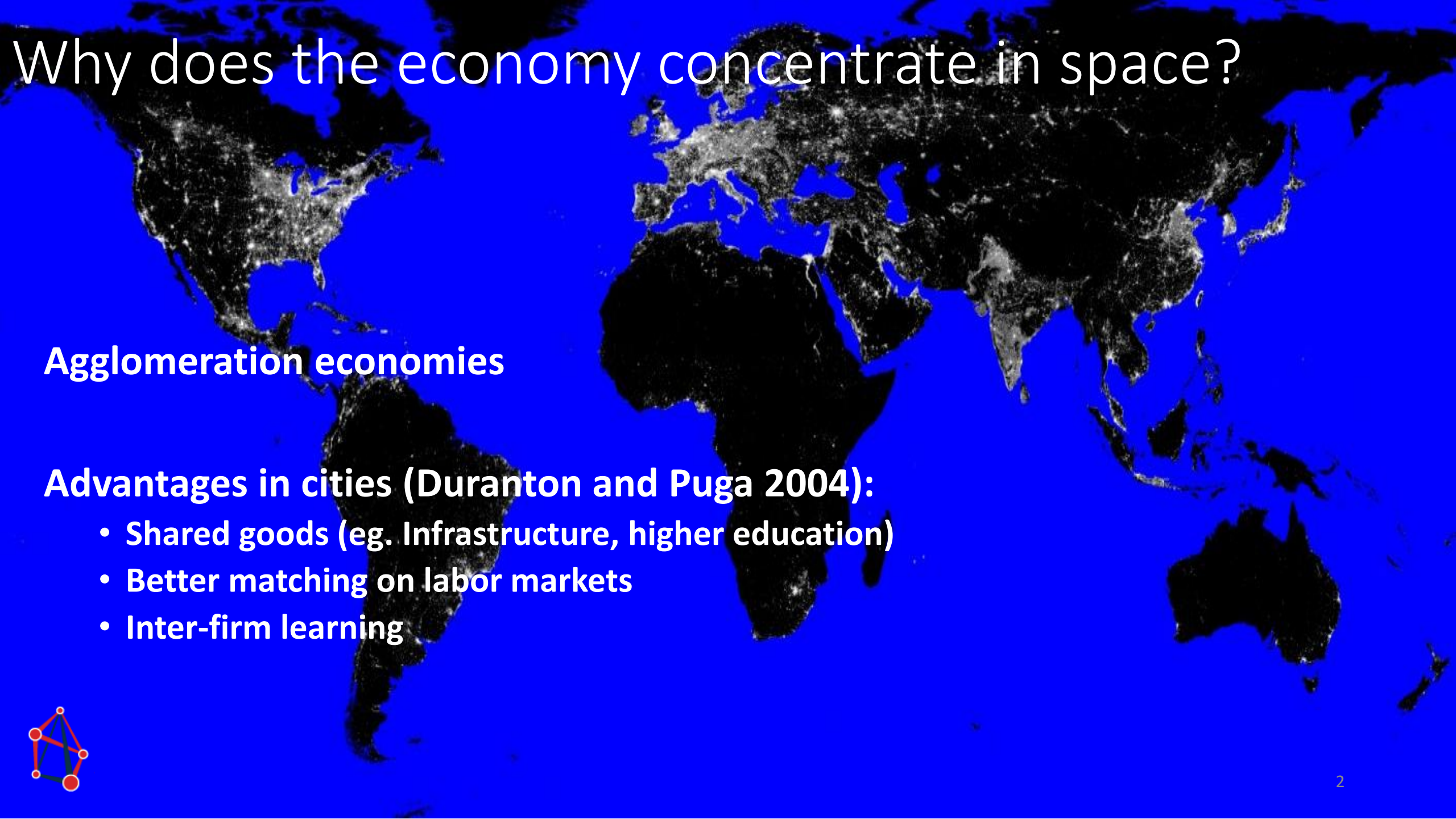
# Social Networks and Economic Geography

Balázs Lengyel
lengyel.balazs@krtk.hun-ren.hu

## Class 2:
## Economic relevanc and Geography of social networks
## Communities

Feb 17 2025

# Why does the economy concentrate in space?



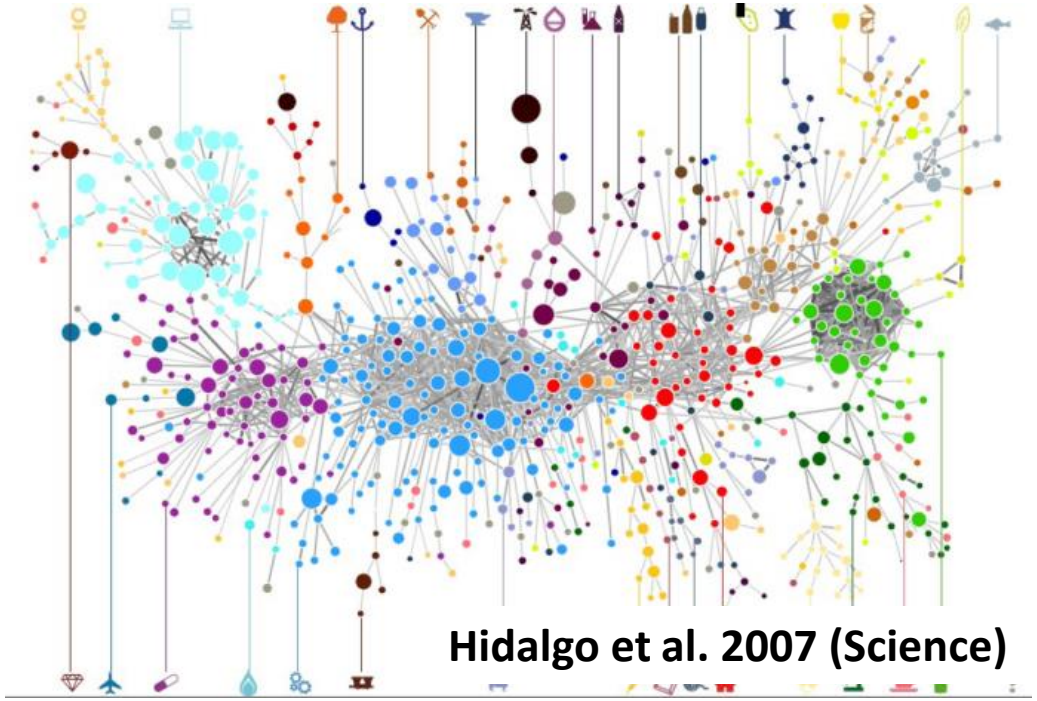
**Agglomeration economies**

**Advantages in cities (Duranton and Puga 2004):**

- **Shared goods (eg. Infrastructure, higher education)**
- **Better matching on labor markets**
- **Inter-firm learning**

# Economic progress and urban success

- Social interaction: population density
- Learning in the city: industry structure
- Related knowledge is easier to learn but contains less novelty



Hidalgo et al. 2007 (Science)

# We live in a connected world.

**Most of our social networks are local**
**but long-distance links establish quick access around the globe.**

**How do social networks induce economic progress?**

# Structure

1. Social network definitions
2. Social networks in geographical space
3. Economic relevance of small world networks
4. Spatial social networks and economic growth
5. Network Science Intro: Community detection

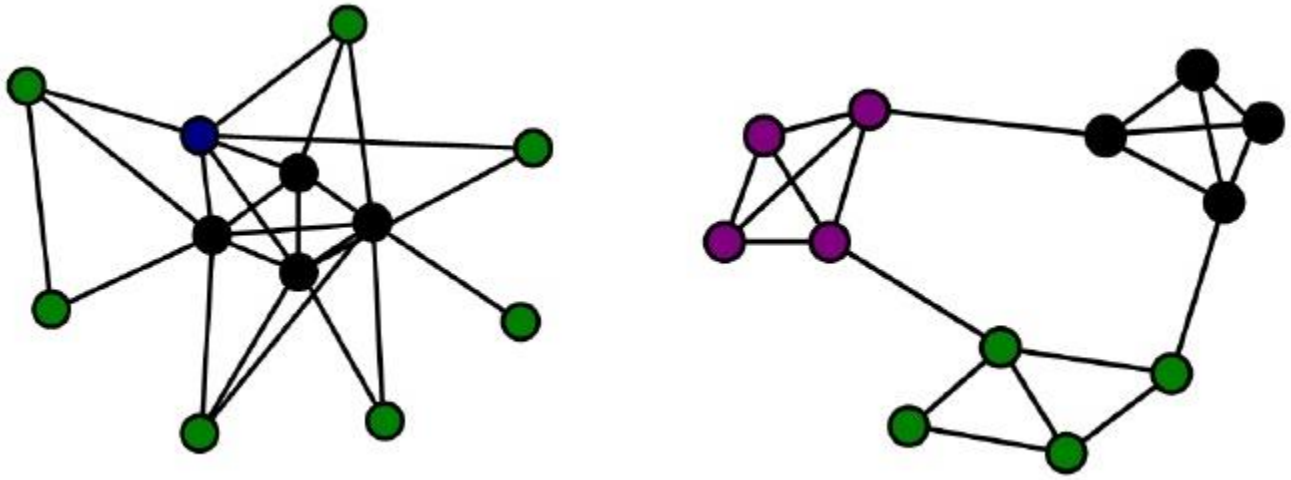
# 1. Social network concepts

# Network definition

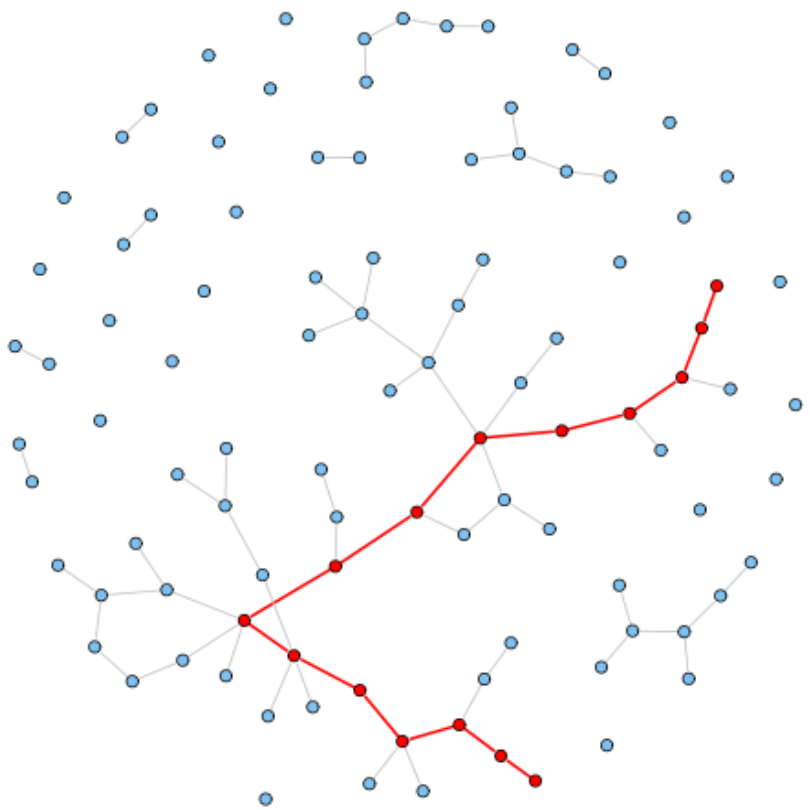
The network consists of a certain number of elements (actor, individual, organization) and the relationships between them (friendship, sales, professional support, etc.).

Dots: Node, vertex

Lines: edge, tie



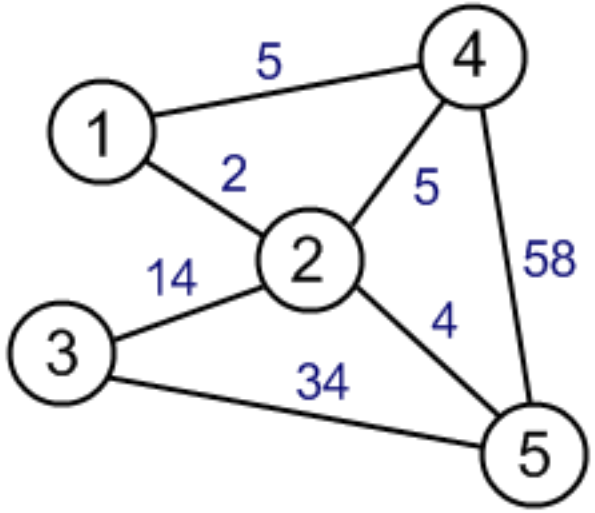
# Path



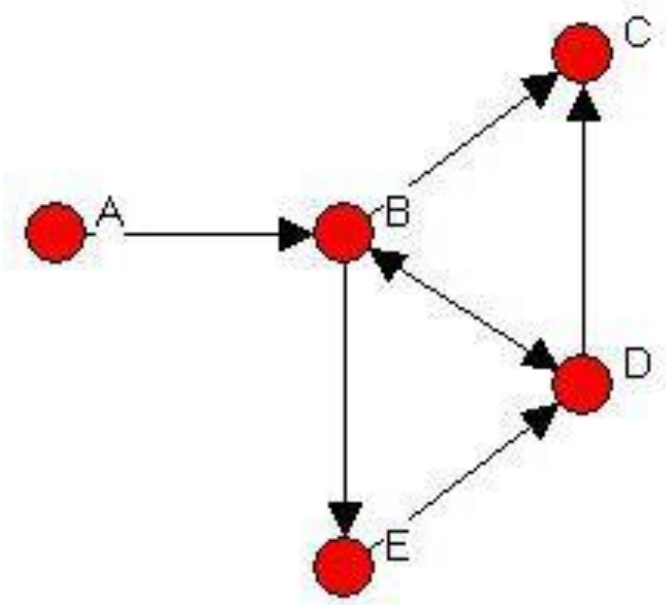
- Path: The set of edges that we can get from one node to any other node
- Distance: the number of steps needed to get from one node to another in the network
- Shortest path: the path between two nodes that includes the fewest steps

