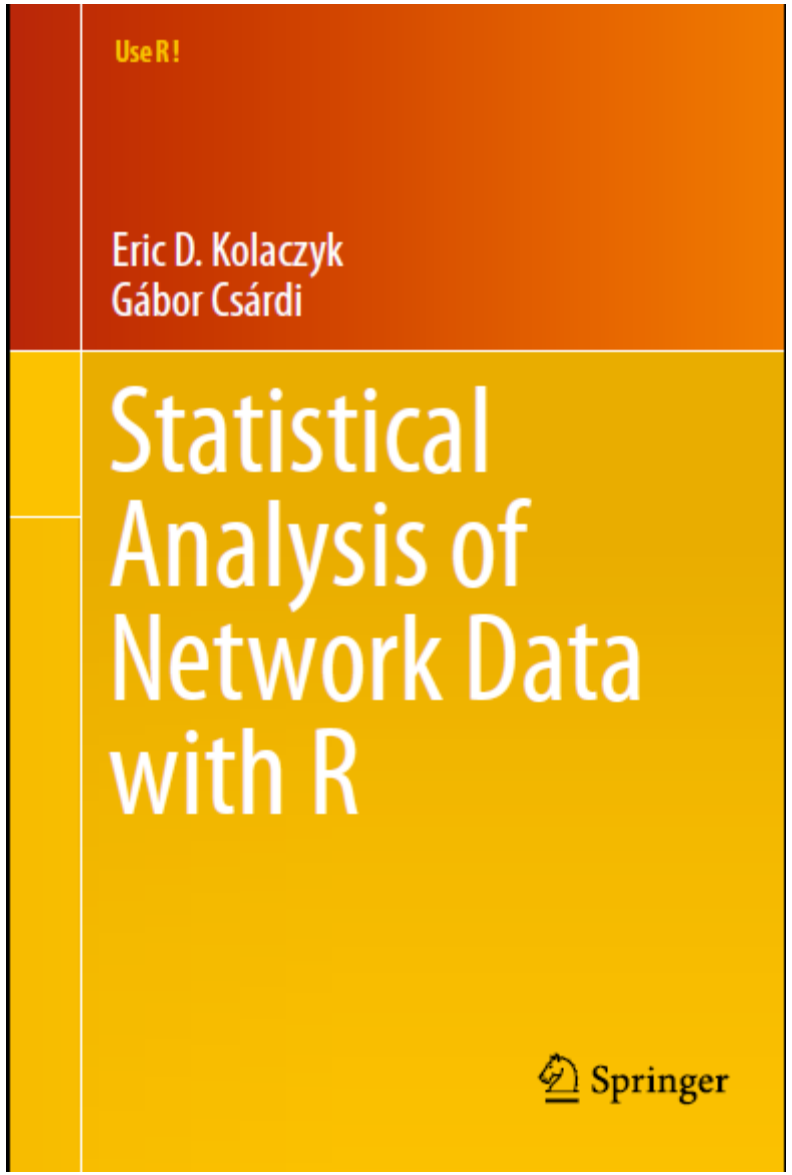
Network and Spatial Analyses
26. February 2020

Seminar: Intro network exercise in R

1. Network data
2. Bi-partite networks
3. Network and node characteristics
4. Exercise: co-inventor networks
5. Community structure
6. Exercise: the global airline network

Course Github page: https://github.com/bokae/anet_course



Introduction to the ‘igraph’ package

‘sand’ package: integration + education data

Codes by chapters: https://github.com/kolaczyk/sand

AT THE END OF THE CLASS, CODES SHOULD BE SAVED AS "NAME_ELTE_ANET.R".
PLEASE SEND THE FILE TO lengyel.balazs@krtk.mta.hu.

