# Analysis of European Projects on Simulation

Vladimir Batagelj

University of Ljubljana

Slovenia

**The Age of Simulation 2006**

Linz, January 12-14, 2006

version: 13. Jan 2006 / 05 : 32

# Outline





# Some technical details

The EU projects datasets gives us an opportunity to apply some of the options implemented in **Pajek** in the last year.

The data were collected by FAS in the form of Excel table stored on the textfile `RuthDELmain.csv`. The rows are the projects and colomns correspond to different properties of projects. I produced from this table three two-mode networks using Jürgen's **Text2Pajek** program:

- `project.net` – idents $\times$ projects $= \mathbf{P}$
- `country.net` – idents $\times$ countries $= \mathbf{C}$
- `institution.net` – idents $\times$ institutions $= \mathbf{U}$

$|\text{idents}| = 8869, \quad |\text{projects}| = 933, \quad |\text{institutions}| = 3438, \quad |\text{countries}| = 60.$

# Some technical details – network multiplication

Since all three networks have the common set (idents) we can derive from them using *network multiplication* several interesting networks:

- `ProjInst.net` – projects $\times$ institutions $\mathbf{W} = \mathbf{P}^T \star \mathbf{U}$
- `Countries.net` – countries $\times$ countries $\mathbf{S} = \mathbf{C}^T \star \mathbf{C}$
- `Institutions.net` – institutions $\times$ institutions $\mathbf{Q} = \mathbf{W}^T \star \mathbf{W}$
- ...

# Some technical details – deleted projects

Some projects (27) from original data set have to be deleted – in the final data set `RuthDELmain` they were marked as deleted. When producing two-mode networks I could first physicaly delete them. Instead of this, I used another approach: I produced the cluster $C_D$ of deleted idents and from it a two-mode matrix $\mathbf{D}$ (idents $\times$ idents; not implemented yet in **`Pajek`**). Matrix $\mathbf{D}$ is a 'diagonal' matrix with value 1 for idents not belonging to $C_D$ and 0 otherwise. Using matrix $\mathbf{D}$ we can determine the network `ProjInst.net` by $\mathbf{W} = \mathbf{P}^T \star \mathbf{D} \star \mathbf{U}$ – the deleted idents don't contribute to the network.

# Analysis of `ProjInst.net`

For identifying important parts of `ProjInst.net` I first computed the 4-rings weights and in the obtained network I determined the line islands

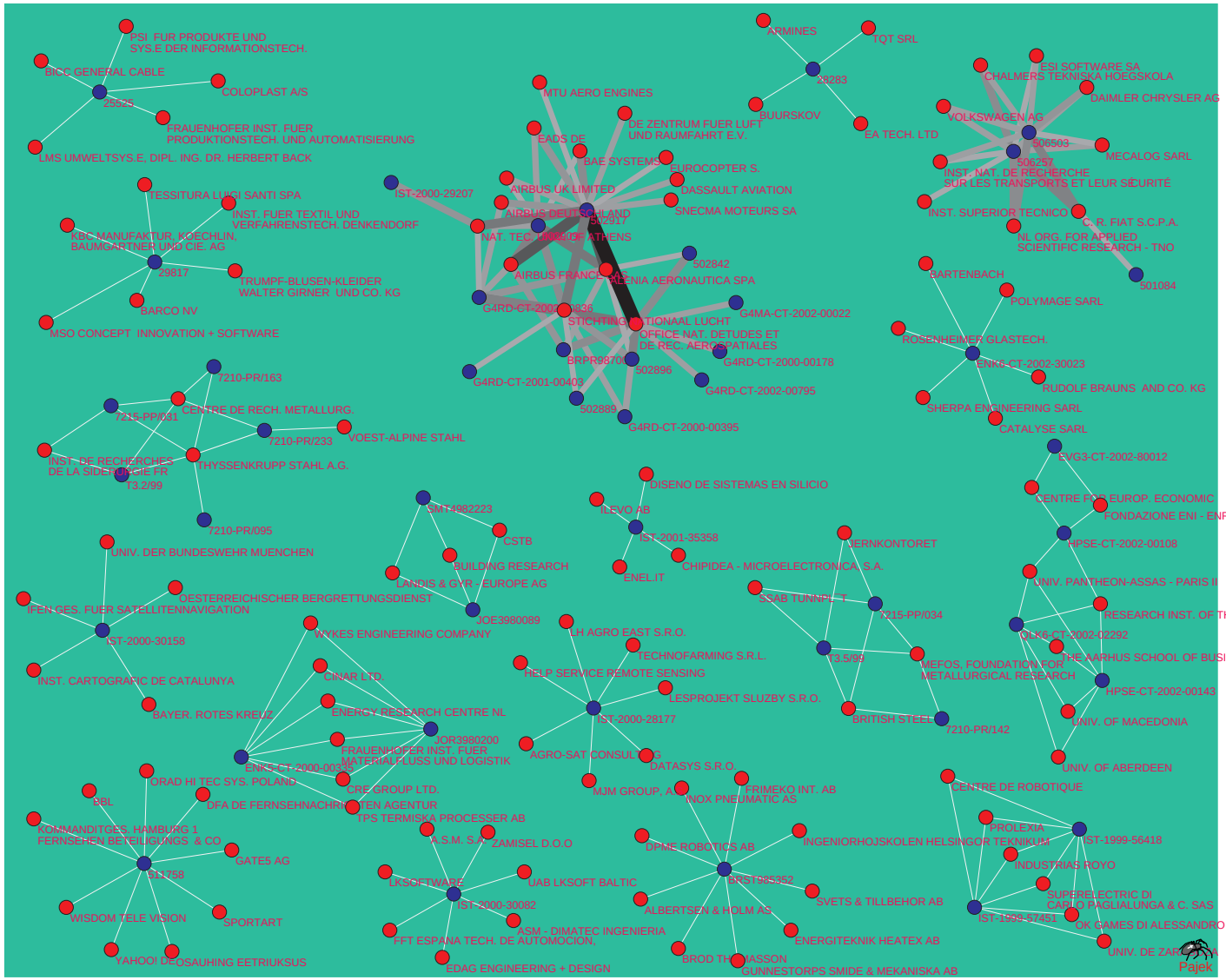
```
Net/Count/4-rings/Undirected
Net/Partitions/Islands/Line Weights[Simple [2,200]
```