# Edge characteristics

**Weighted – unweighted**
Friendship vs. the quality of friendship
Meetings vs. frequency of meetings




**Directed-undirected**
Collaboration vs. information diffusion

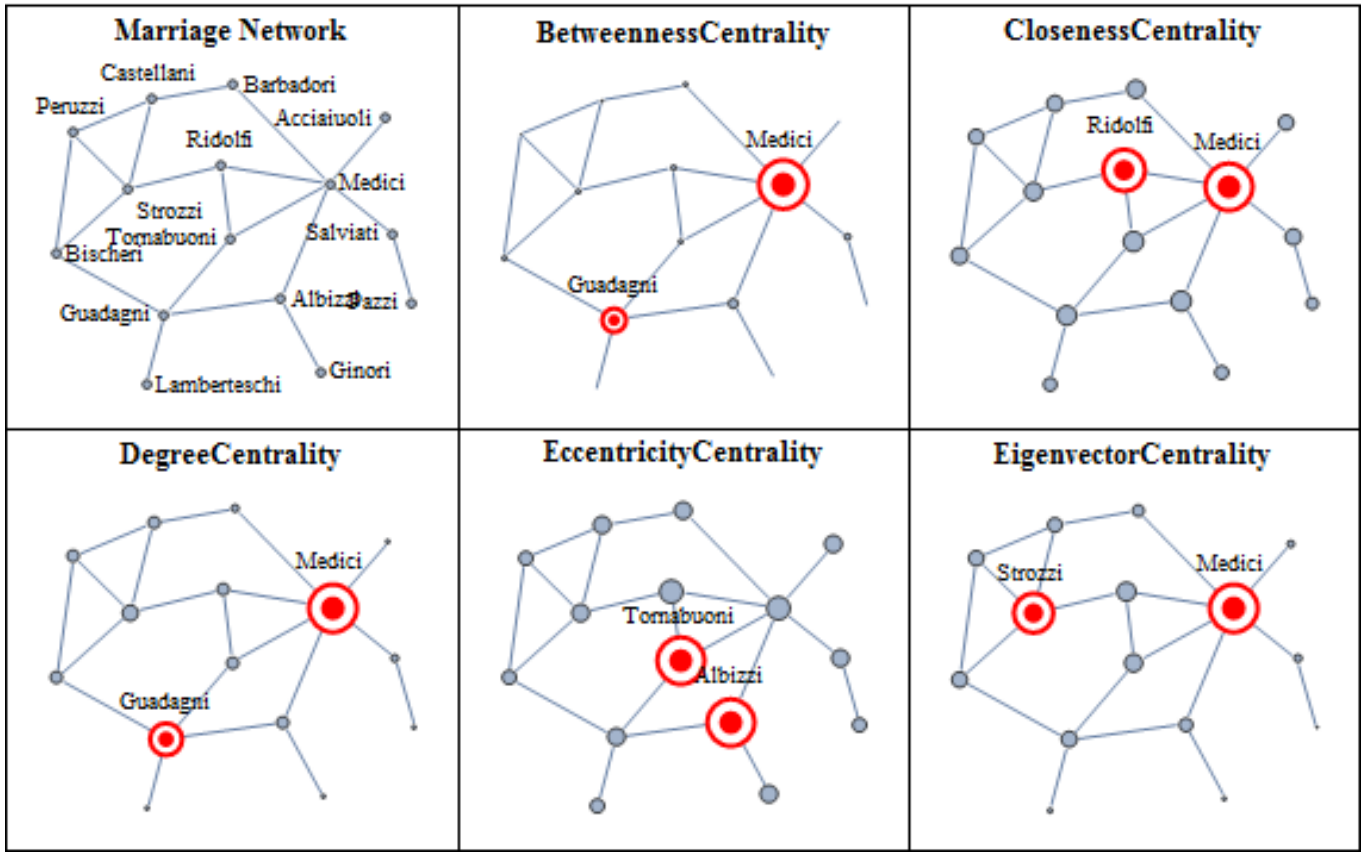


# Network features: centrality of nodes

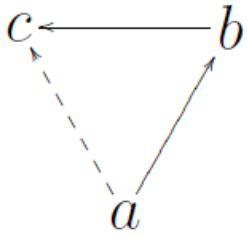
Degree centrality: The number of connections of nodes. For directed networks, we distinguish indegree and outdegree.

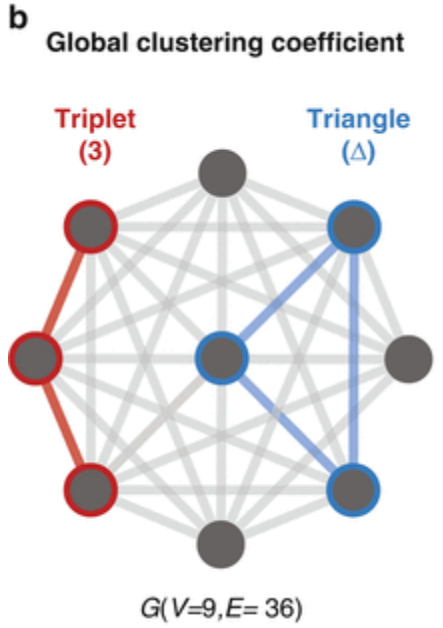
Betweeness centrality: the number of shortest paths that pass through the vertex.

Closeness centrality: the inverse of the sum of distances from the node to all other nodes.

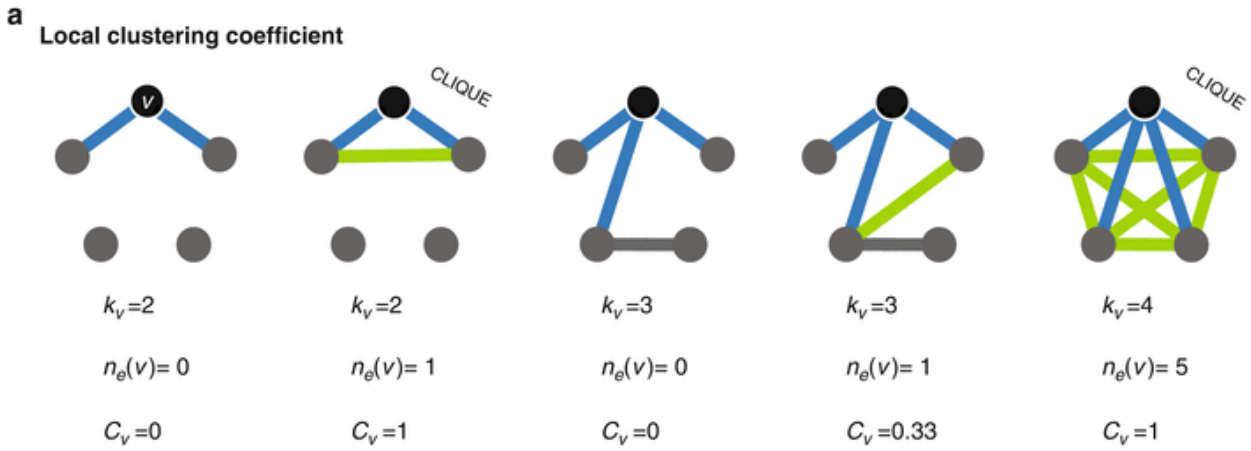


# Clustering



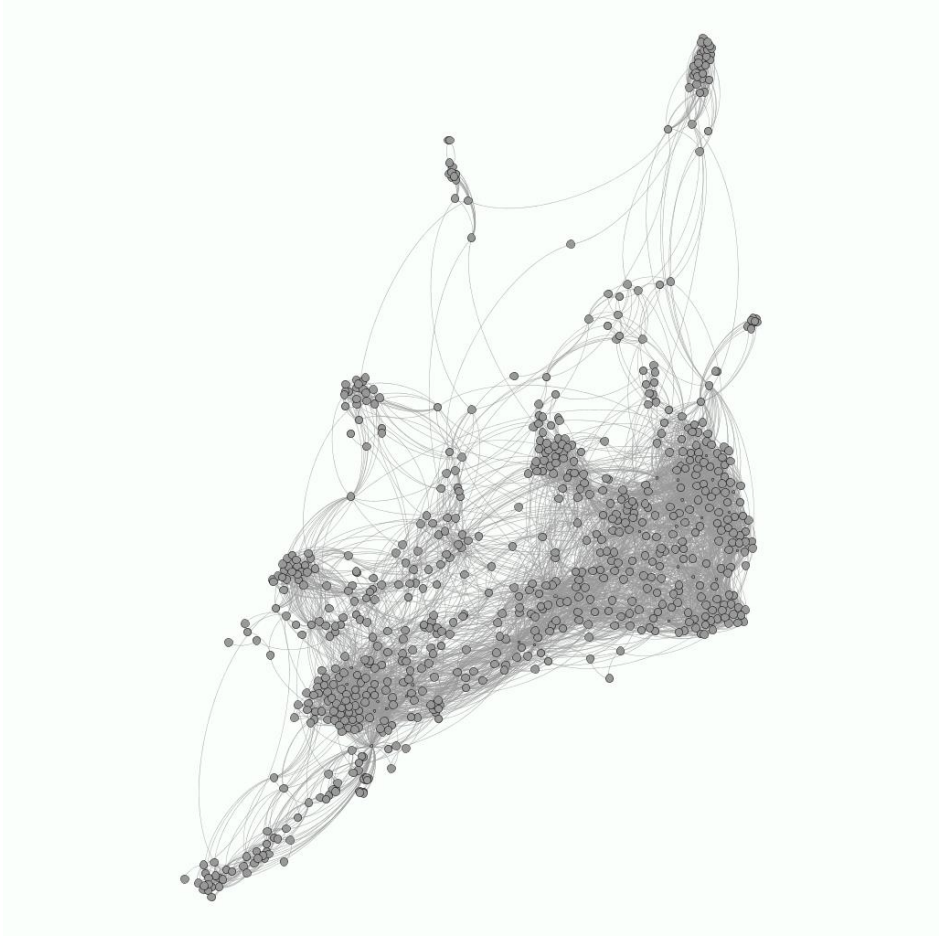

- Transitivity: if ‘a’ knows ‘b’ and ‘b’ knows ‘c’-t, then ‘a’ is likely to know ‘c’ as well
  - Note that we concentrate on tiadic relations in this measure!
- Global clustering: how many triads are closed out of all possible triads?
- Local clustering: for all *j,k* neighbor-pairs of *I, how many times does the j-k* edge materialize? (one can sum this up for the full network)



# Modularity

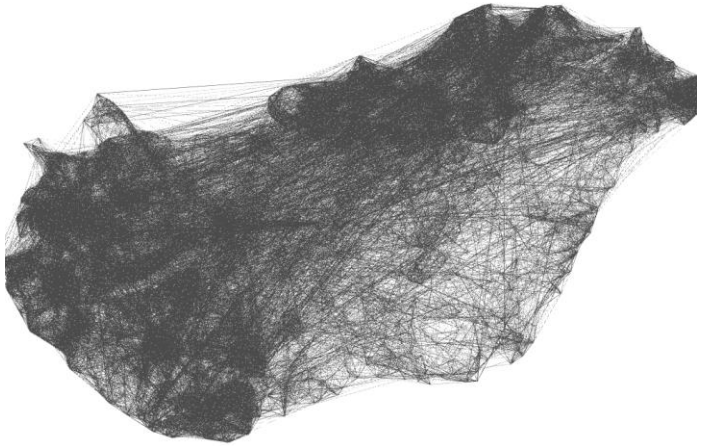
**Moduls**

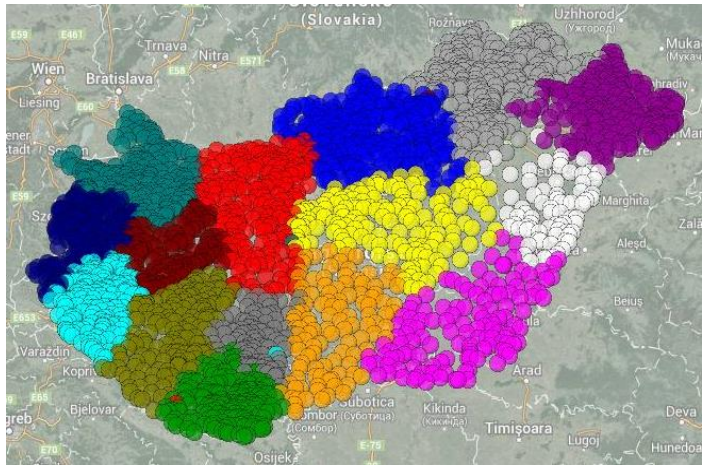
Groups or subnetworks emerge due to transitivity