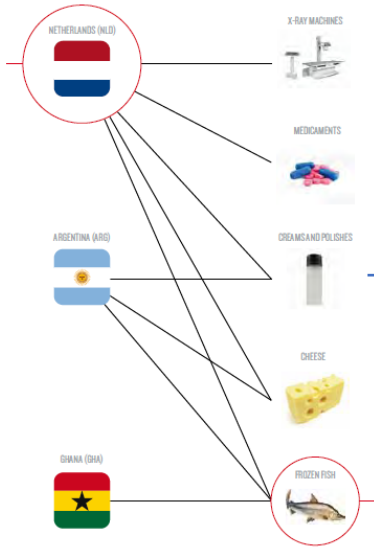
1. Network data

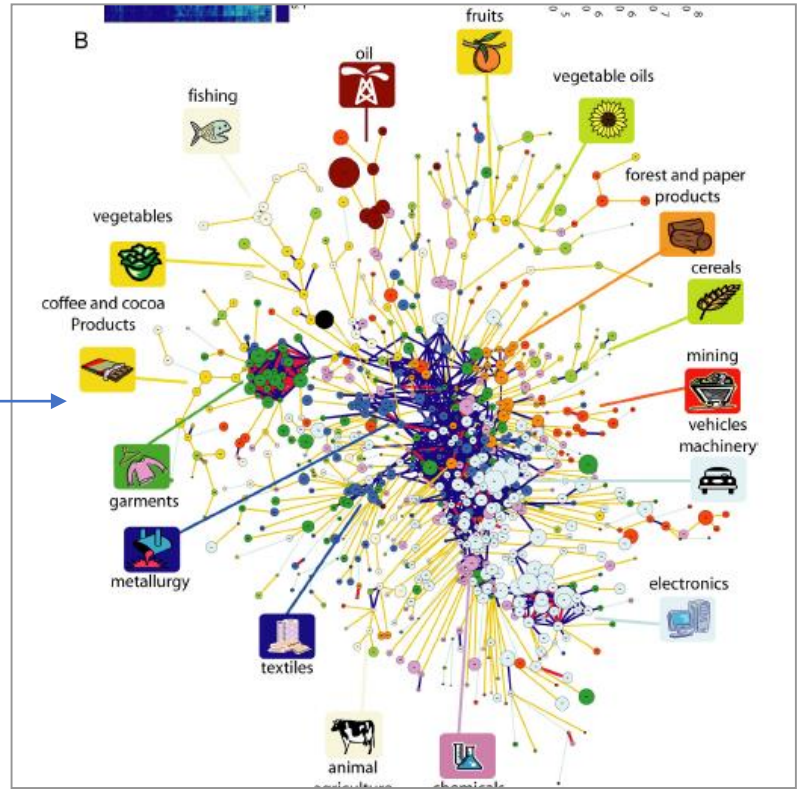
- Store in matrices
  - BUT: matrices cannot store node characteristics
    matrices are often sparse, which makes data-management inefficient

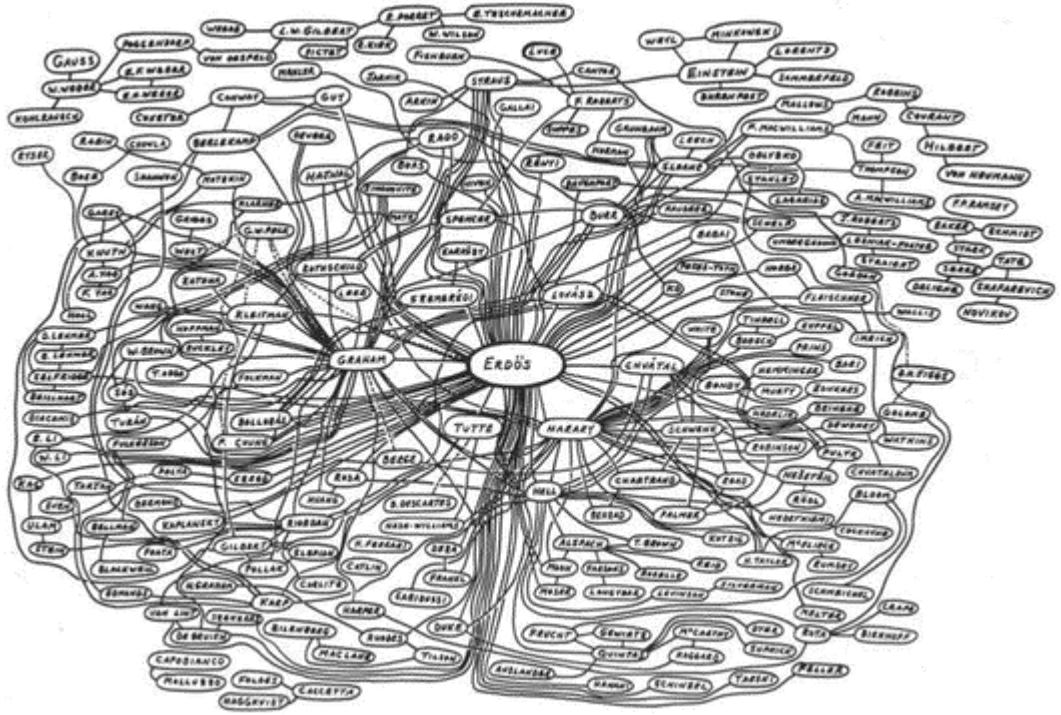
- Store in two files
  - Node list: ID (must correspond to IDs in edgelist), node characteristics
  - Edge list: ID1, ID2, edge characteristics