We obtain 101 islands. I extracted 18 islands of the size at least 5. There are two most important islands: aviation companies and car companies.

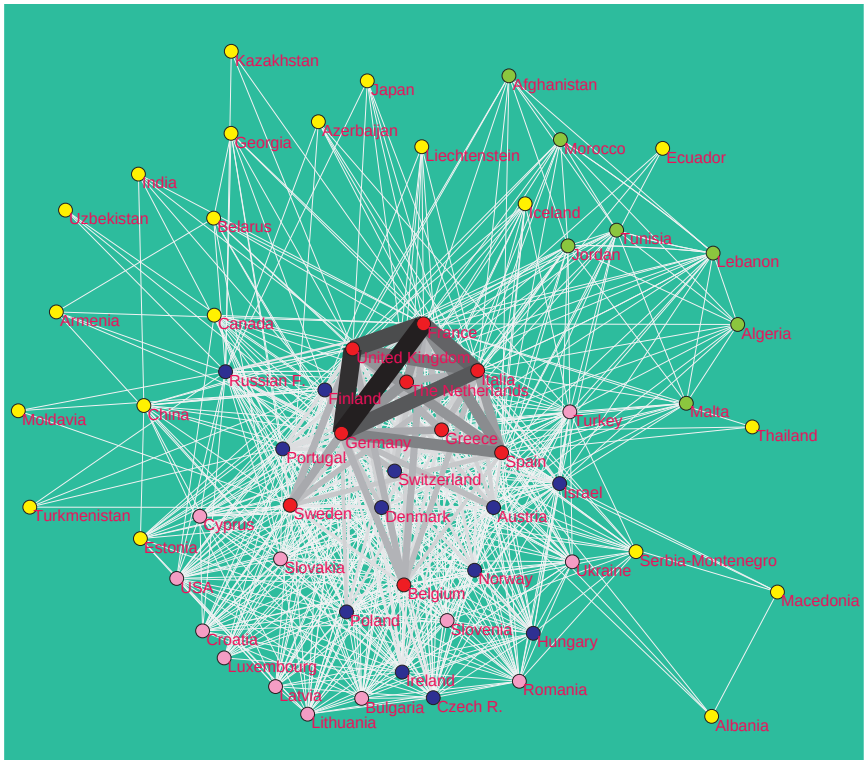
In labels I used a new option `\n`.

For analysis of two-mode networks we can use also $(p,q)-$cores.