**Modularity**

Probaility (density) of edges is higher within the moduls than across the moduls

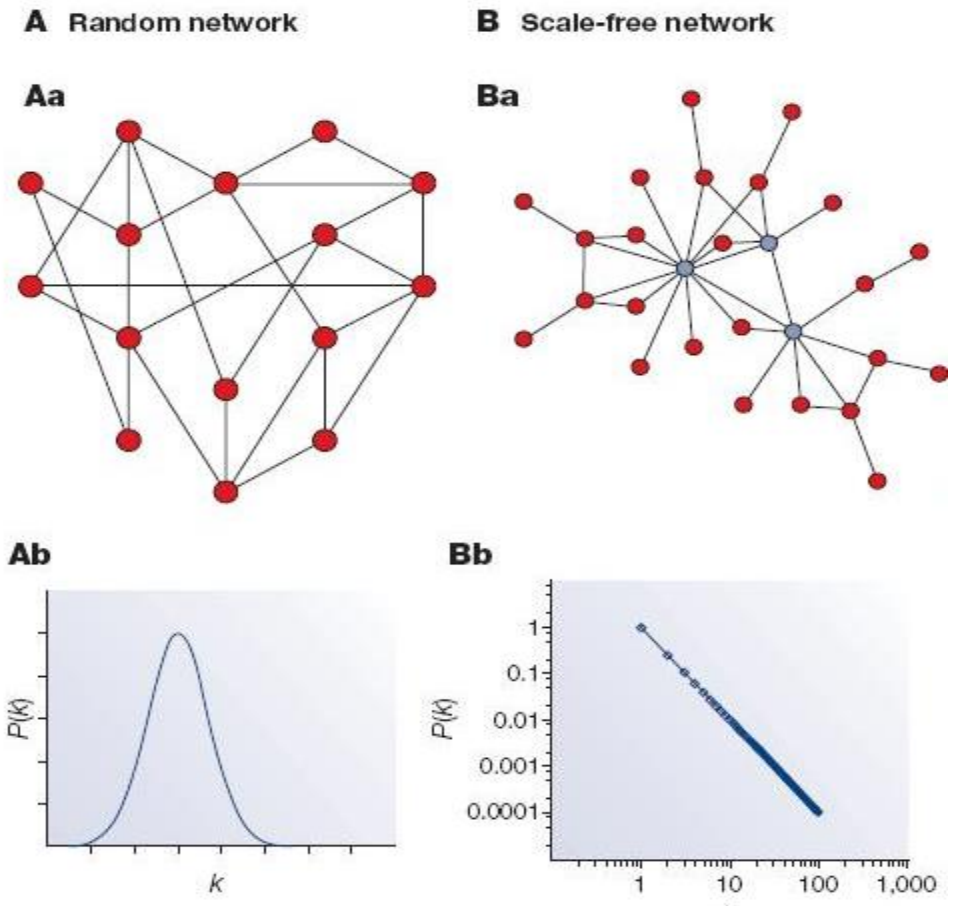




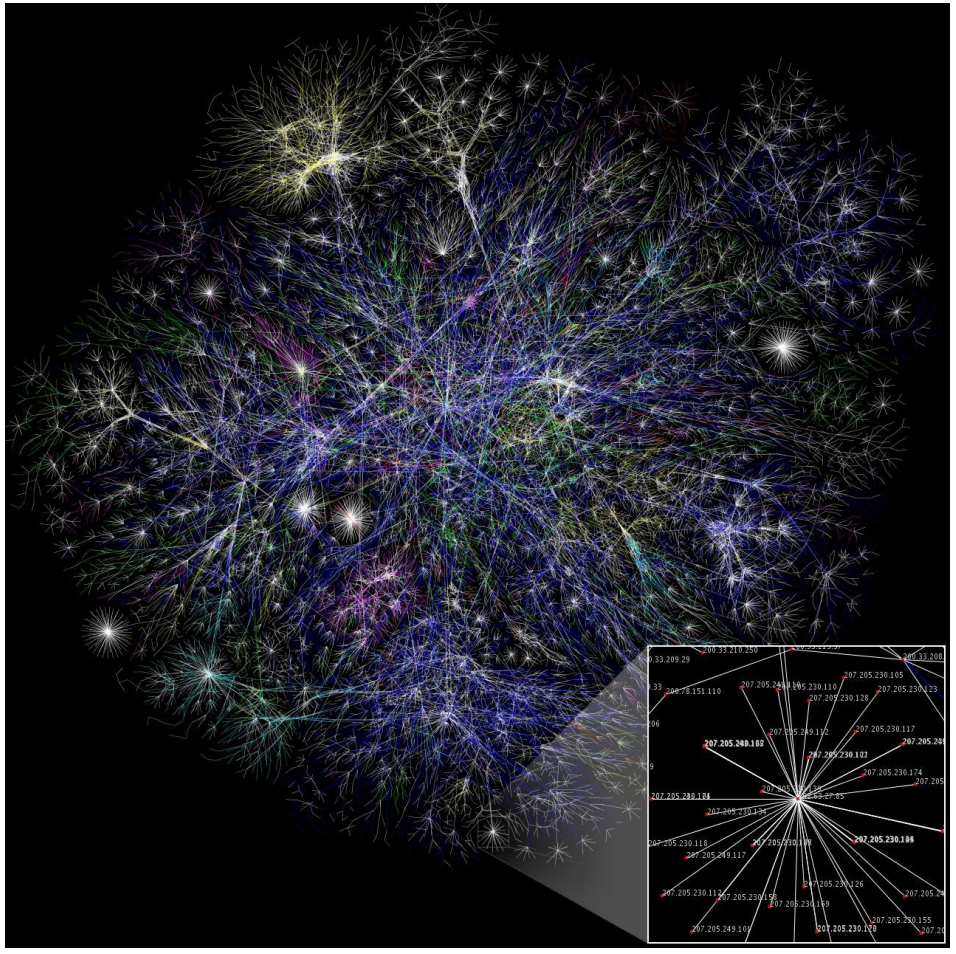
# Degree distribution

**Random network** (Erdős-Rényi)

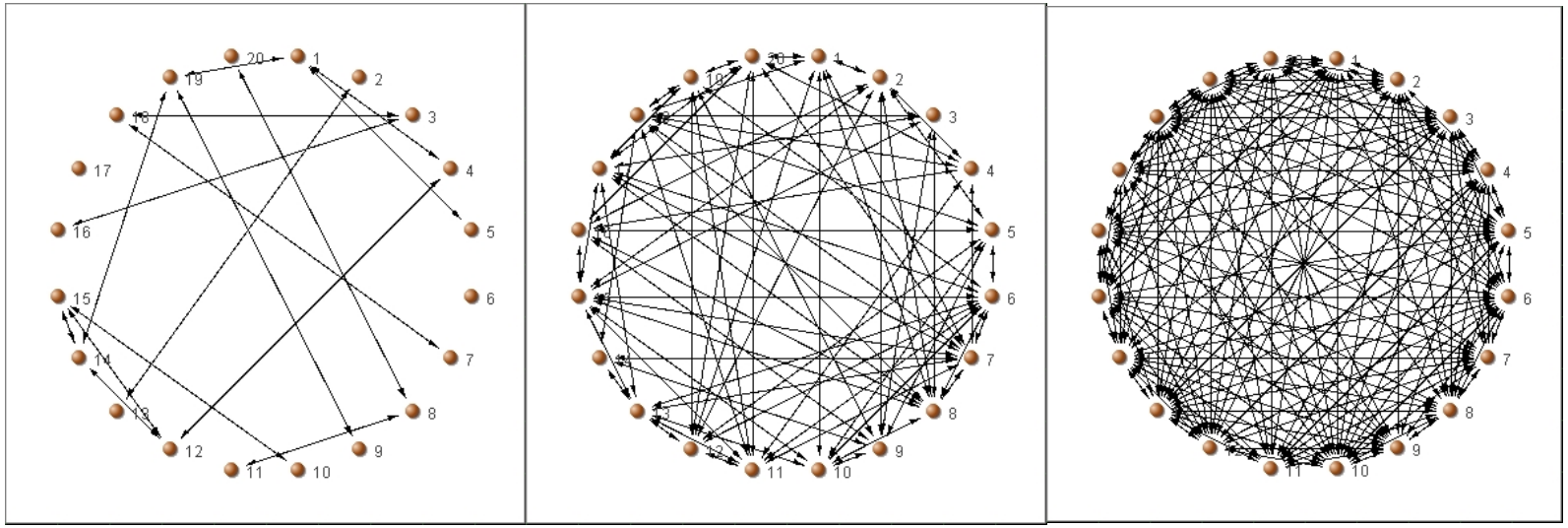
- Normal degree distribution (symmetric to the mean)
- Nodes are similar to each other in terms of degree centrality




**Scale-free network** (Barabási-Albert)
Degree distribution is Poisson, its' logarithmic version is linear
Some nodes are emerging as hubs in the network



# Random networks(*N* = 20)



$p = 0{,}1$

$\bar{d} = 1{,}6$

$\Delta = 0{,}084$

$p = 0{,}4$

$\bar{d} = 7{,}8$

$\Delta = 0{,}41$

$p = 0{,}8$

$\bar{d} = 15{,}2$

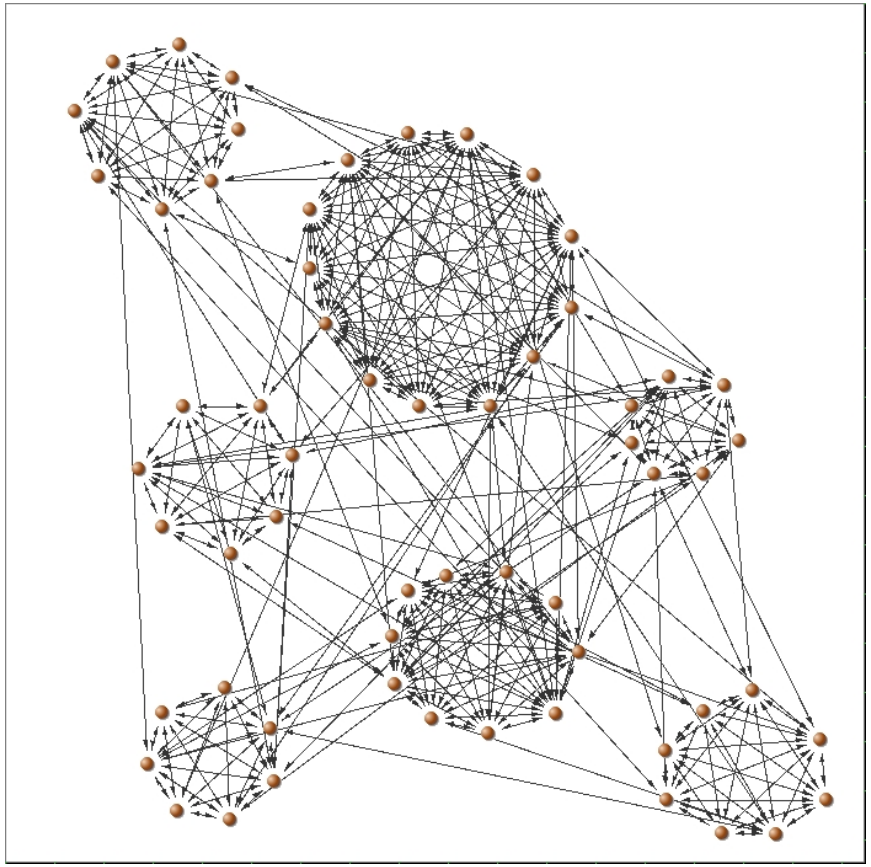
$\Delta = 0{,}8$

# Small worlds

- Social networks cannot be described by random ties:

  Dense local relations and

  few bridges between isolated groups

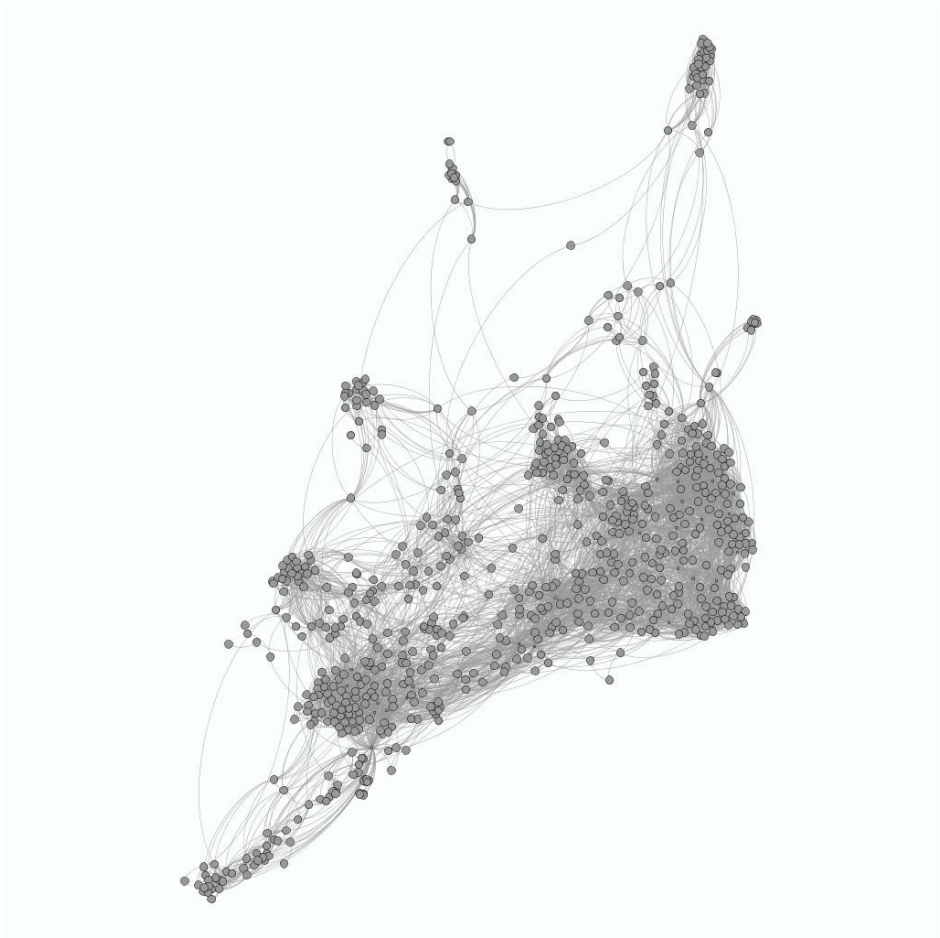


- Small worlds
  - High clustering – due to dense local networks
  - Short paths – due to bridges

