Network Analysis with Pajek

BOJAN SRBINOSKI (BSRBINOSKI@LIUC.IT)

UNIVERSITY CATTANEO – LIUC (CASTELLANZA, ITALY)

But first… You said… Macedonia?? Where is that???

My home - Ohrid

Agenda

While I am talking…

Go to http://mrvar.fdv.uni-lj.si/pajek/ or just be smarter and ask Google (“pajek download”)

Click on the green “Pajek” or “Install-Zip”

for MAC and LINUX, scroll down and you will find instructions

If it doesn’t work, don’t panic. Probably your friend next to you has it.

and yes.. The agenda:

Properties of a network

Types of networks

Measures of the properties

Two mode network

Applications

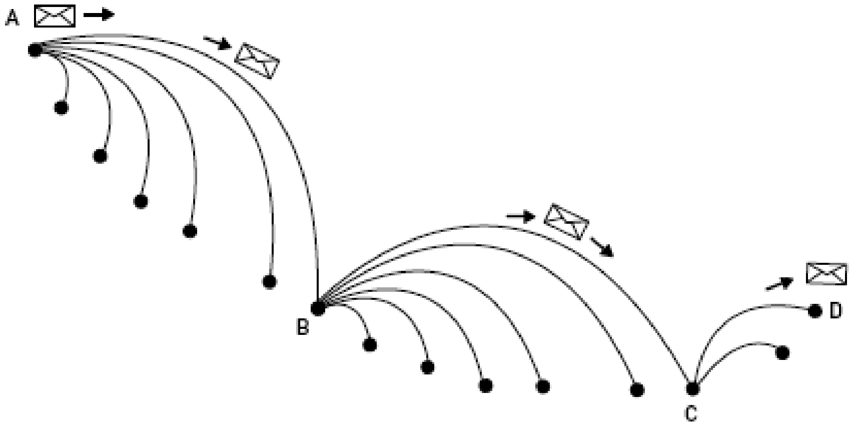
Pajek

Properties of a network: Small World

Example from S. Milgram, *The small world problem,* Psych. Today, 2 (1967), pp. 60-67.

A group of people from Omaha (Nebraska) was asked to send a letter to an unknown person in Boston (Massachussets)

Rules of the experiment: people should forward the letter to a person that they consider closer to the target person



Results of one experiment (there were more):

- 232 out of 296 letters never reached the target
- 64 letters reached the target (with path from two to 10)
- The average path length was 5.2 steps
- **Question:** Assuming the same rules of the experiment apply and you need to send a letter to President Trump, who would you choose first to give the letter and how many steps do you think you need to accomplish the task?