# Analysis of ProjInst.net



# Analysis of `Countries.net`




To obtain picture in which the stronger lines cover weaker lines we have to sort them

`Net/Transform/Sort`

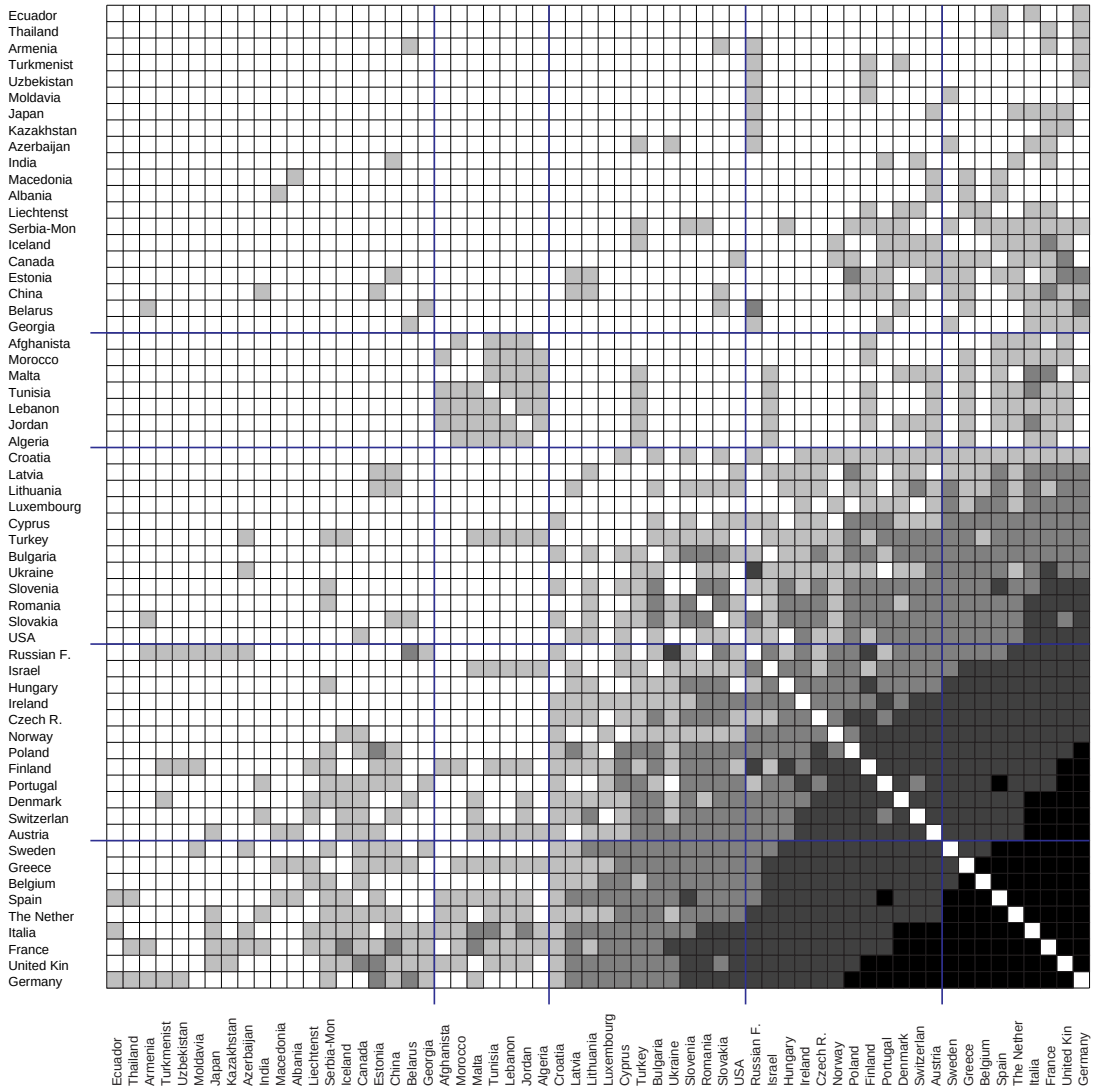
`lines/Line values/Ascending`

For dense (sub)networks we get better visualization by using matrix display. In this case I also recoded values (2,10,50). To determine clusters I used Ward's clustering procedure with dissimilarity measure $d_5$ (corrected Euclidean distance).