# Modularity

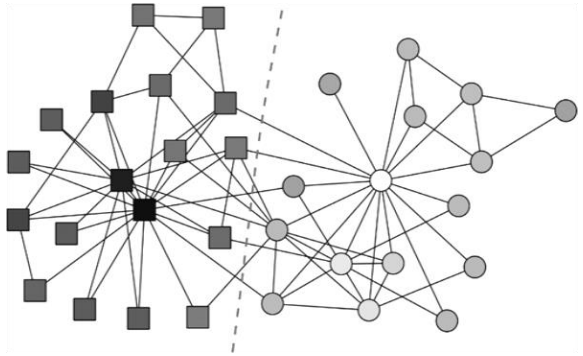
**Moduls**

Groups/subnetworks/moduls/communities emerge due to transitivity



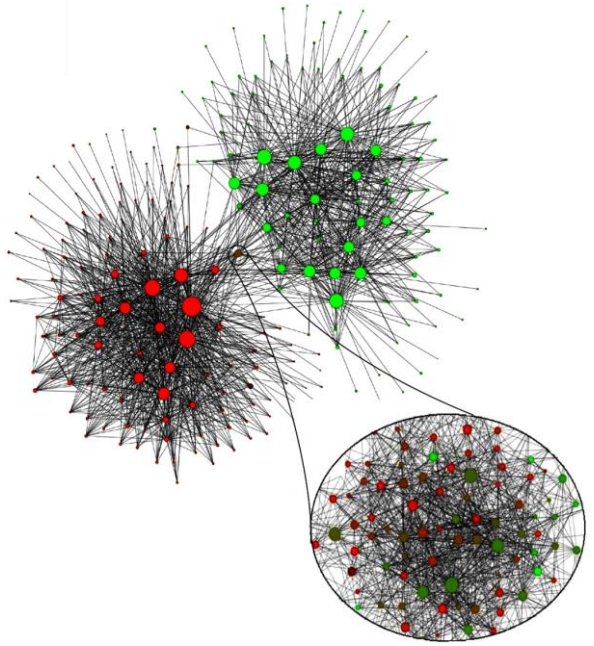
**Density Hypothesis of Community detection:**

Communities correspond to locally dense neighborhoods of a network

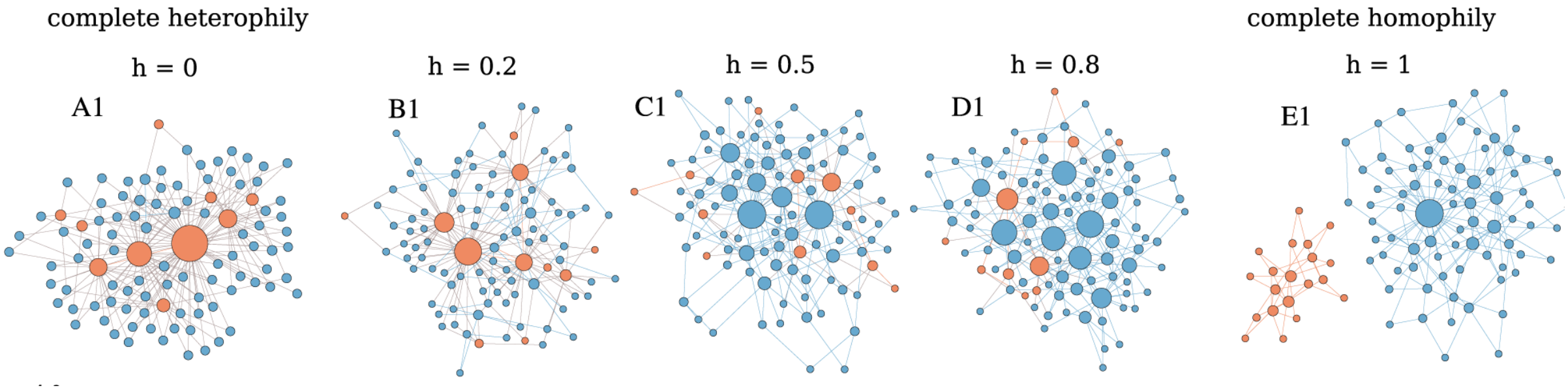


→ Karate Club: Breakup of the club

→ Belgian Phone Data: Language spoken



# Homophily / assortativity



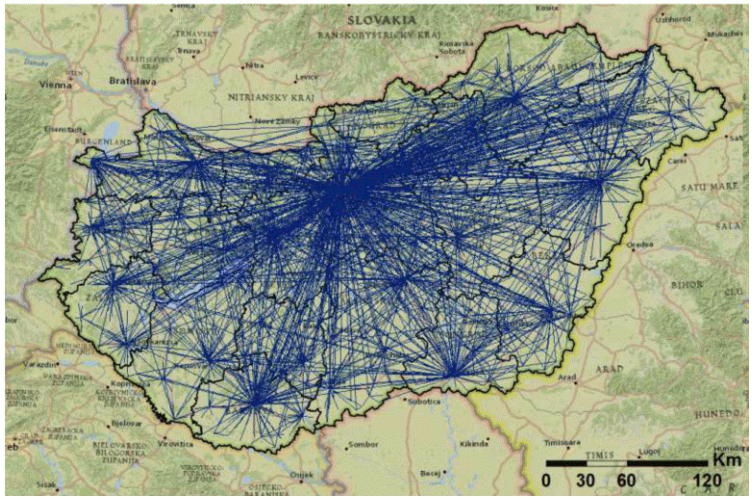
# 2. Social networks in geographical space

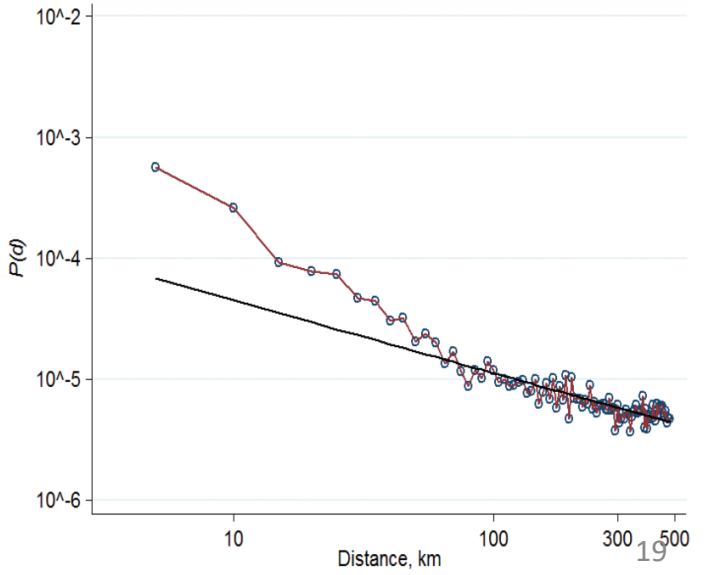
# Distance effect

- Creation and maintenance come with costs
- Costs increase with distance
- Gravity models help us analyze the distance-cost relation

$$P = \Sigma_d L_{ij} / \Sigma_d N_i \times N_j$$

- The probability of triadic closure decreases with distance

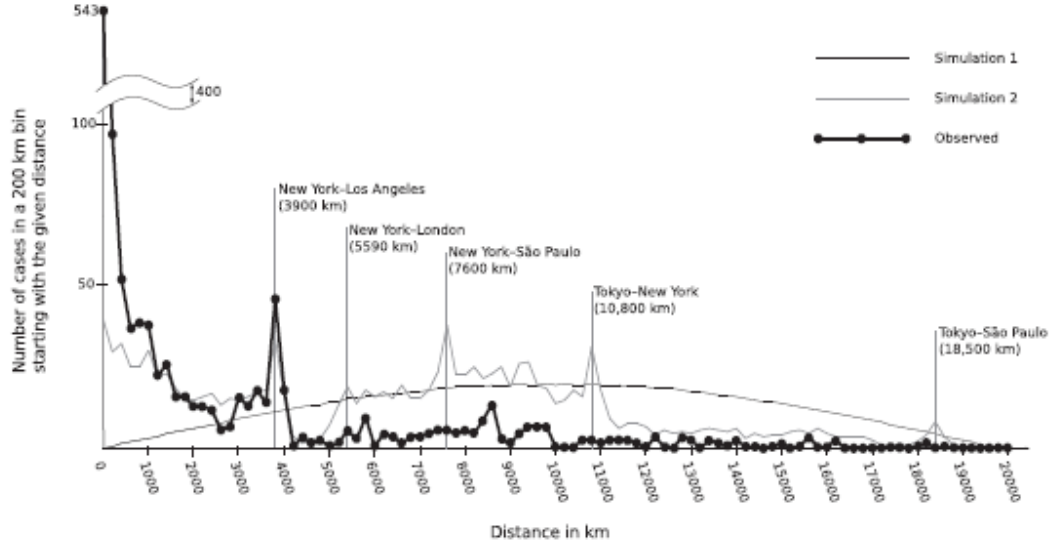


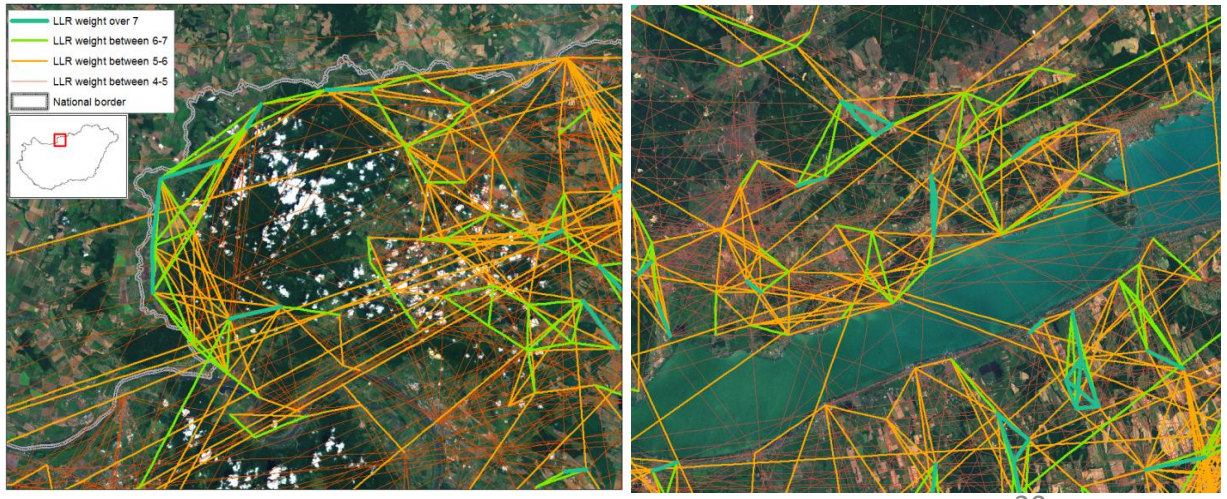




Liben-Nowell et al. (2005) PNAS; Lambiotte et al. (2008) Physica A; Lengyel et al (2015) PLOS ONE

# Spatial barriers

- Spatial mobility determines the geography of social networks
- Spatial barriers hinder mobility and influence network structure