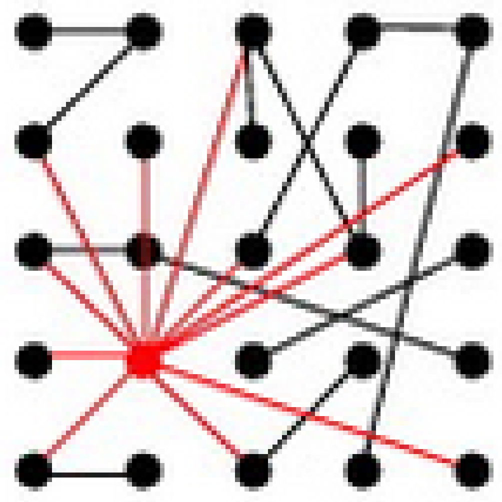
Properties of a network: Hubs

Few nodes with high degree

Robustness to random attacks

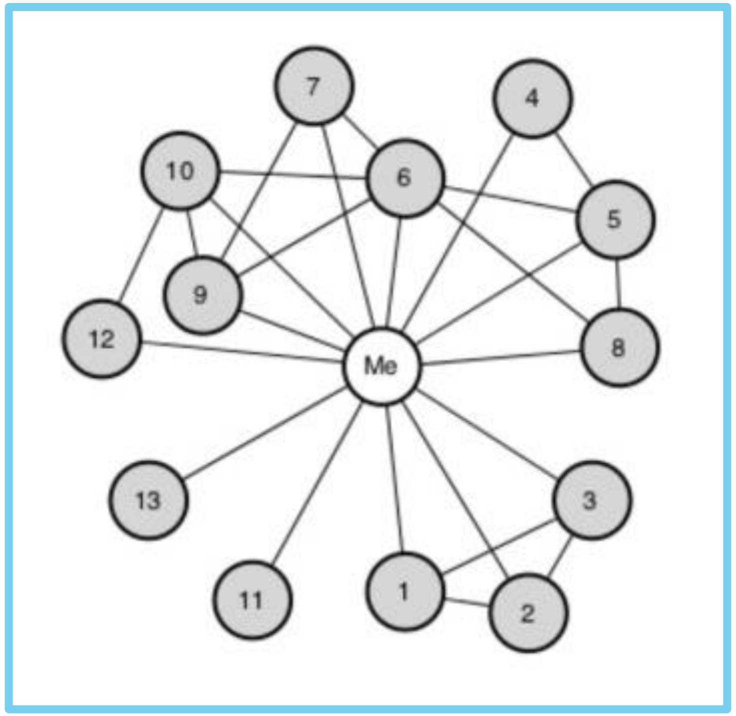
Vulnerability to selective attack

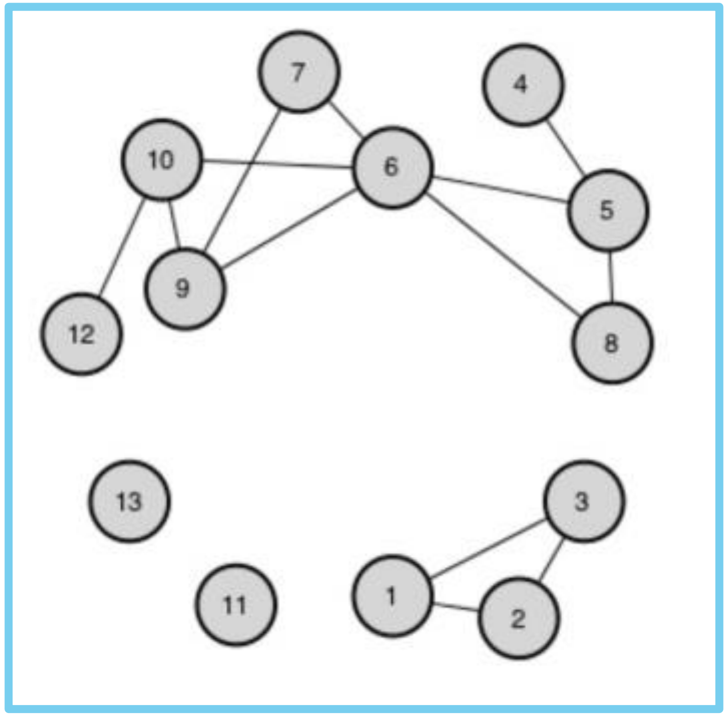
Example: Airlines network



Properties of a network: Clustering

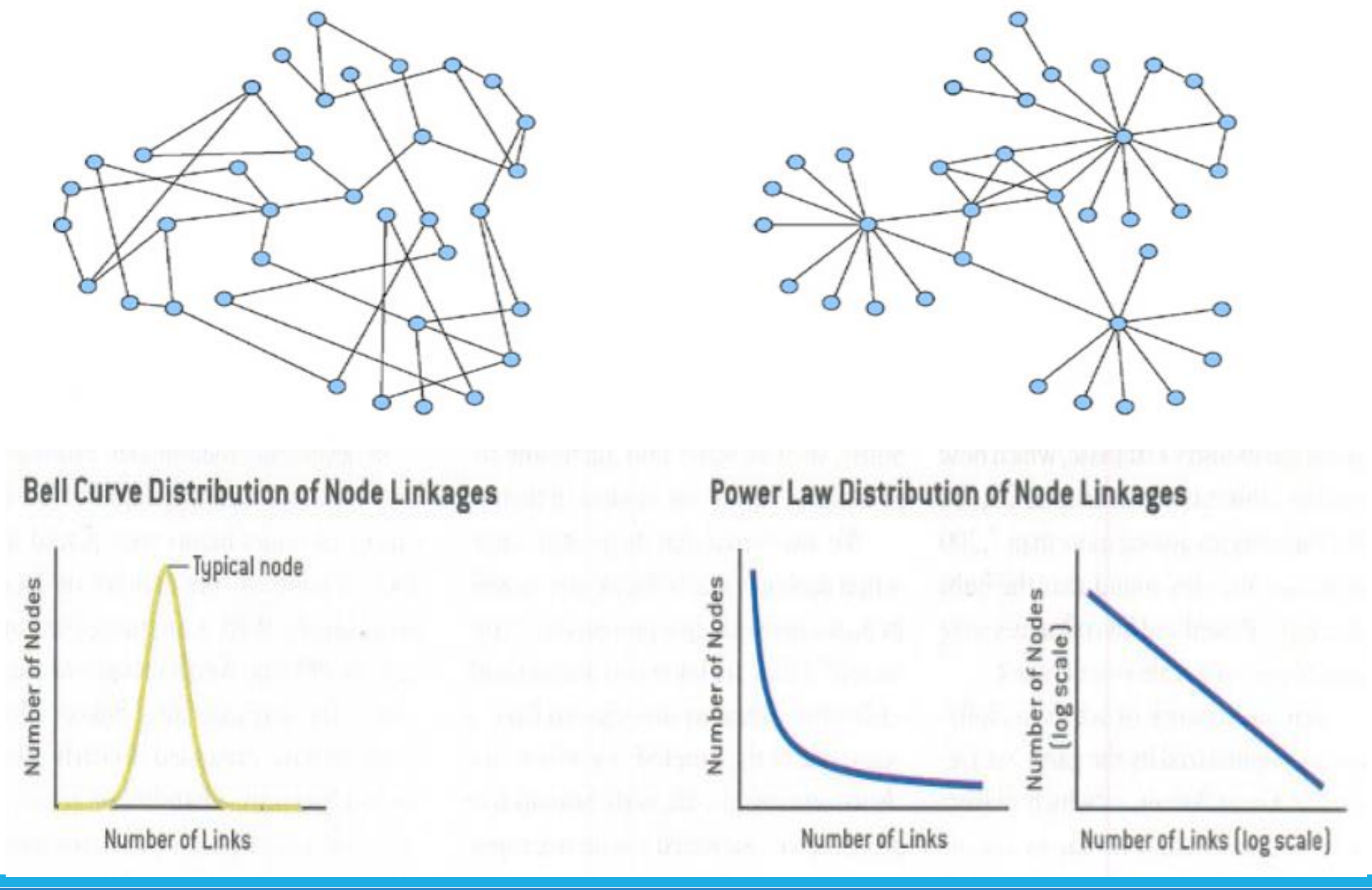