  - Edges are usually not repeated: if there is A-B link, no need for B-A
  - Rational for repeated edges:
    - Directed edges: A-B is different from B-A
    - Data is stored in many files -> table operations are needed

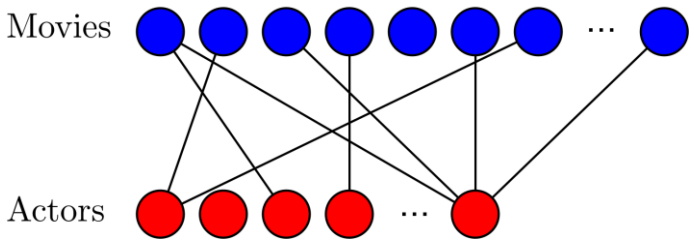
2. Bipartite networks







Hidalgo-Klinger-Barabási-Hausmann (2007) The product space conditions the development of nations. *Science*



3. Network and node measures

SEE LECTURE!

1. Degree
2. Betweenness
3. Closeness centrality
4. Density
5. Transitivity
6. Average path length

4. Exercise: co-inventor networks

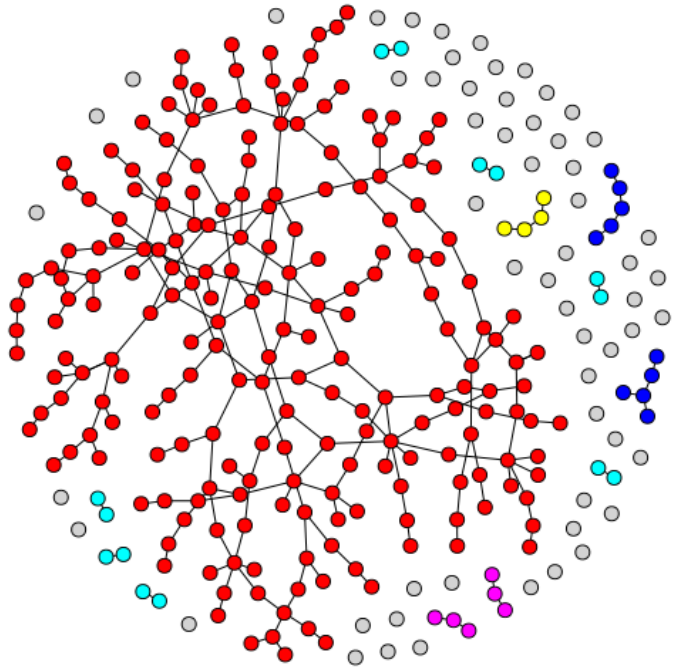
1. Create a simple co-inventor network from “patents.csv”, in which two inventors are linked in case they co-author a patent. The weight of every collaboration in a single patent is 1.