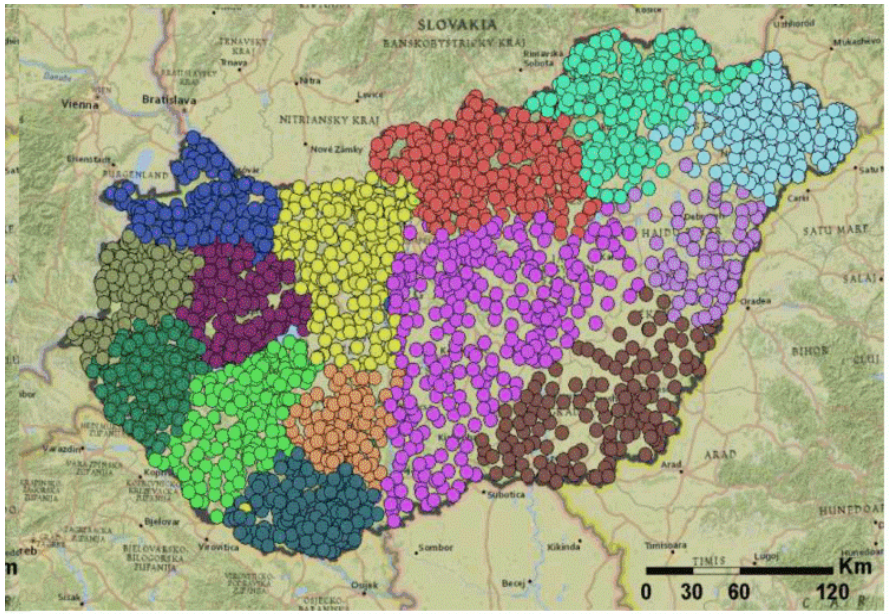


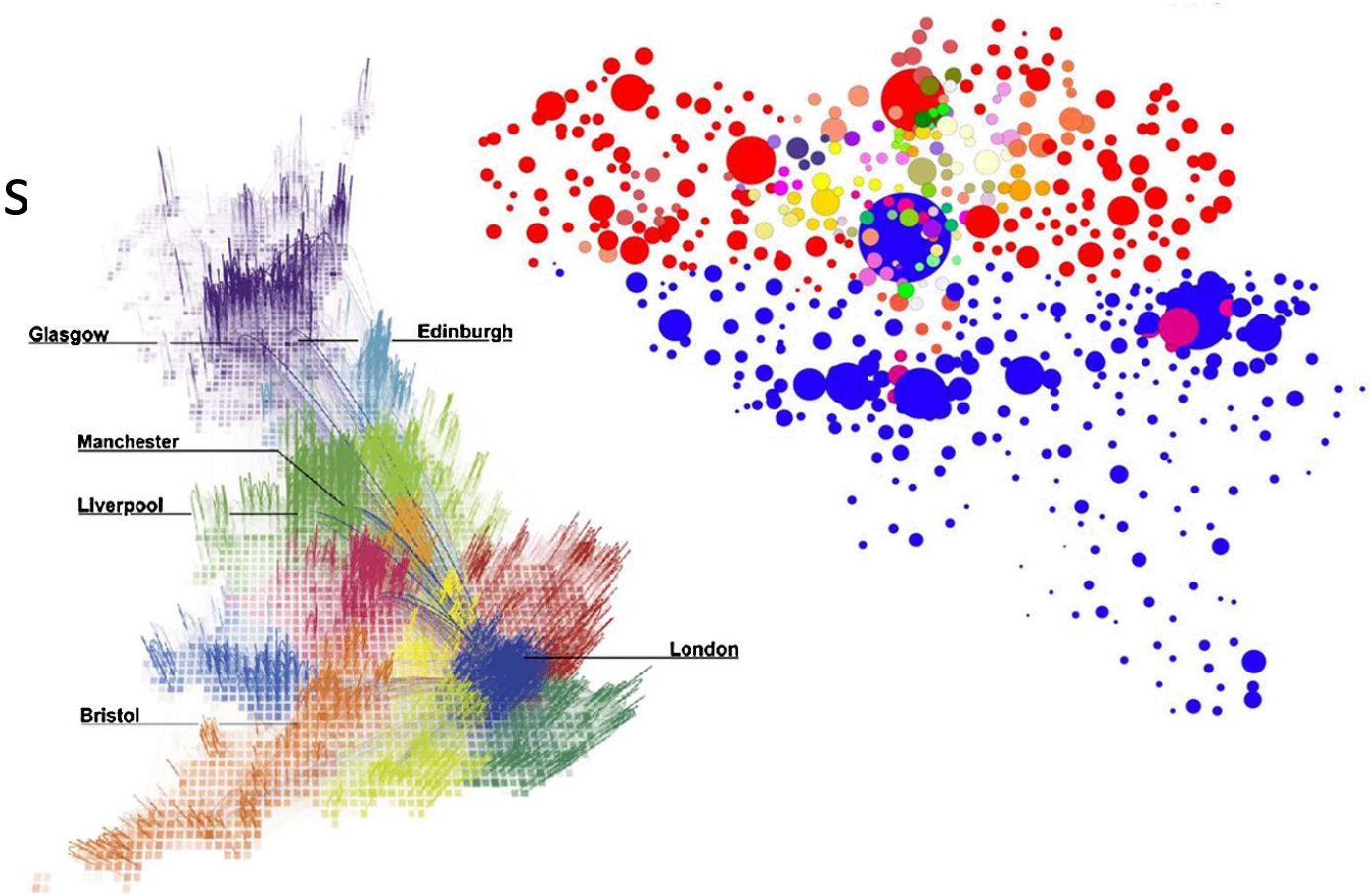




Thaktayev et al. (2012) Social Networks; Jakobi and Lengyel (2016) Springer

# Spatial modules

- Social networks break down by administrative or cultural regions




Expert et al. (2011) PNAS; Sobolevsky et al. (2013) PLoS ONE; Lengyel et al (2015) PLOS ONE

# 3. Economic relevance of spatial networks


December 2010

# Stanley Milgram: 6 degrees of separation / The Experimenter (2015)

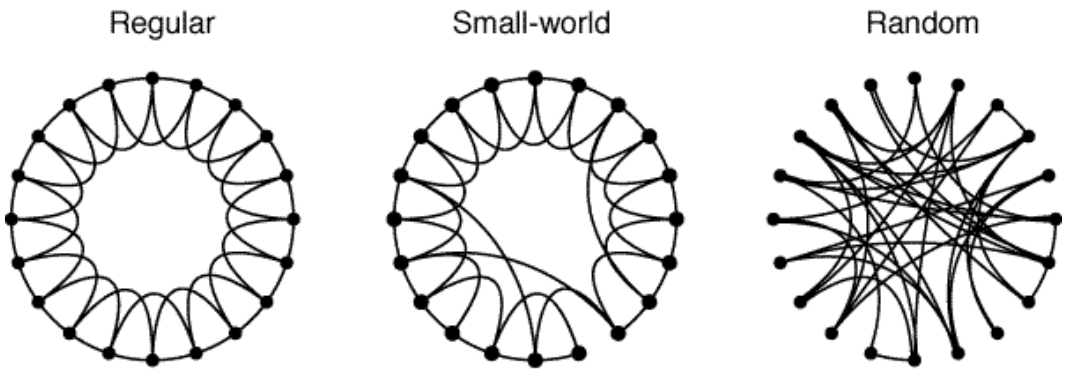
# Small worlds

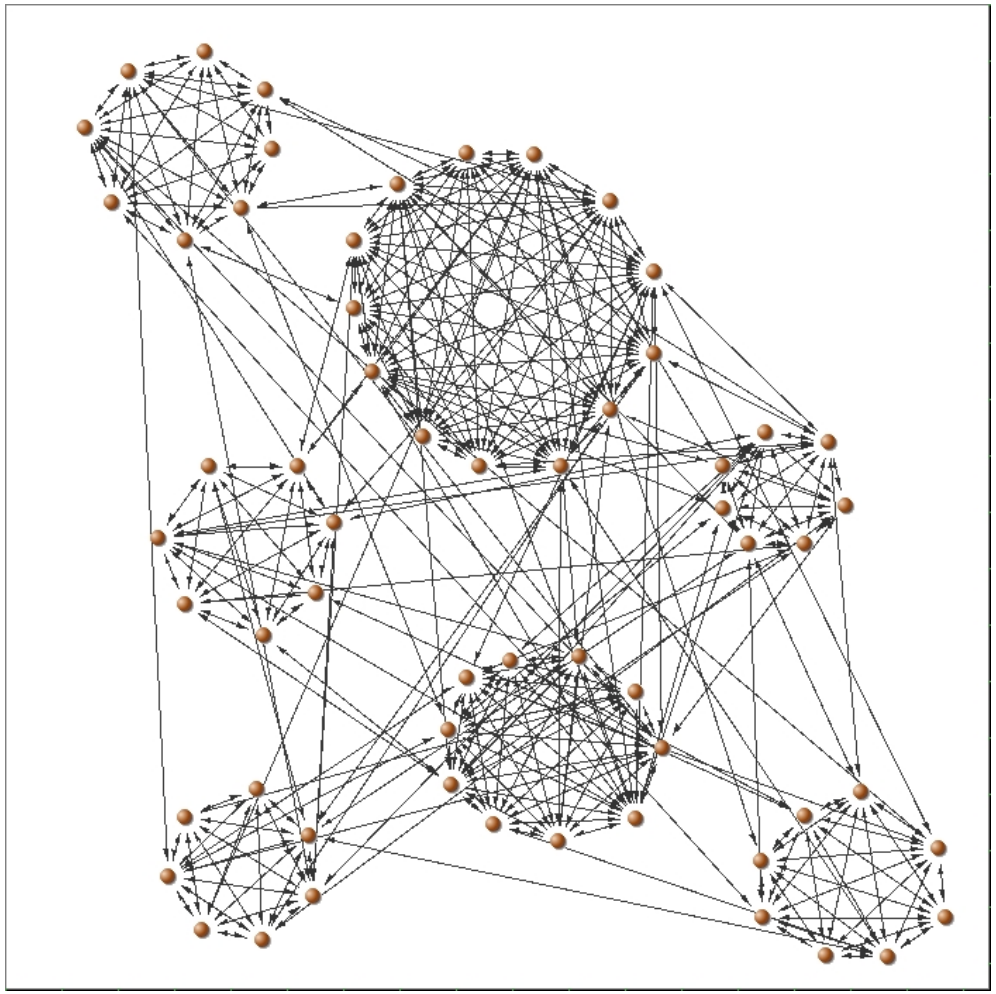
- Social networks cannot be described by random ties:

  Dense local relations and

  few bridges between isolated groups

- Small worlds
  - High clustering – due to dense local networks
  - Short paths – due to bridges



Watts-Strogatz (1998) Nature

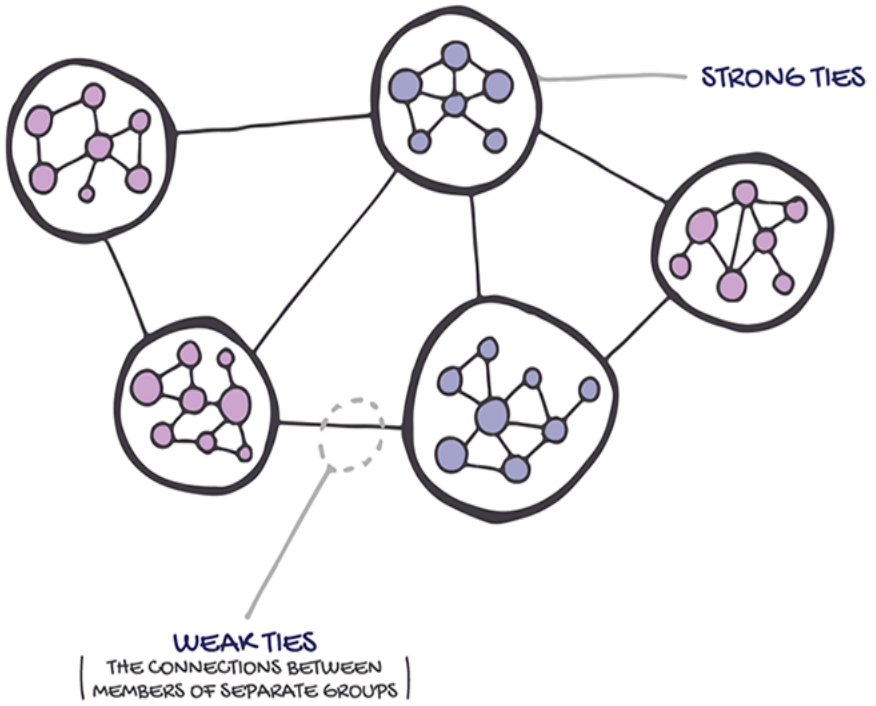
# Weak ties and Brokers

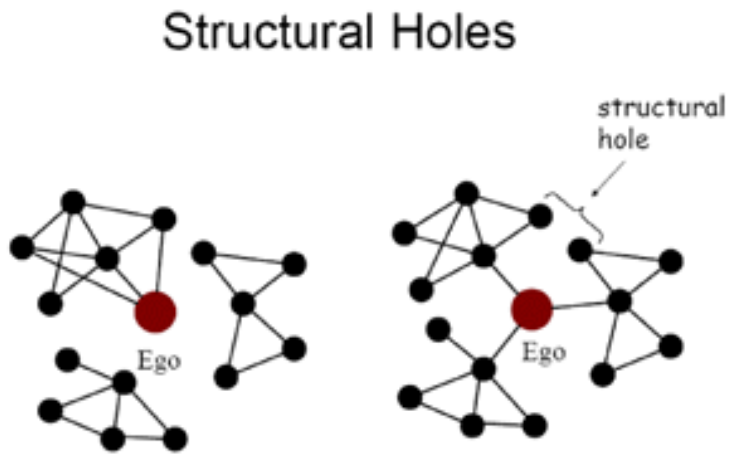
Weak ties (Granovetter, 1973):

- Valuable information comes from those contacts who are in occasional/not frequent relation with us.

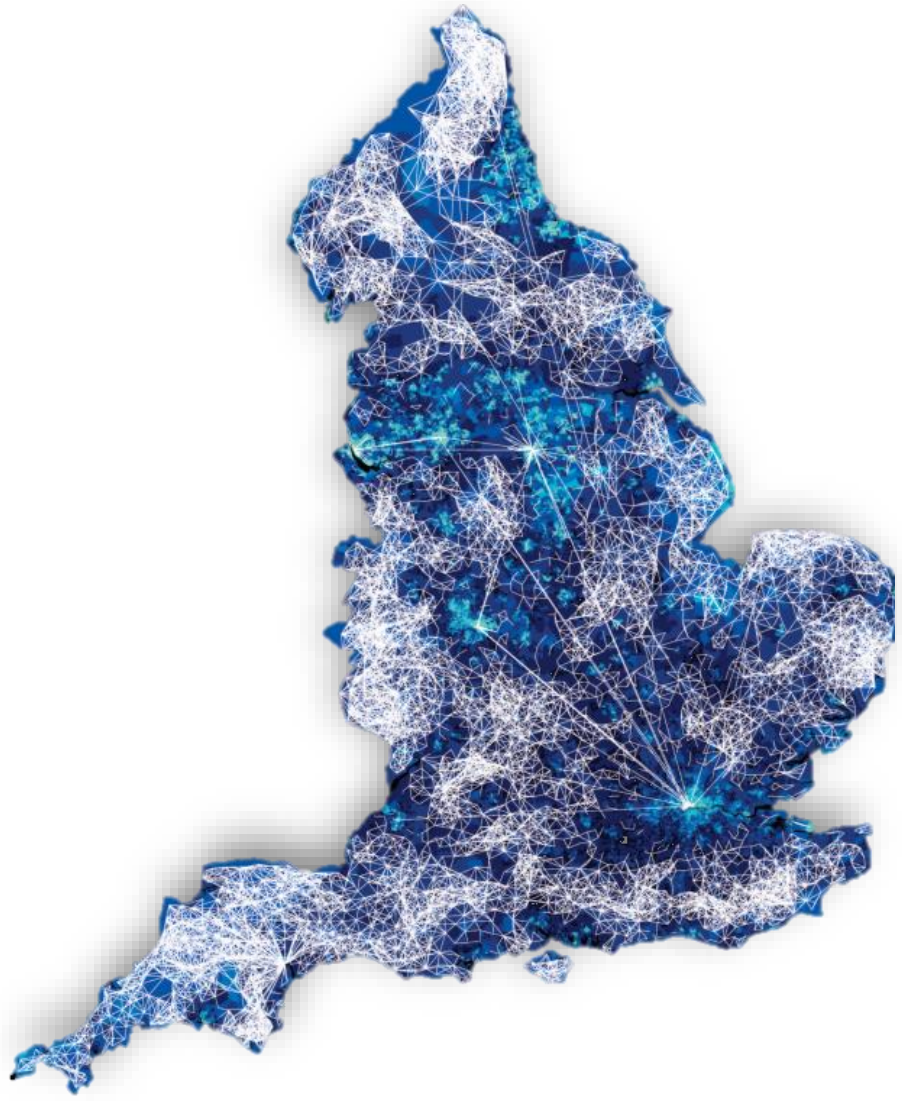
Advantage of Brokers (Burt, 1992)