My friends are friends with each other



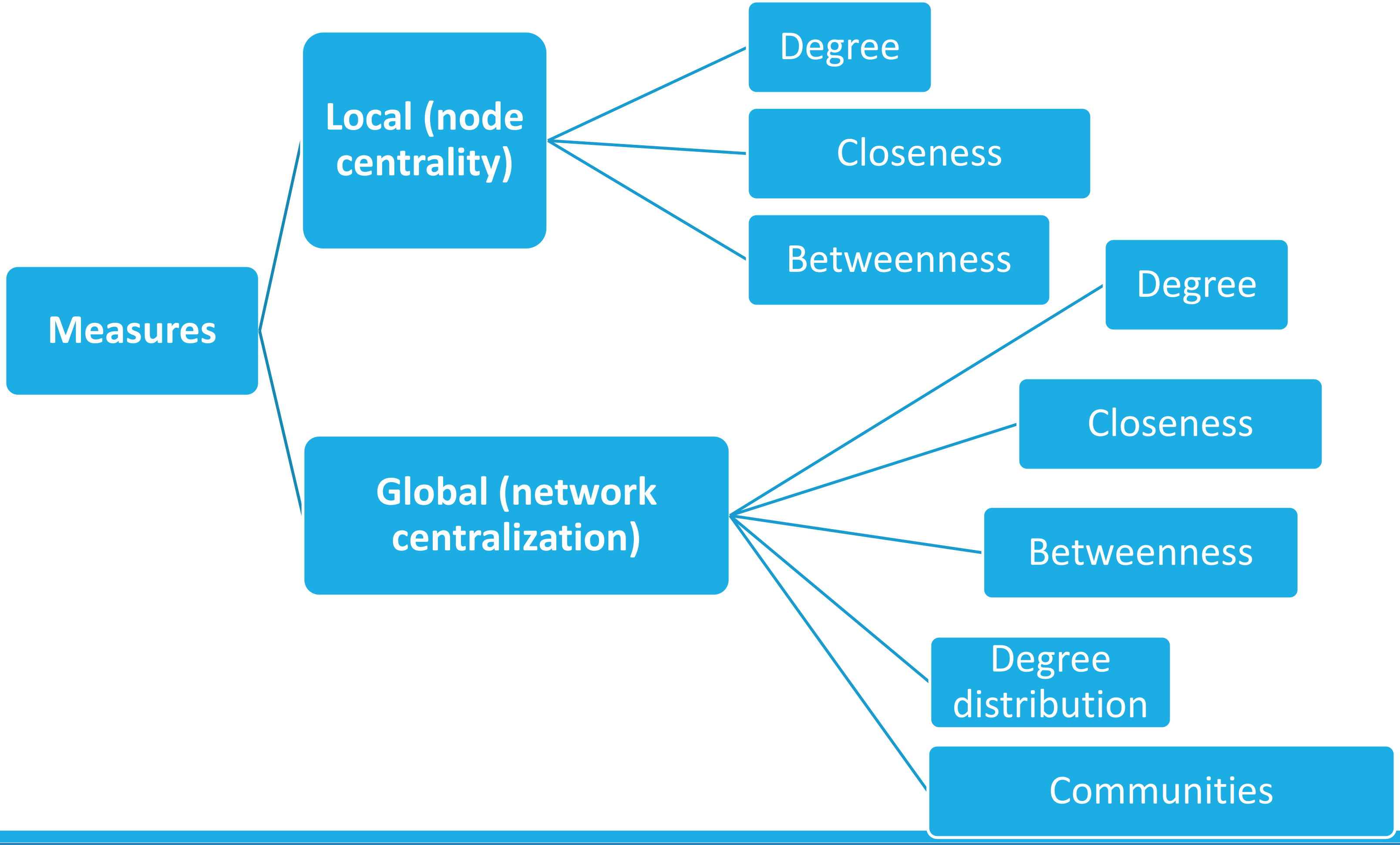


Facebook ego-network

Types of networks: Small world network (left) and Scale free network (right)



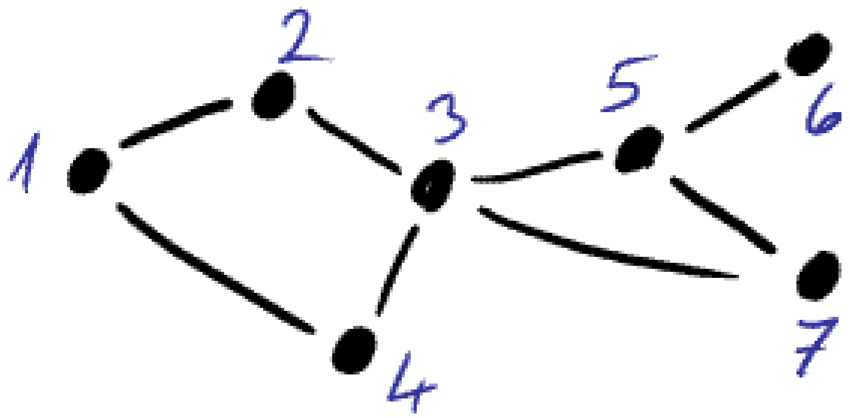
Measures


Measures
Local (node centrality)
Degree
Closeness
Betweenness
Global (network centralization)
Degree
Closeness
Betweenness
Degree distribution
Communities