2. Plot degree distribution and strength distribution of the network. Calculate transitivity, density of the network.

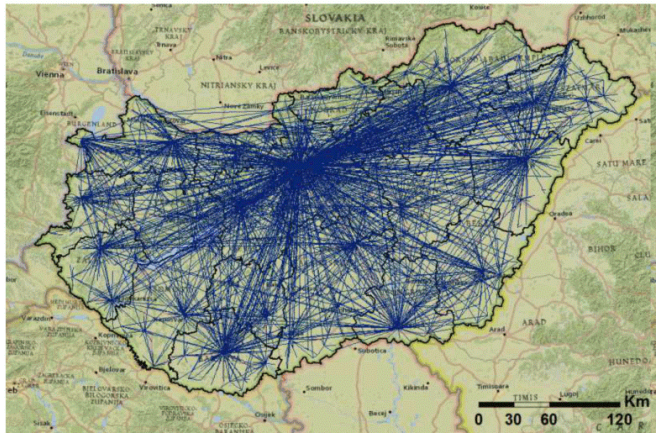
3. Compare co-inventor networks in the US versus in Germany in terms of degree distribution, transitivity, density.

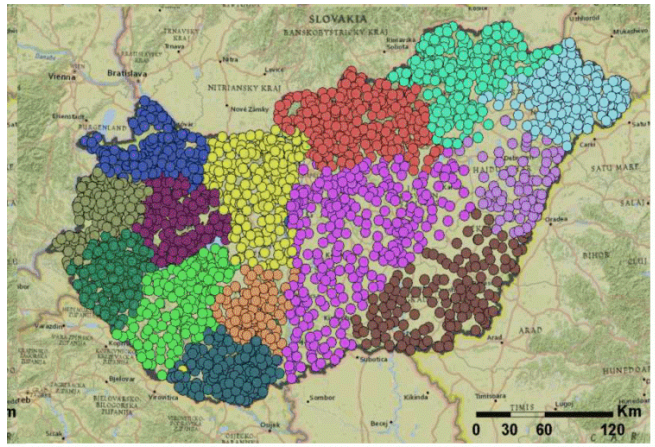
5. Community structure

Components (giant component)



