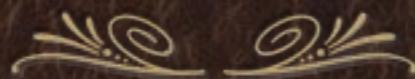
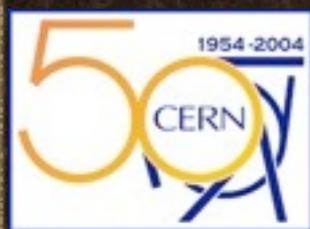




Do Regions of ALICE Matter?



*Federico Carminati (CERN)
ACAT 2011. Brunel. UK*



The Authors

- ❖ E.D.Widmer¹, F.Carminati², C.Grigoras²,
G.Viry¹, G.Galli Carminati³
- ❖ ¹Psychology department. University of Geneva.
Switzerland
- ❖ ²CERN-PH. Geneva. Switzerland
- ❖ ³Psychiatry Department. Adult Psychiatry.
UPDM University of Geneva. Switzerland

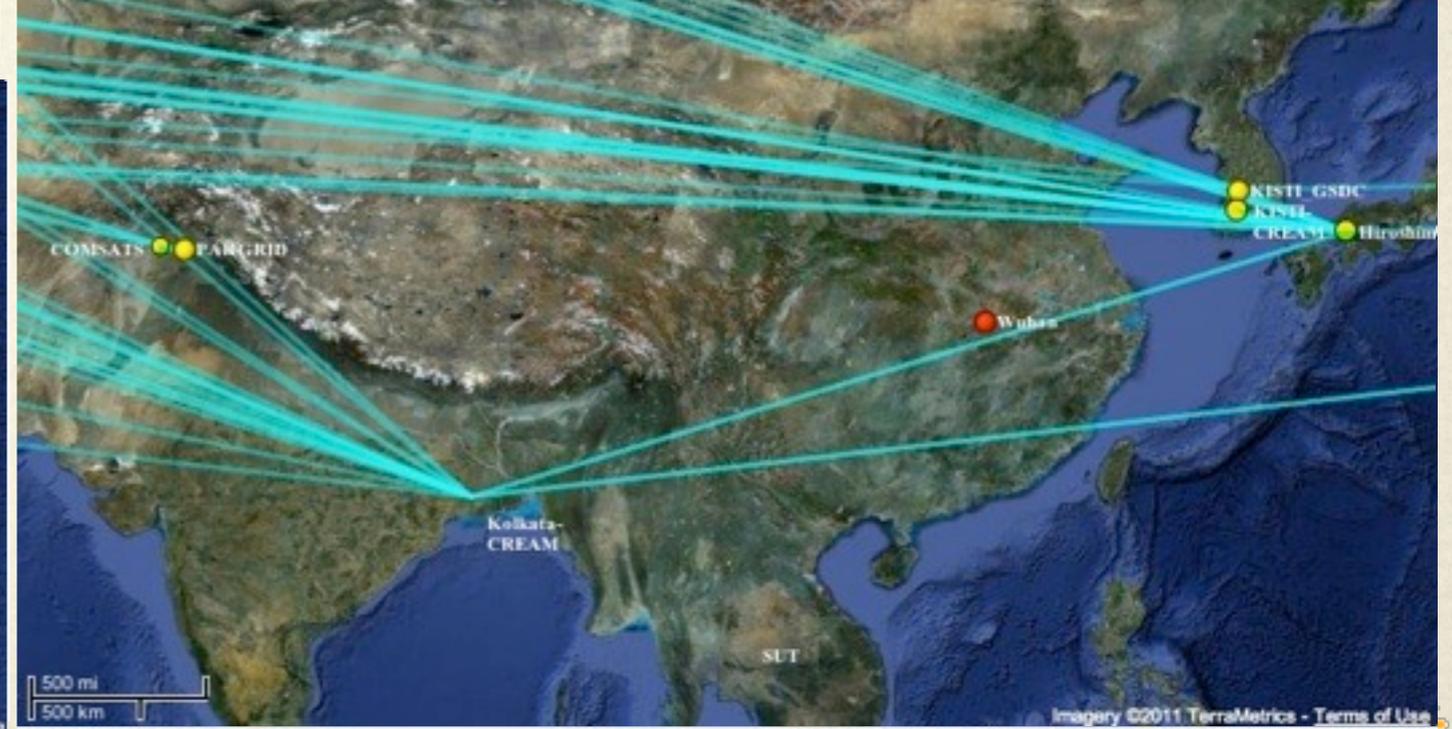
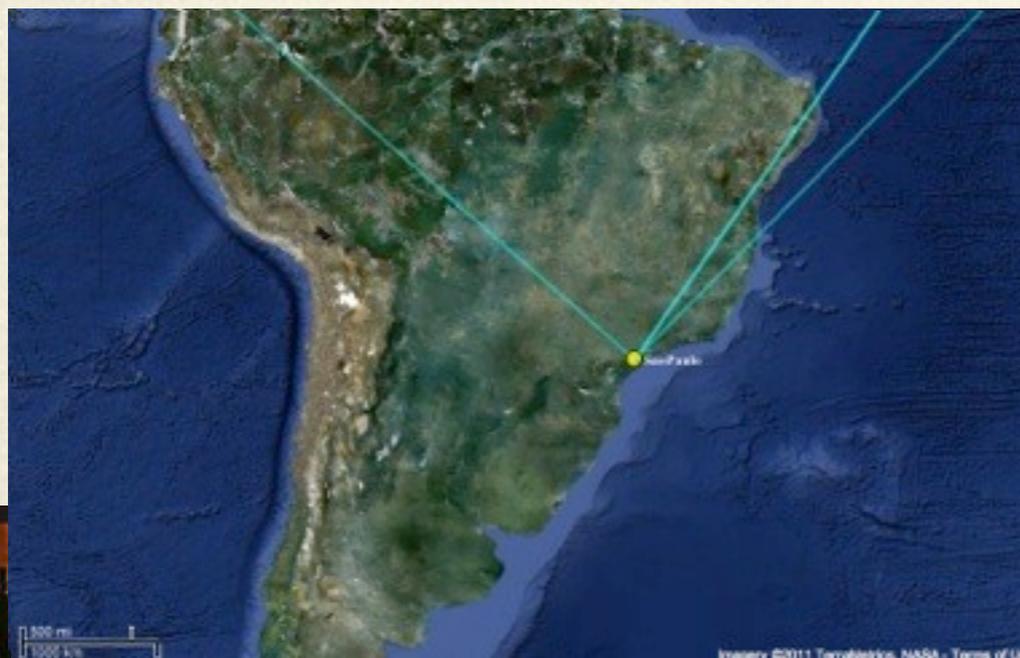
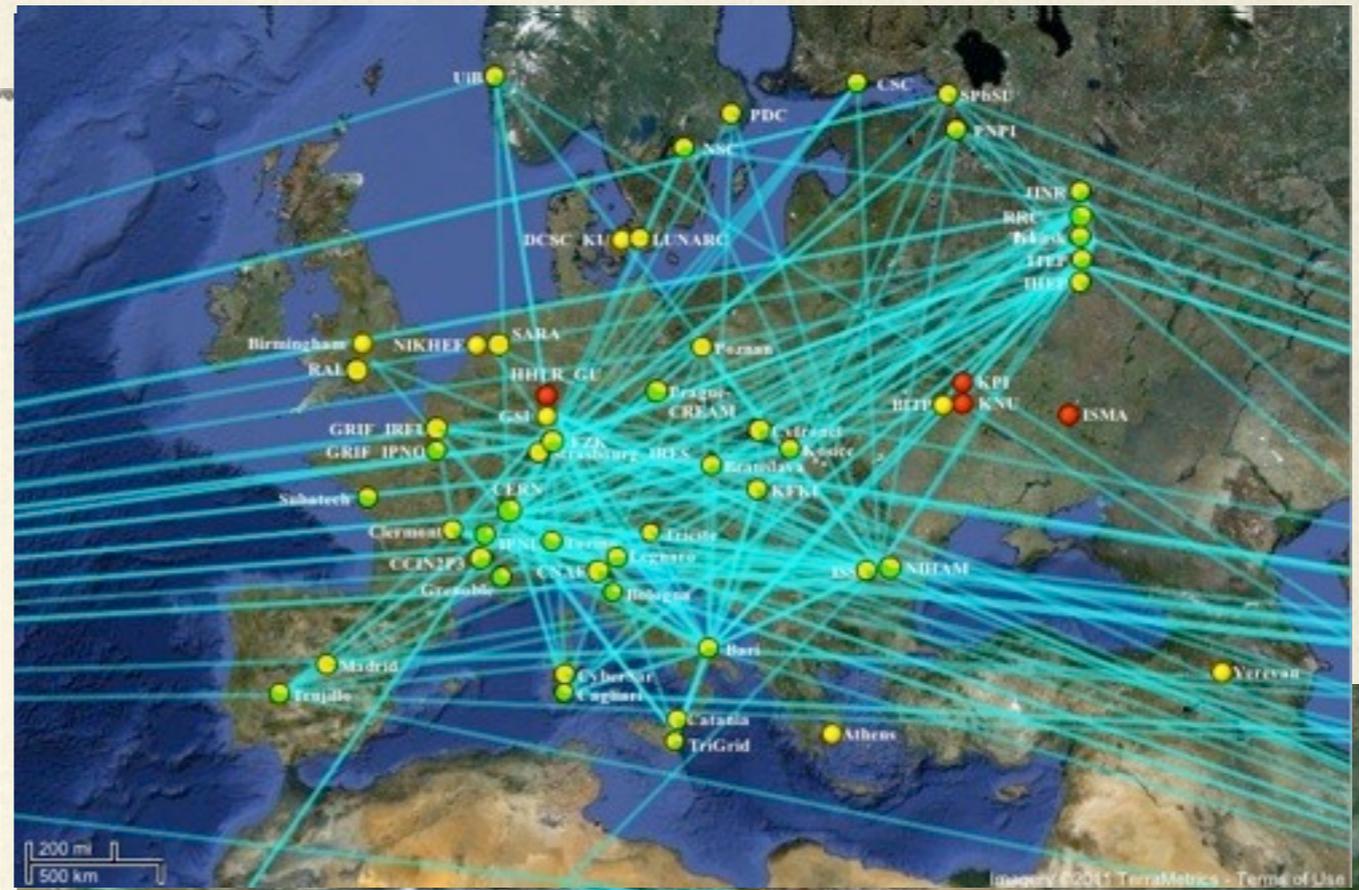