Distance between nodes

The ***path length*** between *i* and *j* is the number of links going from *i* to *j*

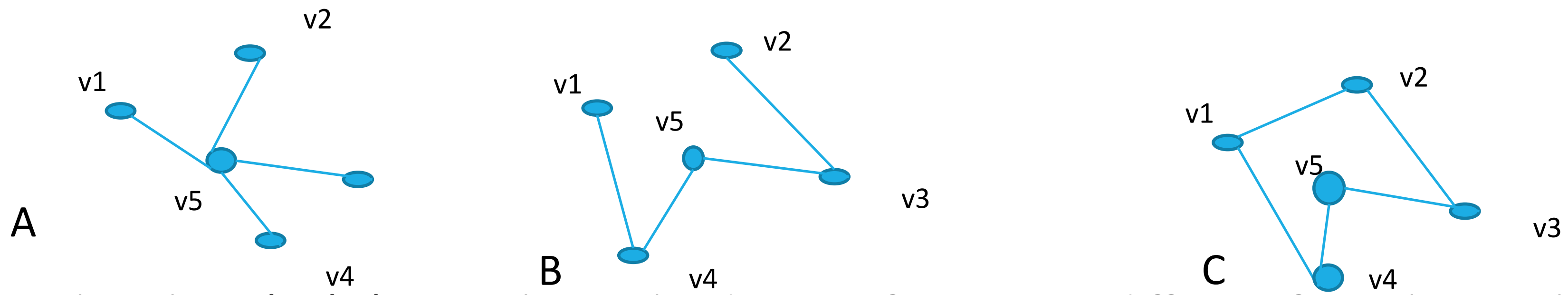
The ***distance*** from *i* to *j* is the length of the minimum path between *i* and *j*.



Degree centrality and degree centralization

The **degree centrality** is the simplest indicator of centrality of a node that is the number of its neighbors (its degree).

The **degree centralization** of a network is the measure of how its structure is close to that of a star network.



In a network with **multiple lines** or loops the degree of a vertex is different from the number of neighbors (and centralization can be > 1). In this case it is not appropriate to use the measure of degree centralization.

Closeness centrality and closeness centralization

In a communication network, the information reaches a person more easily and more correctly if the path he has to do is shorter.

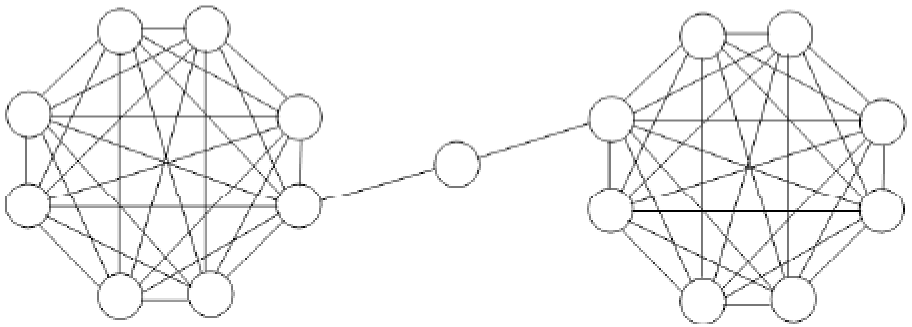
The **Closeness Centrality** of a vertex is the inverse of the sum of the distances of the vertex from the others divided by the number of vertices:

The **Closeness Centralization** is the variation of closeness centrality of vertices divided by the largest such measure can be in a network with the same number of nodes (star network, closeness centralization=1)

Betweenness centrality and betweenness centralization

The **betweenness** of a node is the number of shortest paths (geodesic paths) between other vertices that pass through it, divided by the total number of shortest paths.