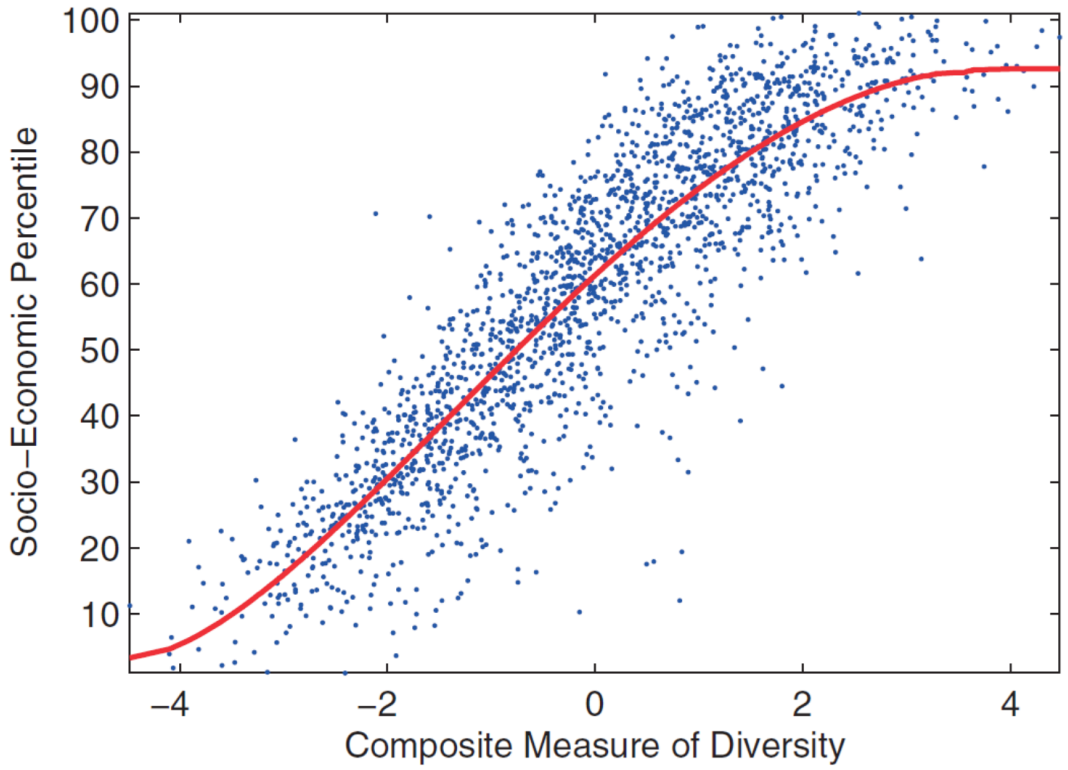
- Information is homogenous and redundant in dense clusters.
- In case information is not overlapping between two loosely knit clusters, there is a structural hole between these clusters.
- Brokers:
  - tertius gaudens -> control of flows
  - tertius iungens -> establish links



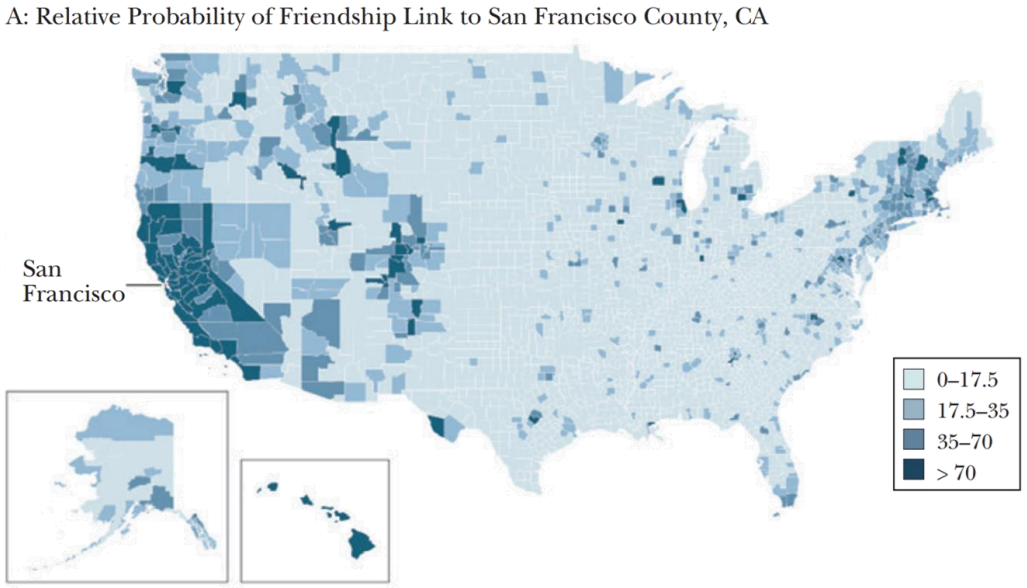
# Social networks, geography and wealth

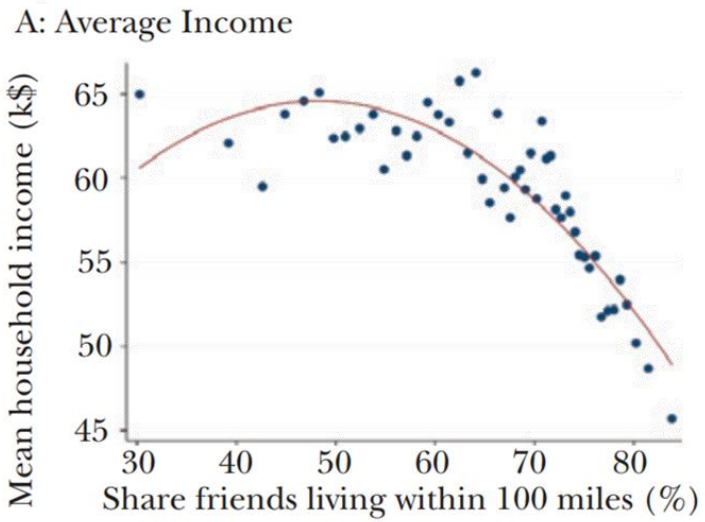


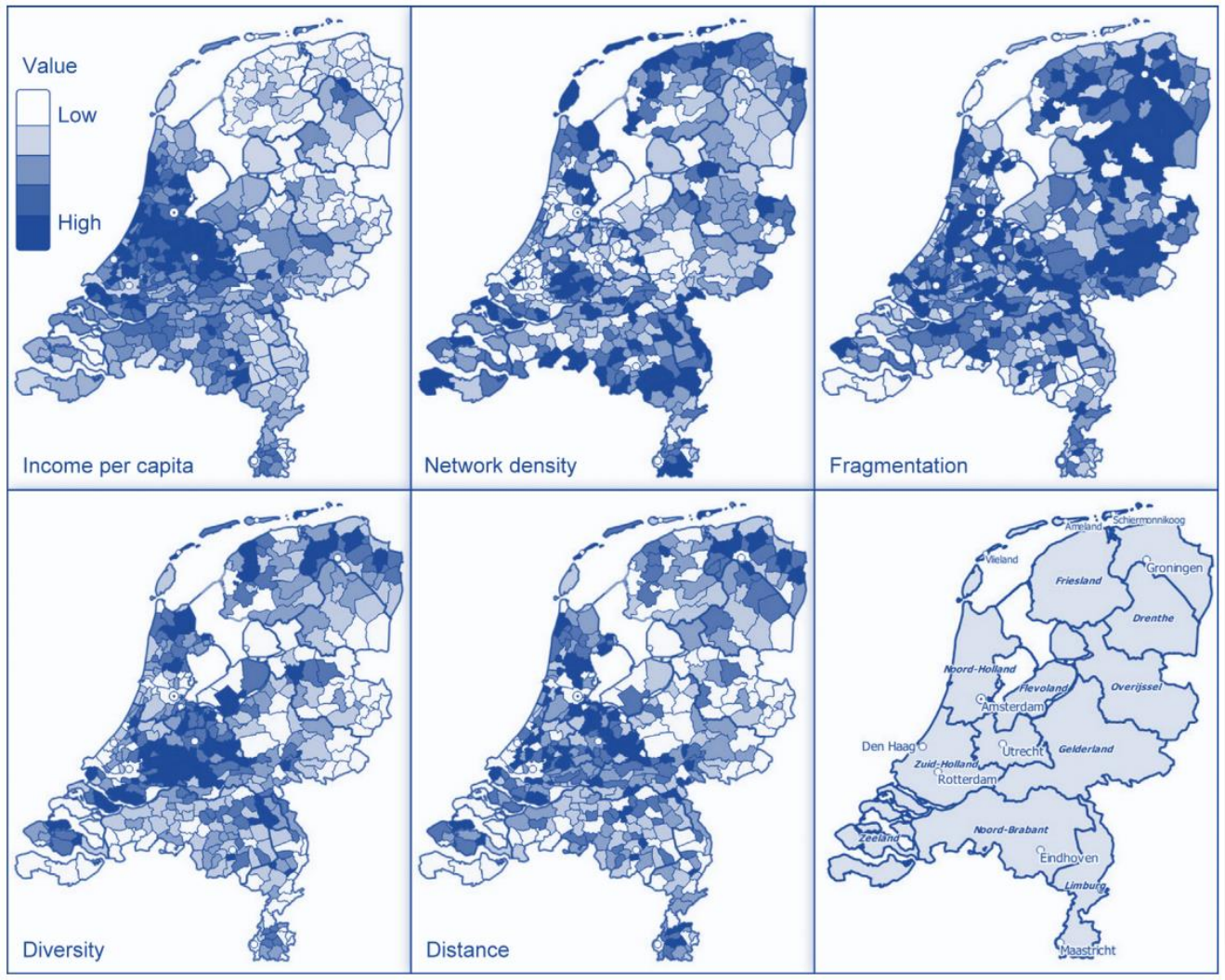



**Eagle-Macy-Claxton (2010) Science**

# Aggregate social network structures correlate with regional average of individual income



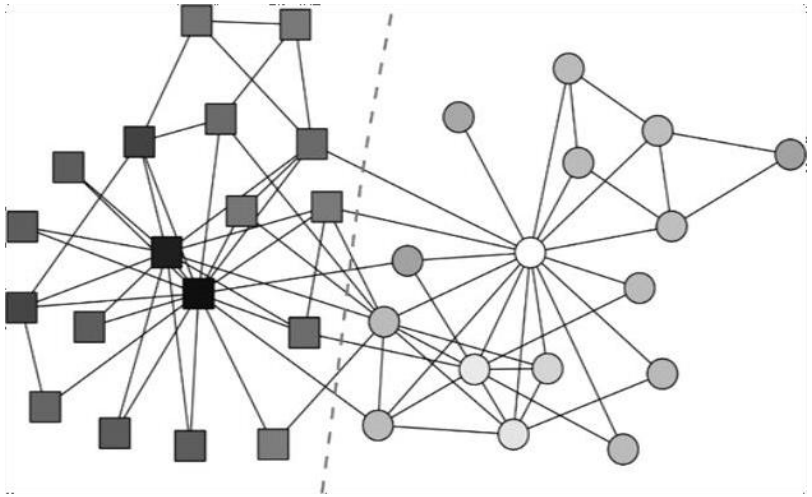



Bailey-Cao-Kuchler-Stroebel-Wong (2018) Journal of Economic Perspectives

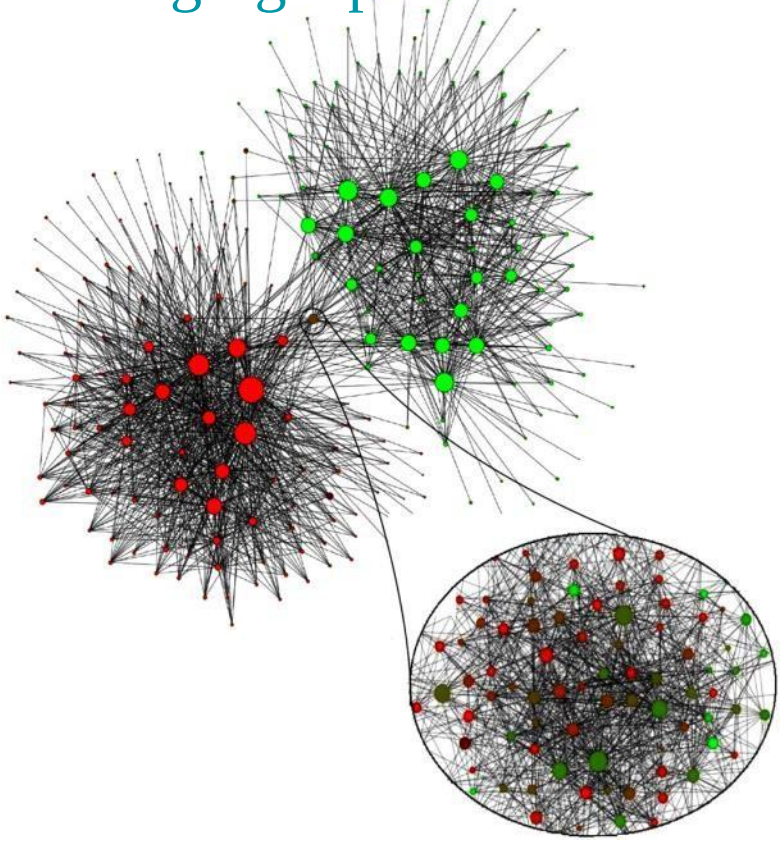
Norbutas-Corten (2018) Social Networks

# 5. Communities

→ Karate Club:
Breakup of the club



→ Belgian Phone Data:
Language spoken



# Basics of Communities

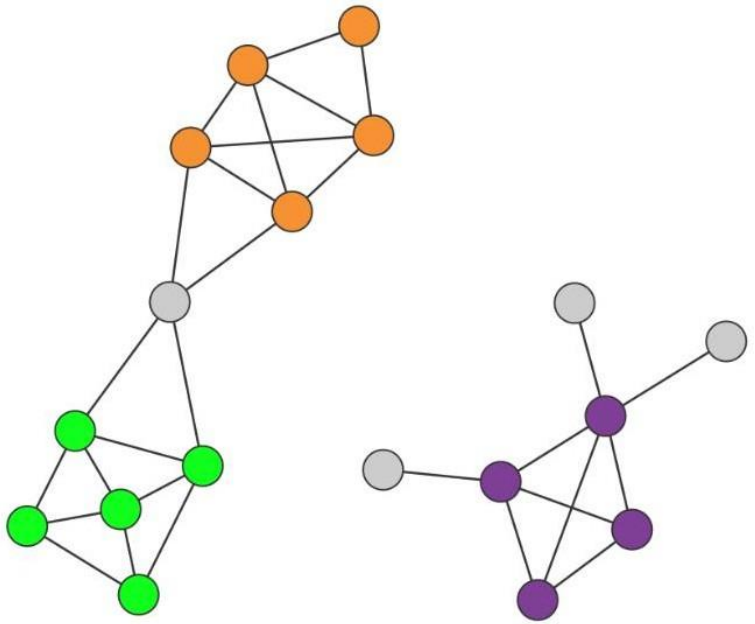
H1: A network's community structure is uniquely encoded in its wiring diagram.