ALICE Computing

- ❖ ALICE computing needs range in the 200 KHEP06/year and 10PB of disk storage
- ❖ These are provided by ~80 computer centres worldwide arranged in Tiers (1 T0, 6 T1s, 73 T2s)
- ❖ A large computational network where data and workload are exchanged, but also a large and complex social network with ~3.000 possible links.
- ❖ 10 years of operation has shown that social relations between the different centres is as important as the material conditions for this complex system extended over different timezones and continents.



The ALICE Grid



This study

- ❖ This study aims at characterising the social network between the ALICE centres
- ❖ Social networks method: Wasserman. S.. and K. Faust (1994). Social Network Analysis: Methods and Applications. Cambridge University Press.
<http://www.insna.org>



Origin of Social Networking

- ❖ Social network analysis aims at recognising and studying the patterns of of people's interaction.
- ❖ The intuitive realisation that these patterns are important elements in the life of the individuals that enact them lead to a systematic approach to theory and research on Social Networks (J.Moreno).
- ❖ After WW2 A.Bavelas establishes the Group Networks Laboratory at M.I.T. to conduct a research guided by formal theory in mathematical terms and based on empirical data.



Theory of Social Networking

- ❖ The networking point of view concentrates on the *relations* between the actors rather than on their *attributes*
- ❖ The objective of the network analysis is to interpret the profile (or structure) of the social links amongst the actors



The *Network* perspective

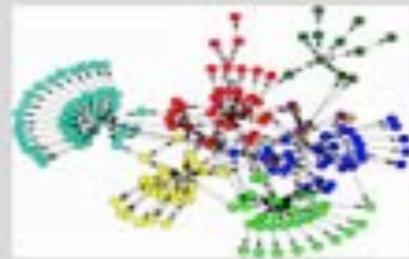
- ❖ Interpret the social links amongst the actors
 - ❖ Network level: e.g. a tightly knit elite and a loosely coupled mass?
 - ❖ Individual level: e.g. are women more marginalised than men?
- ❖ Explore the links between the relational profile and other variables
 - ❖ Network level: e.g. a loosely connected network is a sign of conflict?
 - ❖ Individual level: e.g. a well-connected individual has a higher satisfaction than one who is marginal in the network?