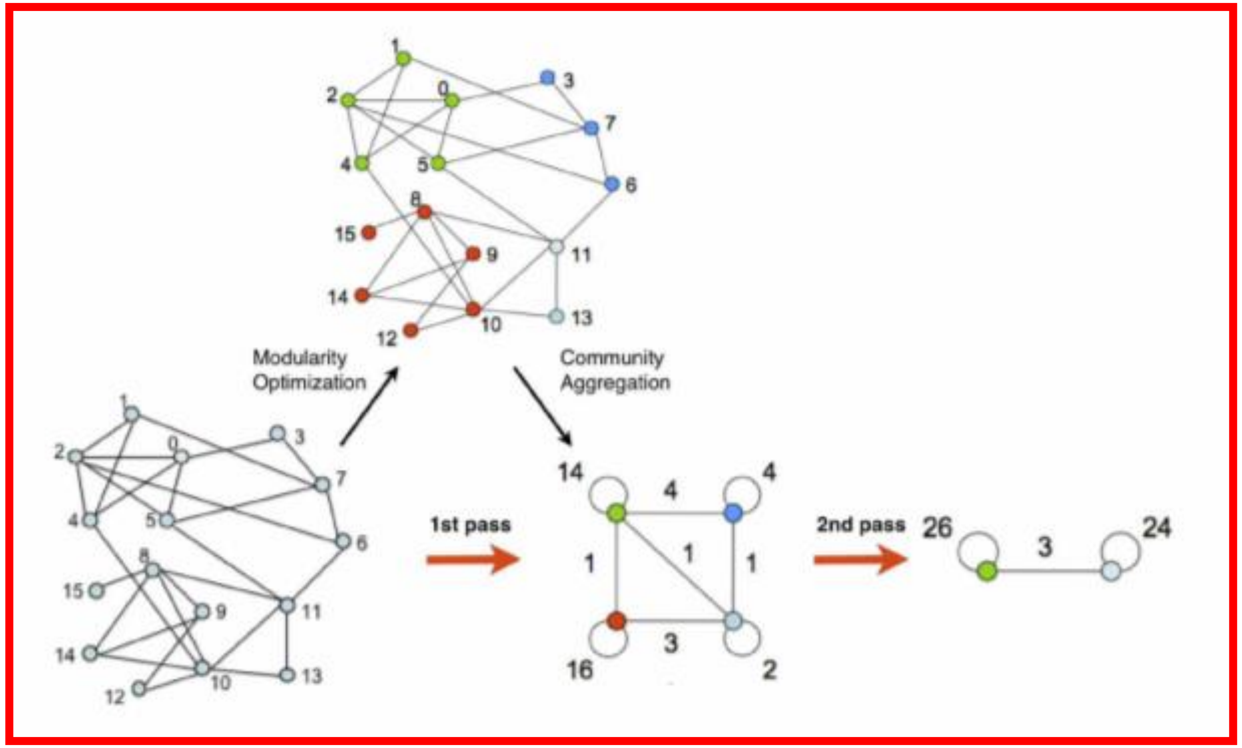
The **betweenness centralization** is the variation of the betweenness centrality of vertices divided by the largest such measure can be in a network with the same number of nodes



Communities

The quality of partition is measured with the Modularity Q is the fraction of the edges that fall within the given groups minus the expected fraction if edges were distributed at random

Networks with **high** modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules.



Two mode network

The measures on the unipartite may have no meaning on bipartite network (e.g. clustering)
Moreover the two projections can have different properties (degree distribution etc.)

By using bipartite network, it is possible to find hidden coalitions between nodes of the other sets

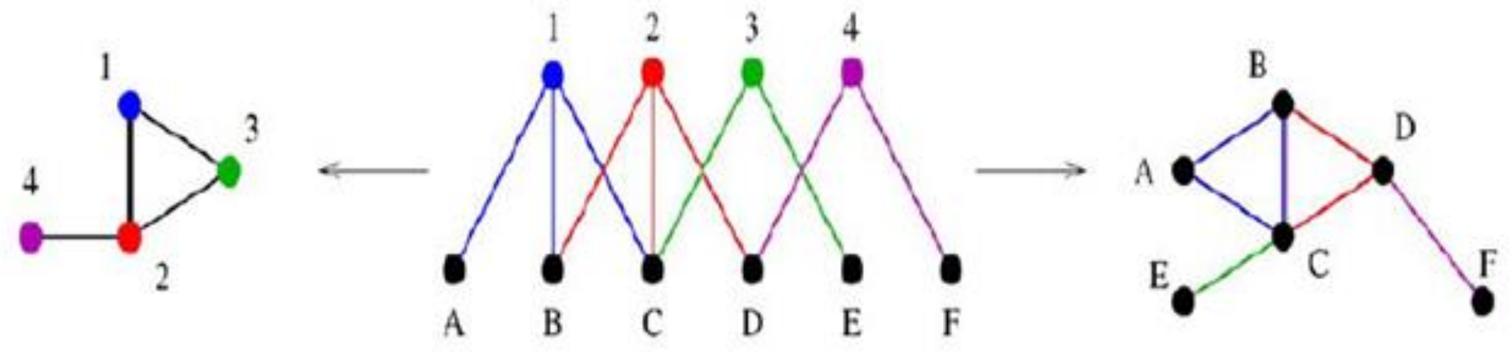


Fig. 1. An example of bipartite graph (center), together with its ⊤-projection (left) and its ⊥-projection (right).