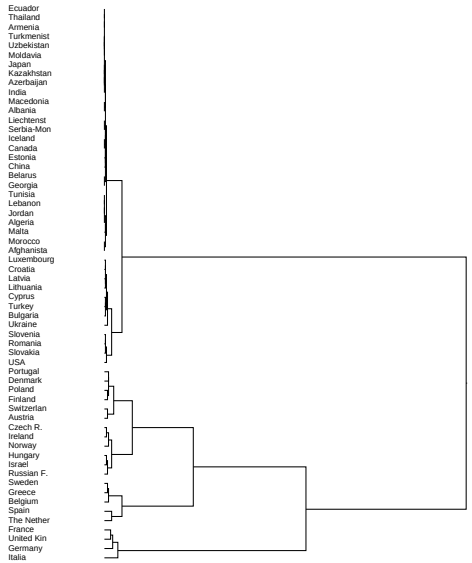
The permutation determined by hierarchy can often be improved by changing the positions of clusters – for the New Year 2006 Andrej added this option in Pajek. We get a typical center-periphery structure.

# Analysis of Countries.net

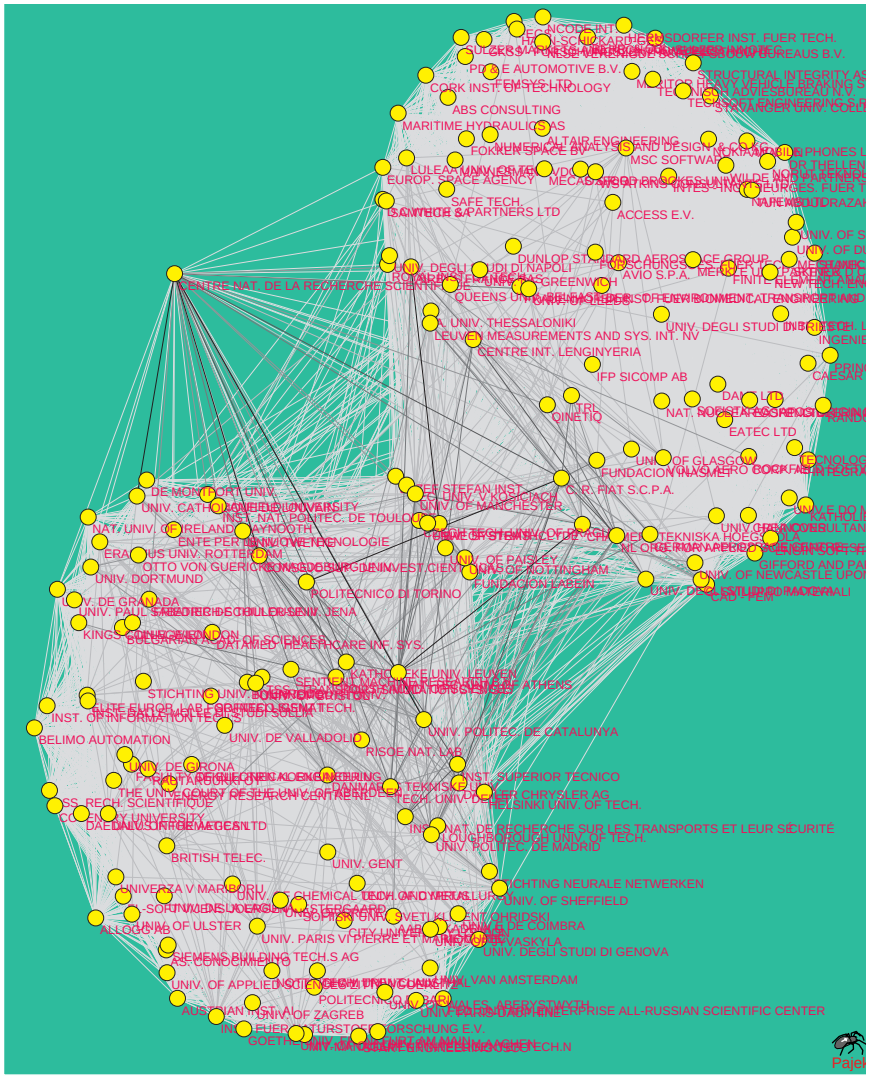
Pajek - shadow [0.00,4.00]




Pajek - Ward [0.00,4785.14]



# Analysis of Institutions.net




To identify the most important institutions I first computed $p_S$-cores vector and use it to determine the corresponding vertex islands. I got essentially one large island. Again the corresponding subnetwork is very dense. I prepared also a matrix display.

# Analysis of Institutions.net