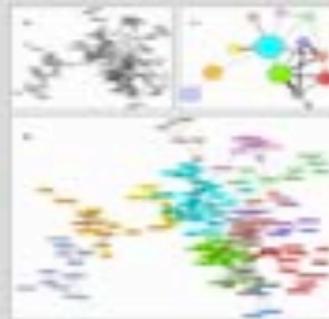
A wide field of application



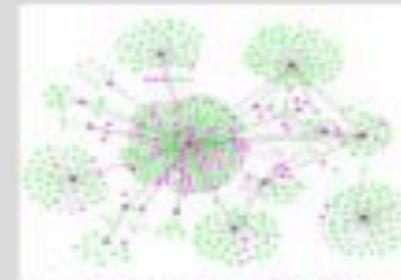
High school dating



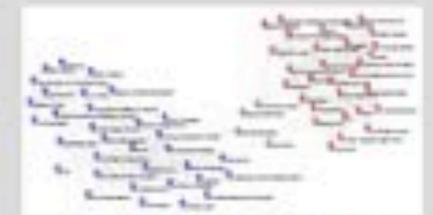
Web site



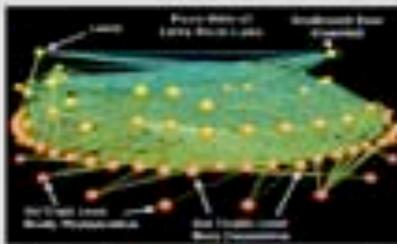
Physicist collaborations



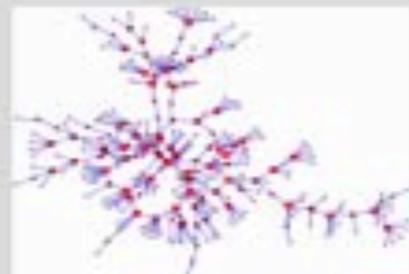
Contagion of TB



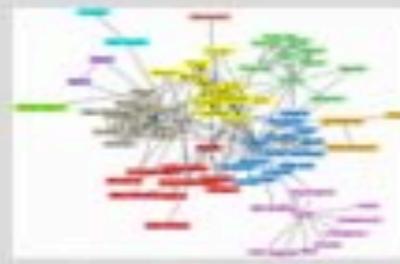
Books on politics



Freshwater food web



Sexual contacts



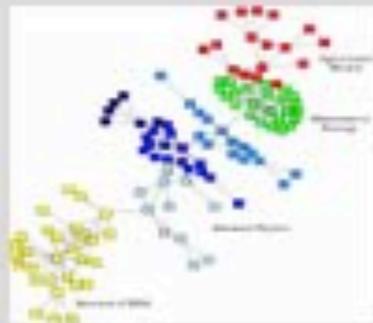
Les Miserables



Network art



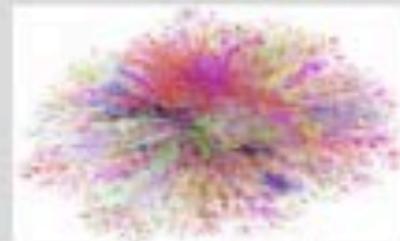
High school friendship



Interdisciplinary collaborations



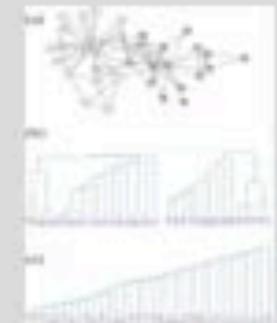
IRC channel



The Internet



Yeast proteins



Karate club



Kind of relations