H2: *Connectedness Hypothesis*

A community corresponds to a connected subgraph.

H3: *Density Hypothesis*

Communities correspond to locally dense neighborhoods of a network.

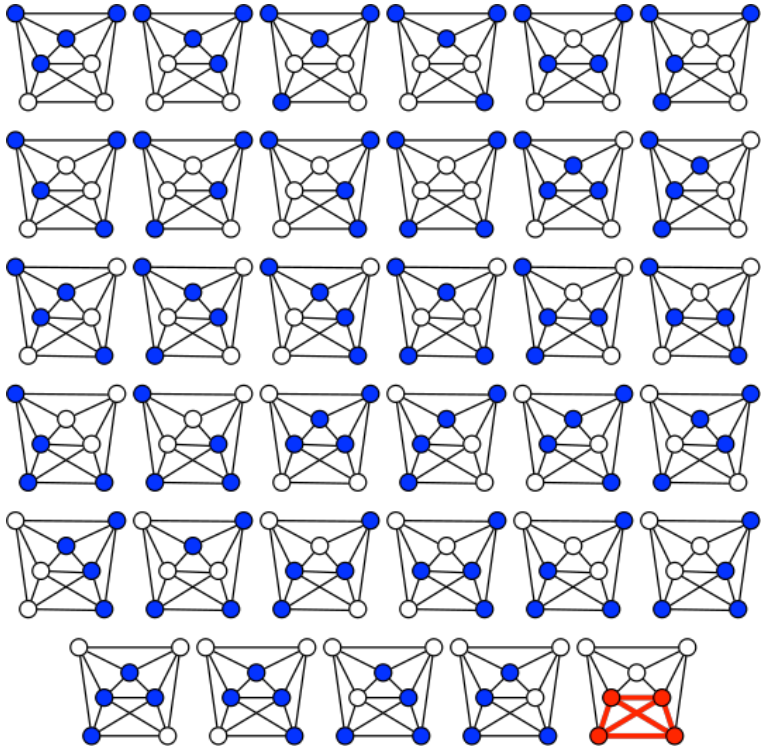


A.-L. Barabási, *Network Science: Communities*.

# Basics of Communities

## Cliques as communities

- Triangles are frequent; larger cliques are rare.
- Communities do not necessarily correspond to complete subgraphs, as many of their nodes do not link directly to each other.
- Finding the cliques of a network is computationally rather demanding.



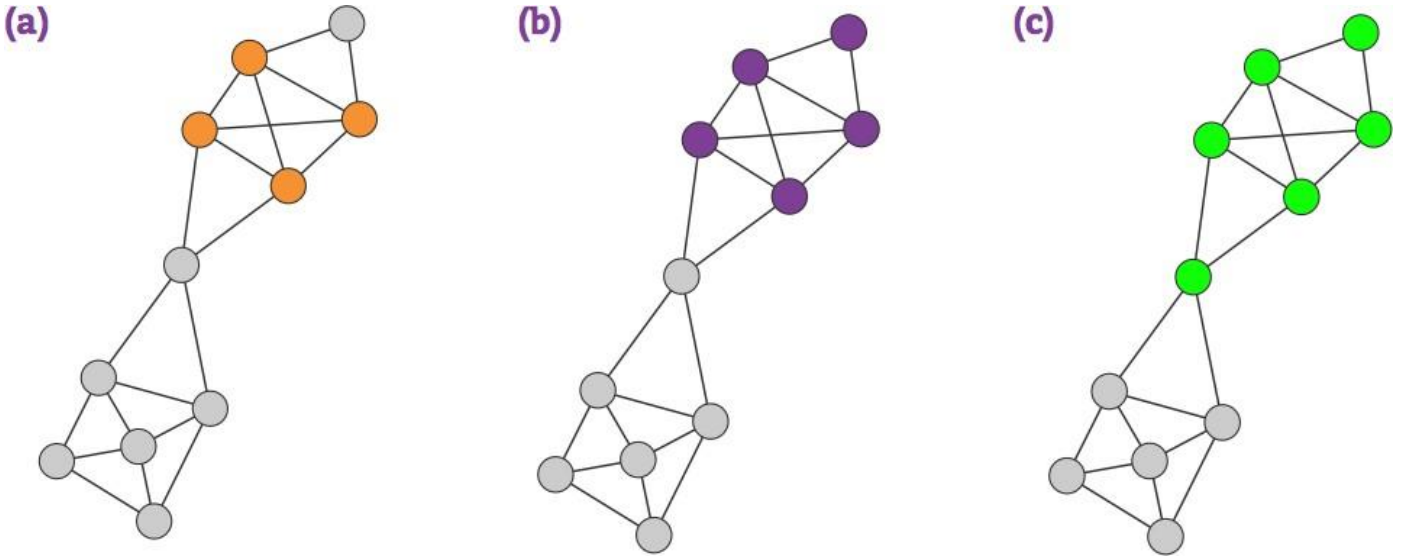
# Basics of Communities

*Strong community:*
Each node of *C* has more links within the community than with the rest of the graph.

$$k_i^{\text{int}}(C) > k_i^{\text{ext}}(C)$$

*Weak community:*
The total internal degree of *C* exceeds its total external degree.

$$\sum_{i \in C} k_i^{\text{in}}(C) > \sum_{i \in C} k_i^{\text{out}}(C)$$




*Clique* *Strong* *Weak*

A.-L. Barabási, *Network Science: Communities*.

# Number of Partitions

How many ways can we partition a network into 2 communities?

Graph bisection

Divide a network into two equal non-overlapping subgraphs, such that the number of links between the nodes in the two groups is minimized.

Two subgroups of size $n_1$ and $n_2$.

Total number of combinations: $$\frac{N!}{n_1!n_2!}$$

Assuming $$n_1 = n_2 = N/2$$

N=10 → 256 partitions (1 ms)

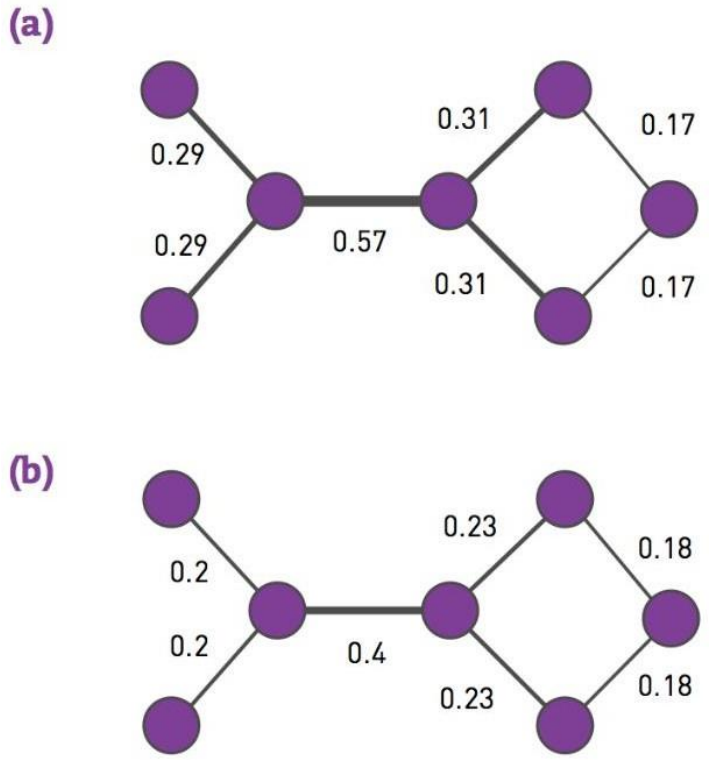
N=100→ $10^{26}$ partitions ($10^{21}$ years)

# Divisive Algorithms

Divisive algorithms split communities by removing links that connect nodes with low similarity.

**Step 1: Define a Centrality Measure (Girvan-Newman algorithm)**

- *Link betweenness* is the number of shortest paths between all node pairs that run along a link.
- *Random-walk betweenness*. A pair of nodes *m* and *n* are chosen at random. A walker starts at *m*, following each adjacent link with equal probability until it reaches *n*. Random walk betweenness $x_{ij}$ is the probability that the link $i \rightarrow j$ was crossed by the walker after averaging over all possible choices for the starting nodes *m* and *n*



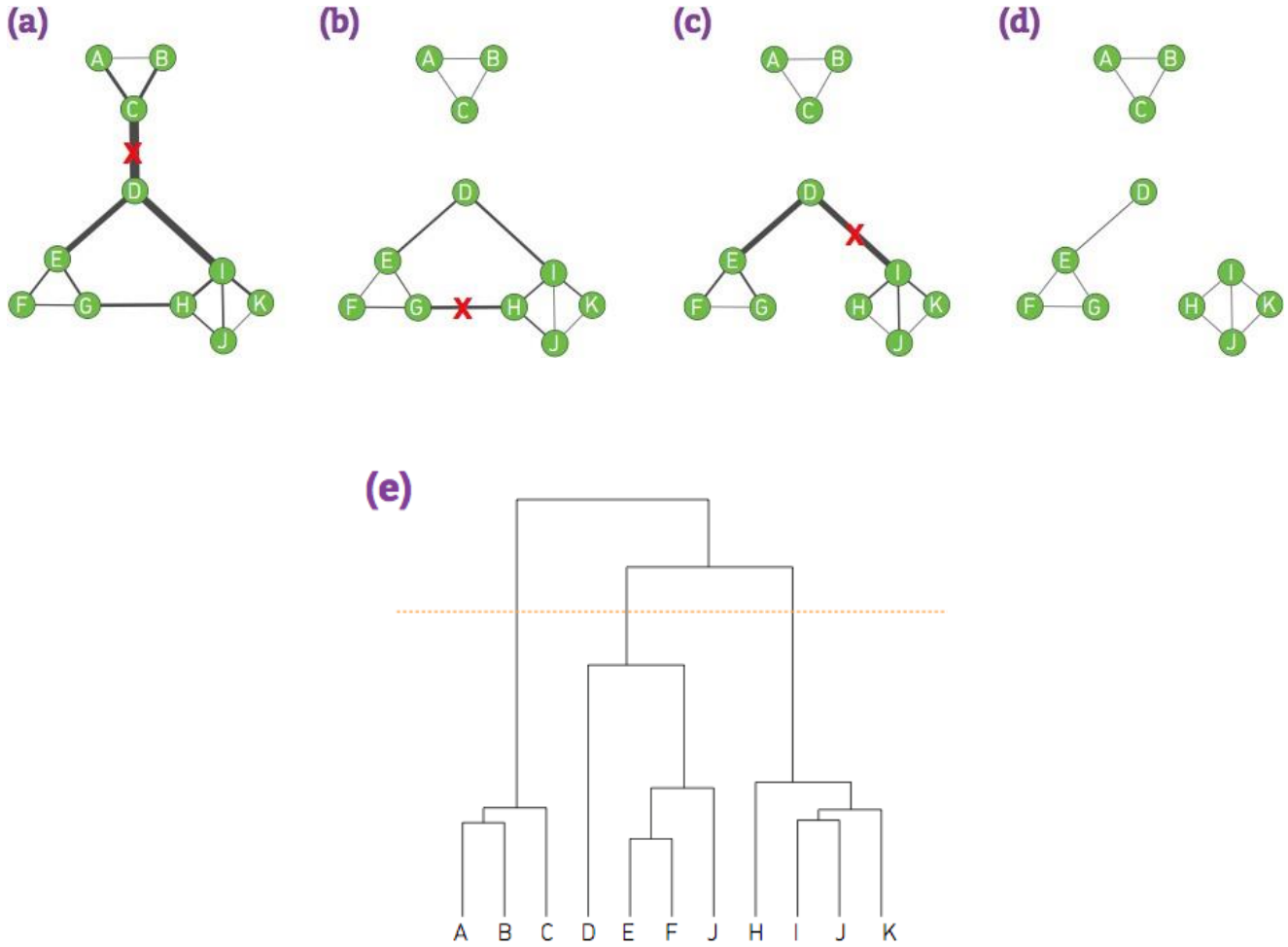

M. Girvan & M.E.J. Newman, *PNAS* 99 (2002).

A.-L. Barabási, *Network Science: Communities*.