Two mode network: Examples

- Scientific collaboration (authoring network)

Collaboration acts=papers

Actors= authors

- Corporate board and director network

Collaboration acts=boards

Actors= directors

- Occurrence networks

Collaboration acts=sentences

Actors= words

- Peer-to-peer exchange networks

Collaboration acts=data base

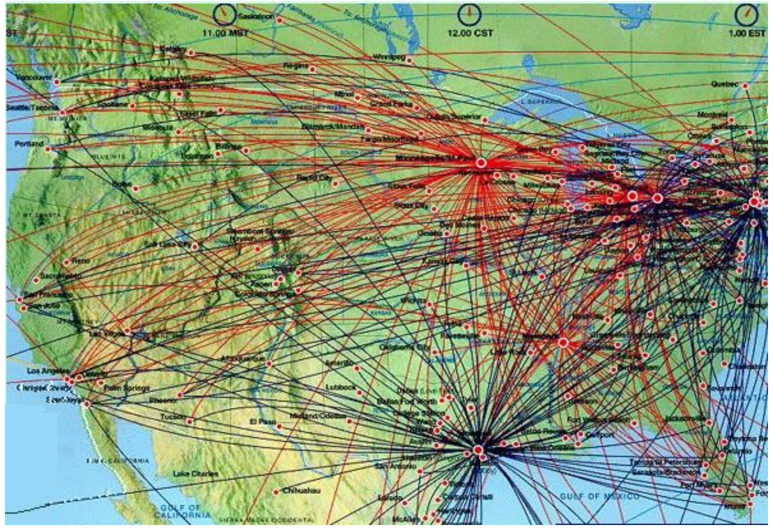
Actors= users

Applications: Air transportation network (Guimera et al., 2005) Communities, degree and betweenness centrality

Incidence: The failures and inefficiencies of the air transportation system have large economic losses. So, why can't we design a better system?

Aim: To analyze the global structure of the worldwide air transportation network

Does this network have small world property (higher clustering)? – Check if the clustering coefficient is significantly higher than the one of random network. The answer is yes.



Applications: Air transportation network
Communities, degree and betweenness centrality

Are the most connected cities also the most important? Compare the degree and betweenness centrality measures of each city. Answer: there cities which are not hubs but are central. Anomaly.

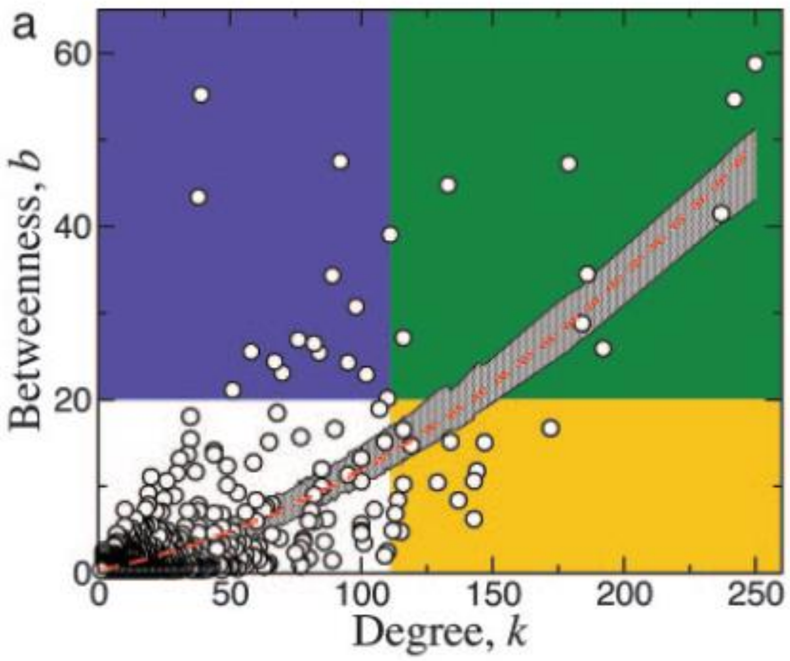


Table 2. The 25 most central cities in the worldwide air transportation network

Rank	City	b	b/b_{ran}	Degree
1	Paris	58.8	1.2	250
2	Anchorage*	55.2	16.7	39
3	London	54.7	1.2	242
4	Singapore*	47.5	4.3	92
5	New York	47.2	1.6	179
6	Los Angeles	44.8	2.3	133
7	Port Moresby*	43.4	13.6	38
8	Frankfurt	41.5	0.9	237
9	Tokyo	39.1	2.7	111
10	Moscow	34.5	1.1	186
11	Seattle*	34.3	3.3	89
12	Hong Kong*	30.8	2.6	98
13	Chicago	28.8	1.0	184
14	Toronto	27.1	1.8	116
15	Buenos Aires*	26.9	3.2	76
16	São Paulo*	26.5	2.8	82
17	Amsterdam	25.9	0.8	192
18	Melbourne*	25.5	4.5	58
19	Johannesburg*	25.4	2.6	84
20	Manila*	24.4	3.5	67
21	Seoul*	24.3	2.1	95
22	Sydney*	23.1	3.2	70
23	Bangkok*	22.9	1.8	102
24	Honolulu*	21.1	4.4	51
25	Miami*	20.1	1.4	110

Cities are ordered according to their normalized betweenness. We also show the ratio of the actual betweenness of the cities to the betweenness that they have after randomizing the network.

*These cities are not among the 25 most connected.

Applications: Air transportation network
Communities, degree and betweenness centrality

Next problem: the existence of nodes with anomalous centrality is related to the existence of regions with a high density of airports but few connections to the outside. Communities???