- ❖ Role based

- ❖ Mather. wife of

- ❖ Boss. teacher of

- ❖ Friend

- ❖ Cognitives perceptional

- ❖ Support

- ❖ Conflict

- ❖ Influence

- ❖ Affiliation

- ❖ Belongs to the same organisation

- ❖ Affective

- ❖ Likes. trusts

- ❖ Interactions

- ❖ Sees regularly. counsels. talks to. quarrels with

- ❖ Has sexual relations with. goes out with

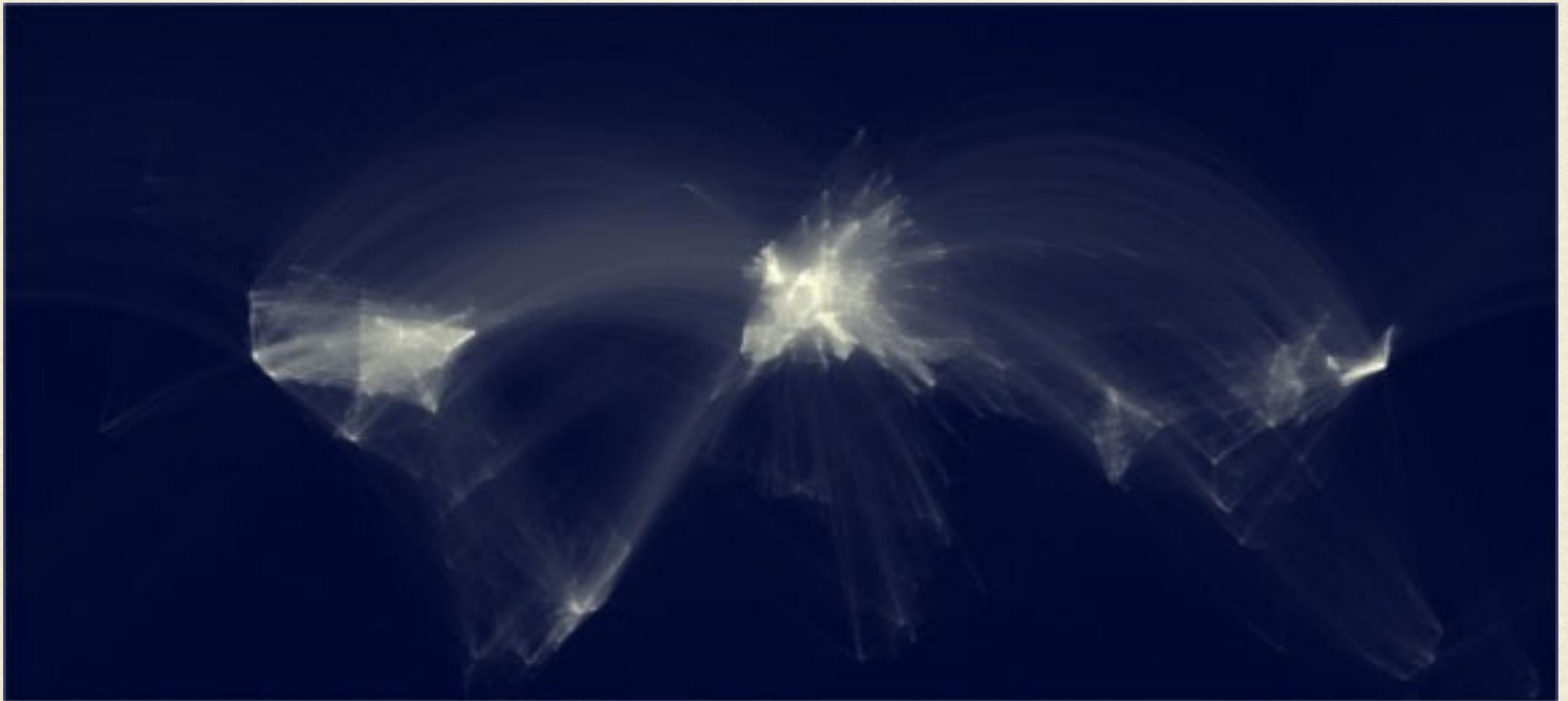


Advantages of the Network perspective

- ❖ Concentrate on observed relations rather than on a-priori categories – non parametric
- ❖ Concentrate on the global structure and not on the single relations
- ❖ No hypotheses on the independence of the observations
- ❖ Relational configurations are complex and opaque. but they obey to rules which can be empirically revealed by the Social Network Analysis