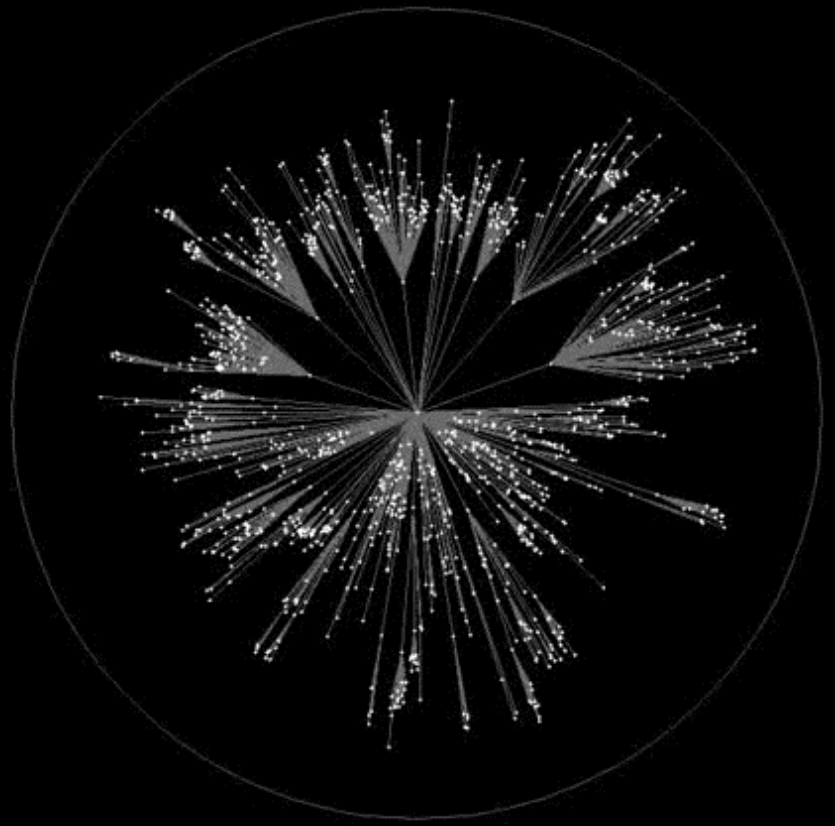
Communities

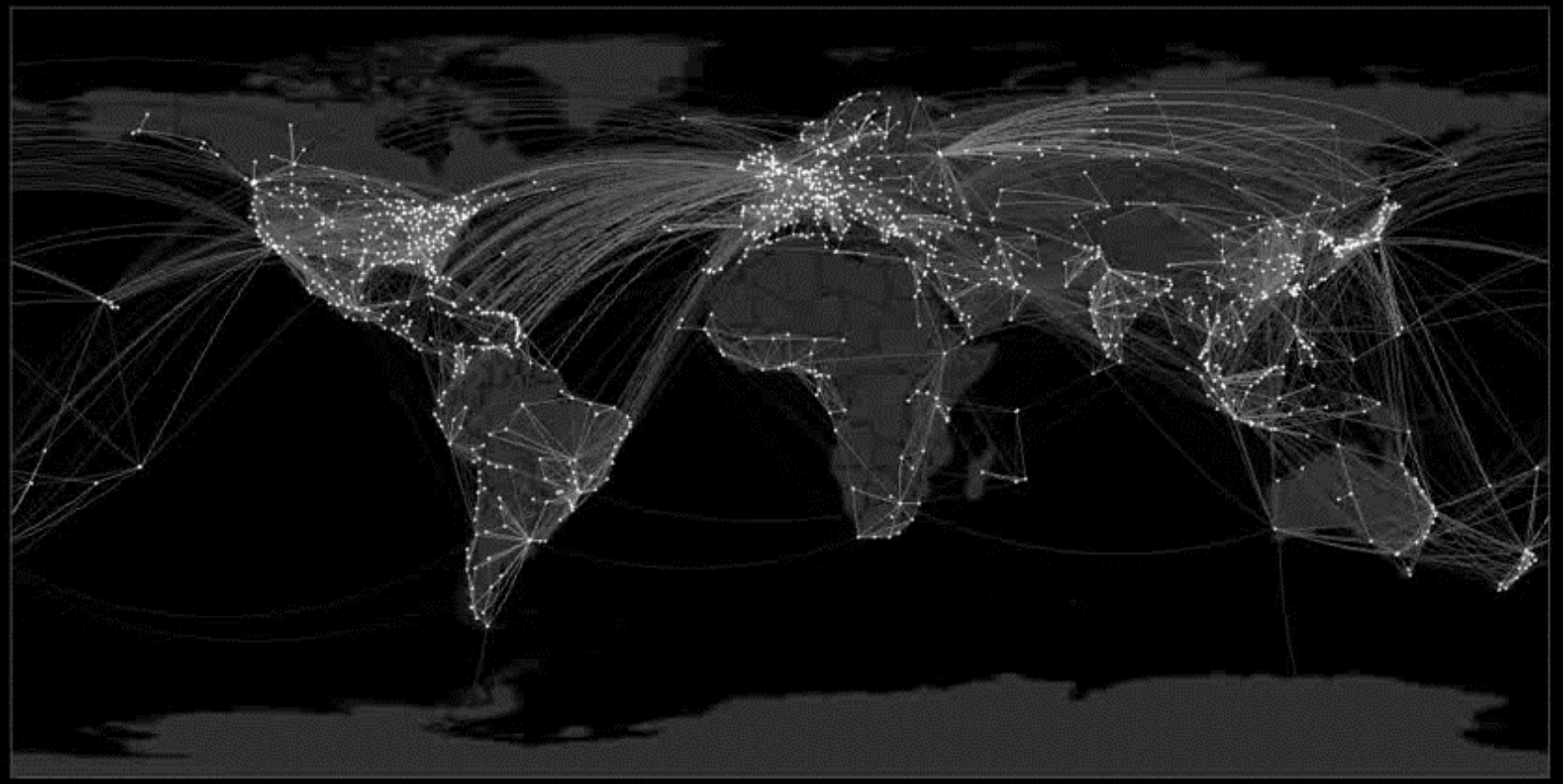




$$Q_i = \sum_{k=1}^{K} \left[\frac{L_k^w}{L_i} - \left(\frac{L_k}{L_i} \right)^2 \right]$$

6. Exercise: the global airline network





Brockman D., Helbing D (2013)The hidden geometry in complex, network-driven contagion phenomena. Science 342 (6164) pp. 1337-1342.

6. Exercise: the global airline network

- In case of a new virus, airlines are the prior global transporter of the disease.

1. How can we characterize the network to describe potential diffusion?
2. What airports should we block to decrease diffusion potential?