Calculate the maximum modularity and then, check the patterns of intracommunity and intercommunity connections (a bit complex procedure)

Conclusion: most of the nodes (95%) are peripheral; that is, the vast majority of their connections are within their own communities. Also, the nodes that connect different communities are typically hubs within their own community, although not necessarily global hubs

Applications: Communities in Italian corporate networks (Piccardi et al., 2010)
Two mode network

Aim: to analyze and compare the community structure of two real-world financial networks (board and ownership networks) based on the firms listed in the Italian Stock Exchange

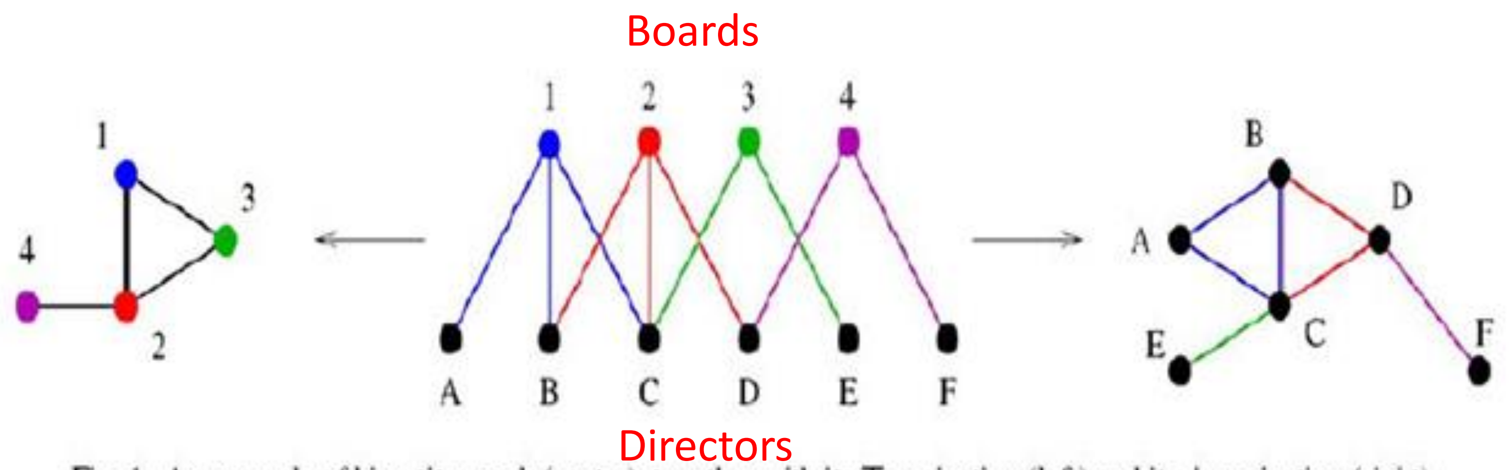
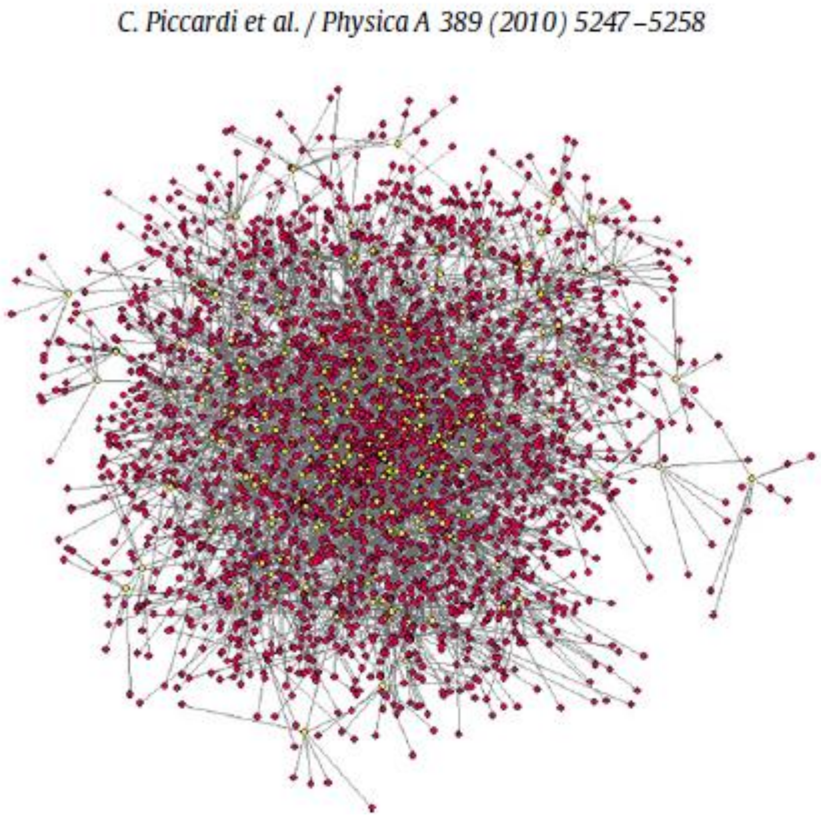
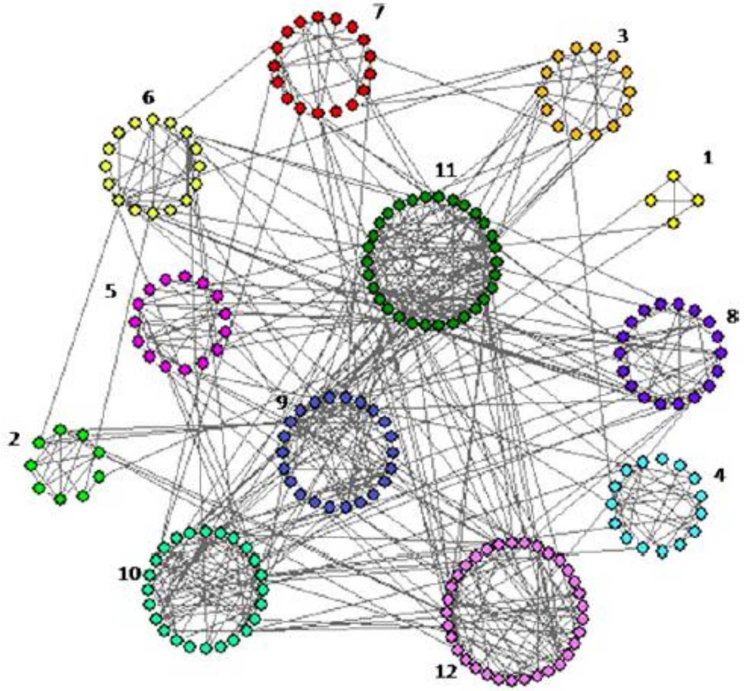