Map of scientific collaborations from 2005 to 2009



- ❖ Computed by O. Beauchesne. Science-Matrix. Inc.
<http://flowingdata.com/2011/01/27/map-of-scientific-collaboration-between-researchers/>



Pitfalls of the Network Analysis

- ❖ Mainly an exploratory analysis. complex inferential statistics
- ❖ Difficulties in the data collection (borders of the network. time for the interviews. confidentiality)
- ❖ High sensitivity to missing data



Network data



Number

Binary

Signed

Quantitative

Non directed

| | | |
|--|--|--|
| | | |
| | | |

Direction

Directed

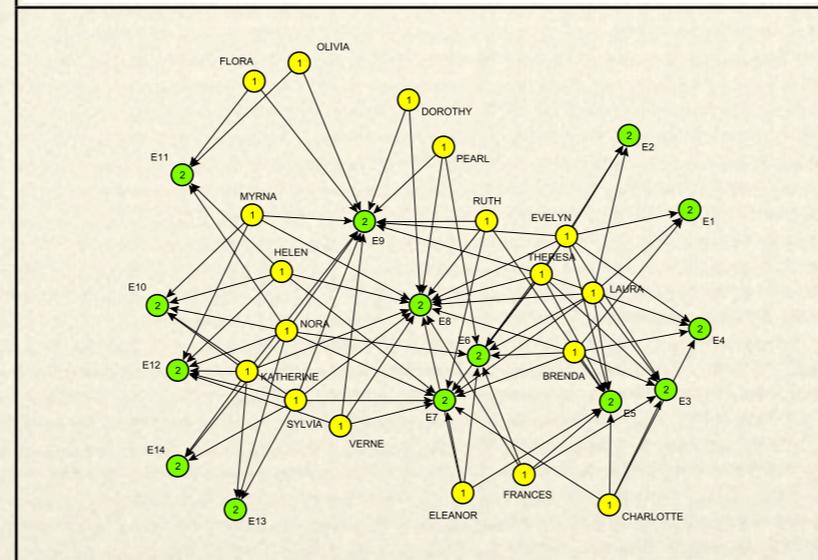
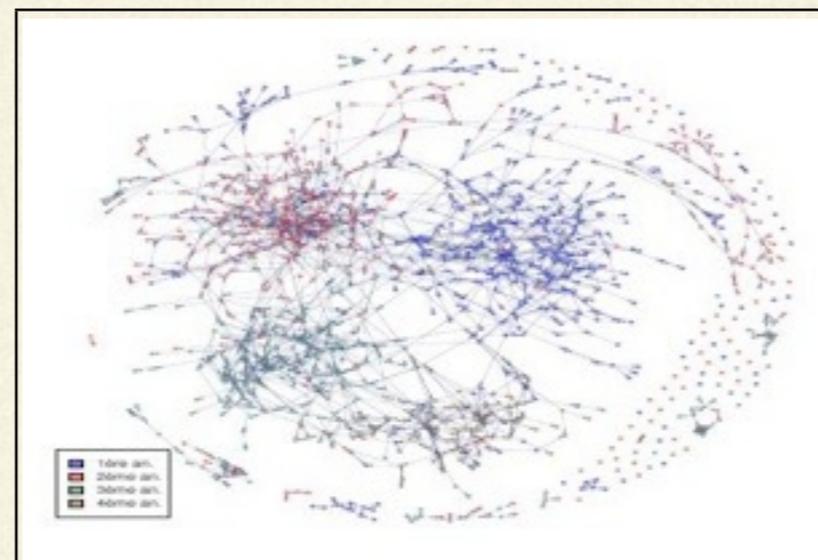