# Divisive Algorithms

**Step 2: Hierarchical Clustering**

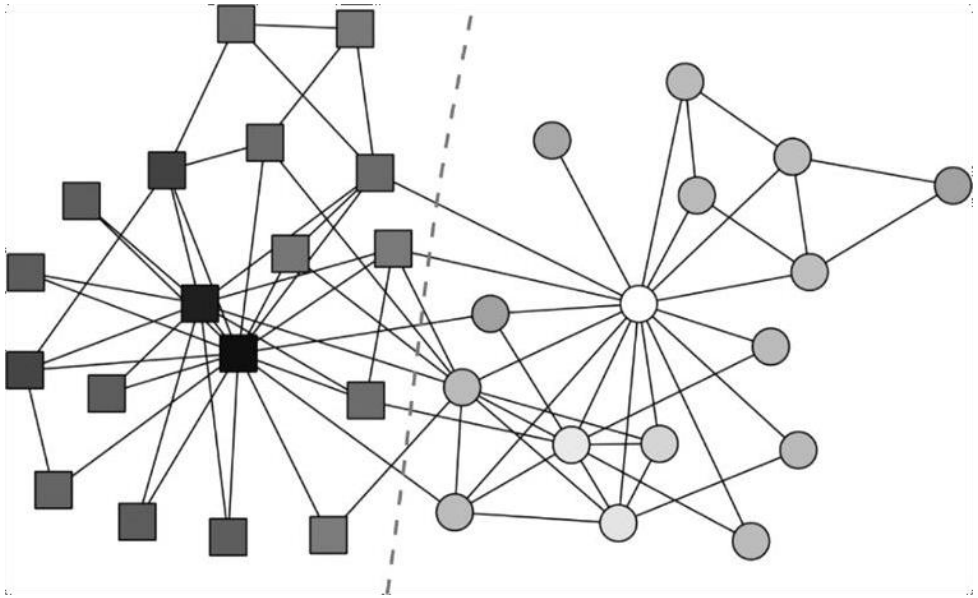
a) Compute of the centrality of each link.
b) Remove the link with the largest centrality; in case of a tie, choose one randomly.
c) Recalculate the centrality of each link for the altered network.
d) Repeat until all links are removed (yields a dendrogram).




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

A.-L. Barabási, *Network Science: Communities*.

# Divisive Algorithms



**Computational complexity:**

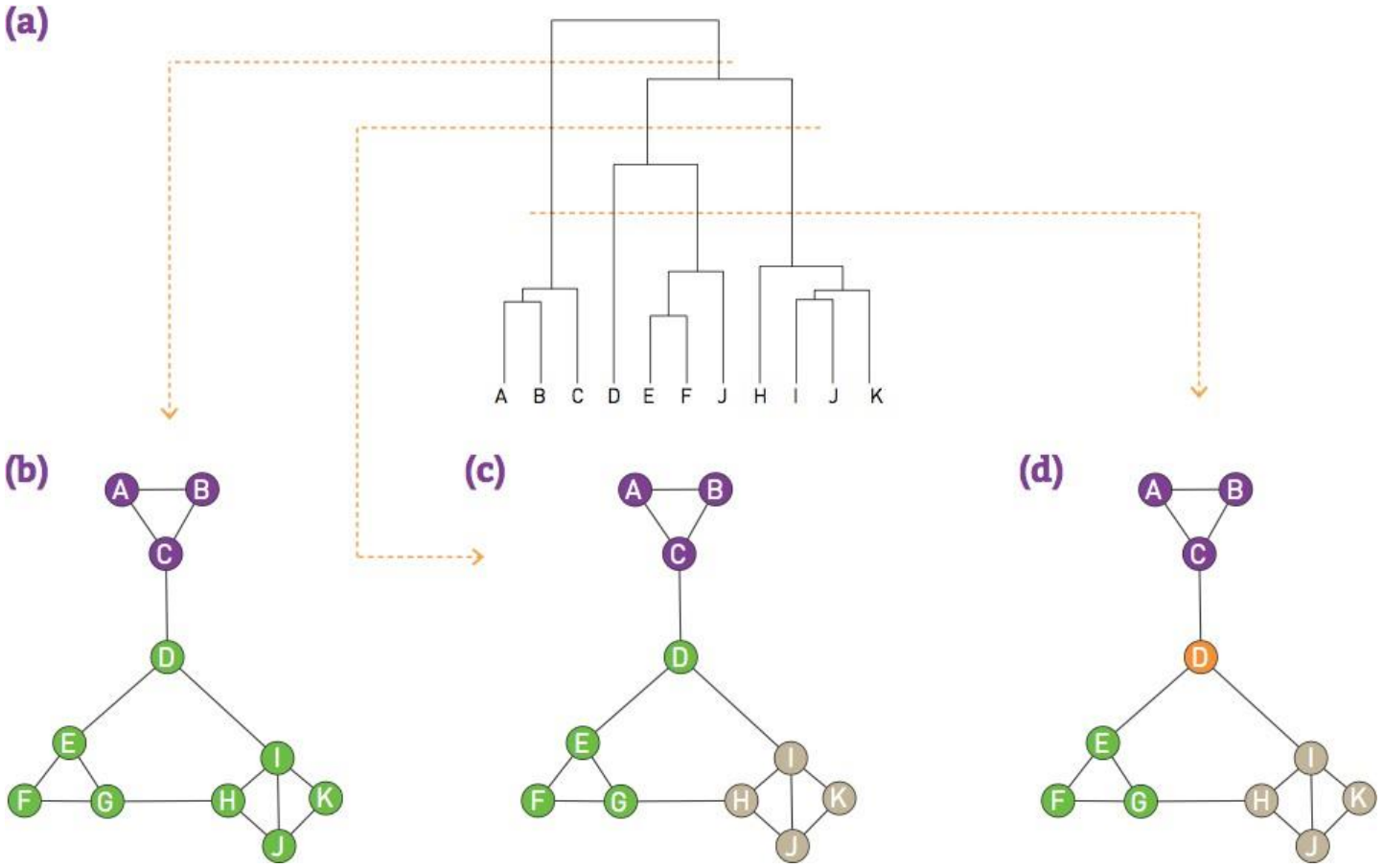
- Step 1a (calculation betweenness centrality): $O\left(N^2\right)$ *for sparse networks* $\Rightarrow$ $O\left(N^3\right)$
- Step 1b (Recalculation of betweenness centrality for all links): $O\left(LN^2\right)$

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

A.-L. Barabási, *Network Science: Communities*.

# Ambiguity in Hierarchical clustering

## Where to "cut"?




A.-L. Barabási, *Network Science: Communities*.

# Modularity

H4: *Random Hypothesis*

Randomly wired networks are not expected to have a community structure.

Imagine a partition in $n_c$ communities $\{C_c,\ c = 1,\ n_c\}$

Modularity $$M(C_c) = \frac{1}{2L}\sum_{i,j=1}^{N}(A_{ij} - P_{ij})\delta(C_i - C_j)$$

Original data

Expected connections, a model

Relative to a specific partition

→ Random network $P_{ij} = \frac{k_i k_j}{2L}$

→ Modularity is a measure associated to a partition

MEJ Newman, *PNAS* 103 (2006).

A.-L. Barabási, *Network Science: Communities*.

# Modularity

## Another way of writing M

$L_C$ is the number of links within $C$ and $k_c$ is total degree of nodes within community

$$M(C_c) = \sum_{c=1}^{n_c} \left[ \frac{l_c}{L} - \left( \frac{k_c}{2L} \right)^2 \right]$$

H5: *Maximal Modularity Hypothesis*

The partition with the maximum modularity $M$ for a given network offers the optimal community structure

**Goal** Find $\{C_c, c = 1, n_c\}$ that maximizes $M$

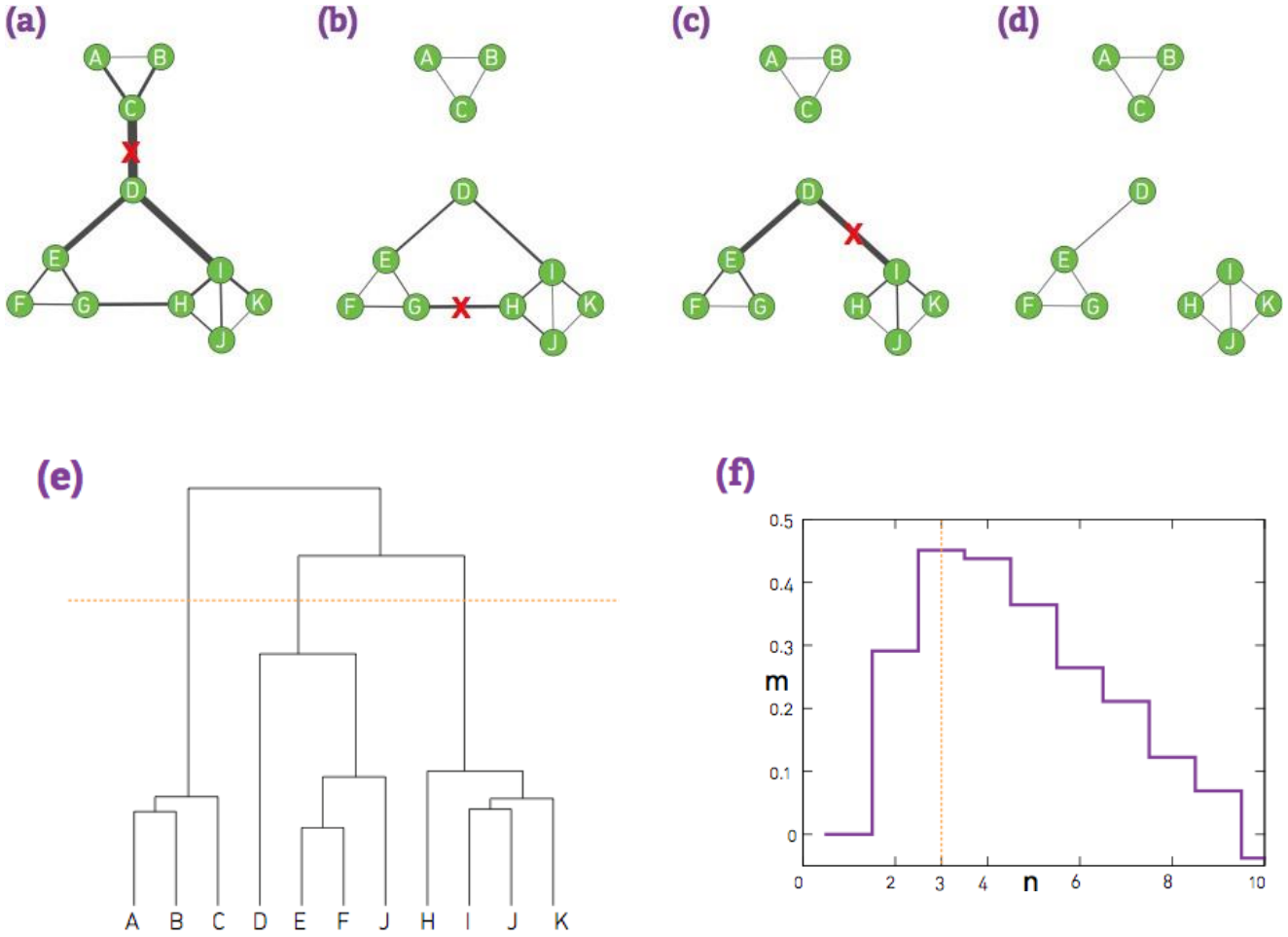
MEJ Newman, *PNAS* 103 (2006).

A.-L. Barabási, *Network Science: Communities*.

# Modularity for the Girvan-Newman

Which partition $\{C_c, c = 1, n_c\}$ ?

$$M(\mathcal{C}) = \sum_{c=1}^{n_c} \left[ \frac{l_c}{L} - \left( \frac{k_c}{2L} \right)^2 \right]$$




A.-L. Barabási, *Network Science: Communities*.

# Modularity based community identification

A *greedy algorithm*, which iteratively joins nodes if the move increases the new partition's modularity.

Step 1. Assign each node to a community of its own. Hence we start with $N$ communities.

Step 2. Inspect each pair of communities connected by at least one link and compute the modularity variation obtained if we merge these two communities.

Step 3. Identify the community pairs for which $\Delta M$ is the largest and merge them. Note that modularity of a particular partition is always calculated from the full topology of the network.

Step 4. Repeat step 2 until all nodes are merged into a single community.

Step 5. Record for each step and select the partition for which the modularity is maximal.

MEJ Newman, *PRE 69* (2004).

A.-L. Barabási, *Network Science: Communities*.

# Louvain method

$$Q = \frac{1}{2m} \sum_{i,j} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

where $A_{ij}$ represents the weight of the edge between $i$ and $j$, $k_i = \sum_j A_{ij}$ is the sum of the weights of the edges attached to vertex $i$, $c_i$ is the community to which vertex $i$ is assigned, the $\delta$ function $\delta(u, v)$ is 1 if $u = v$ and 0 otherwise and $m = \frac{1}{2} \sum_{ij} A_{ij}$.

Modularity gain of moving isolated node *i* to community *C:*

$$\Delta Q = \left[ \frac{\sum_{\text{in}} + 2k_{i,\text{in}}}{2m} - \left( \frac{\sum_{\text{tot}} + k_i}{2m} \right)^2 \right] - \left[ \frac{\sum_{\text{in}}}{2m} - \left( \frac{\sum_{\text{tot}}}{2m} \right)^2 - \left( \frac{k_i}{2m} \right)^2 \right]$$

where $\sum_{\text{in}}$ is the sum of the weights of the links inside $C$, $\sum_{\text{tot}}$ is the sum of the weights of the links incident to nodes in $C$, $k_i$ is the sum of the weights of the links incident to node $i$, $k_{i,\text{in}}$ is the sum of the weights of the links from $i$ to nodes in $C$ and $m$ is the sum of the weights of all the links in the network.

# Louvain method