Fig. 1. An example of bipartite graph (center), together with its ⊤-projection (left) and its ⊥-projection (right).

Applications: Communities in Italian corporate networks (Piccardi et al., 2010)
Two mode network



Two mode network



One mode network (communities)

Applications: Communities in Italian corporate networks (Piccardi et al., 2010) Two mode network

- Ownership network: From two mode (shareholders and companies) to one mode network (node = company; link = shareholder)

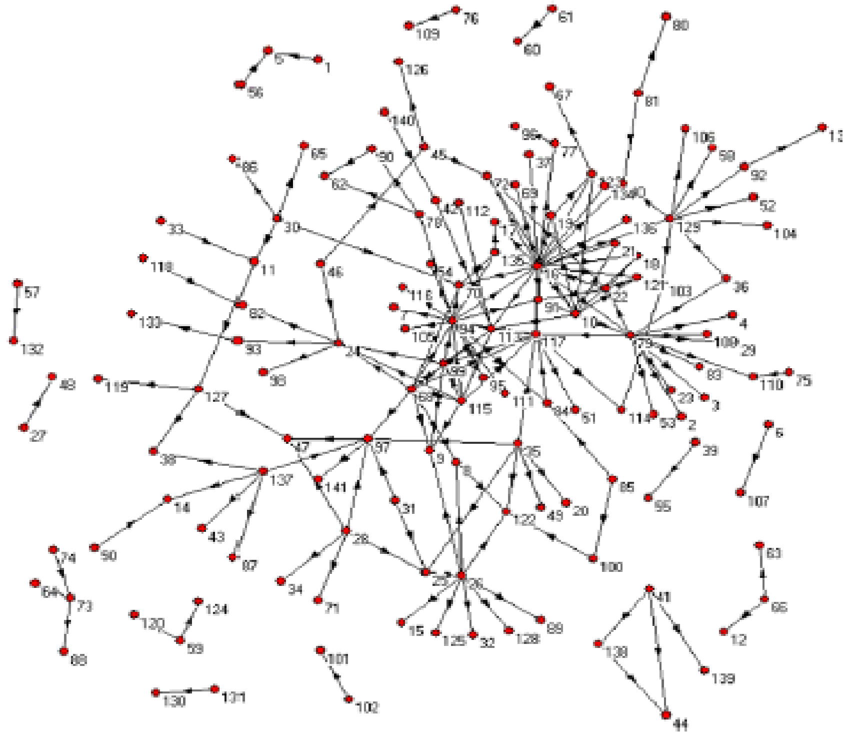


Fig. 5. The ownership network (isolated nodes have been removed).

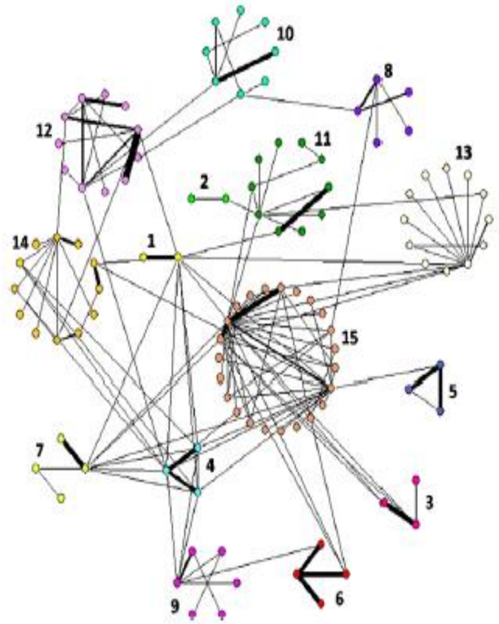


Fig. 7. The community structure of the giant component of the undirected ownership network. The thickness of the links is proportional to the weight.

Applications: Communities in Italian corporate networks (Piccardi et al., 2010)
Two mode network

Both (board and ownership) networks show similarities

Conclusion: existence of business groups (pyramidal structure) sharing common interests, but also existence of softer coalitions (control in decision making, historical dependence)

Move to Pajek!!!