Network data

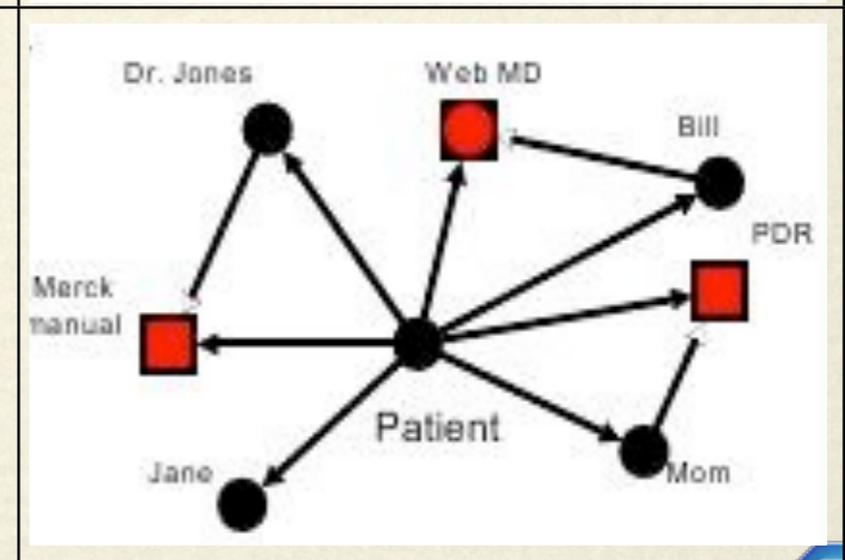
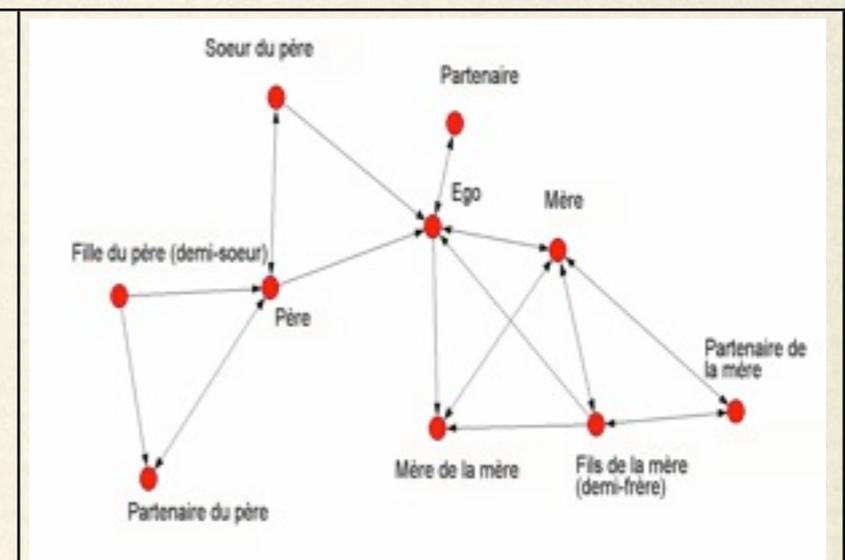
Mono-modal Net

Bi-modal Net
(affiliation)

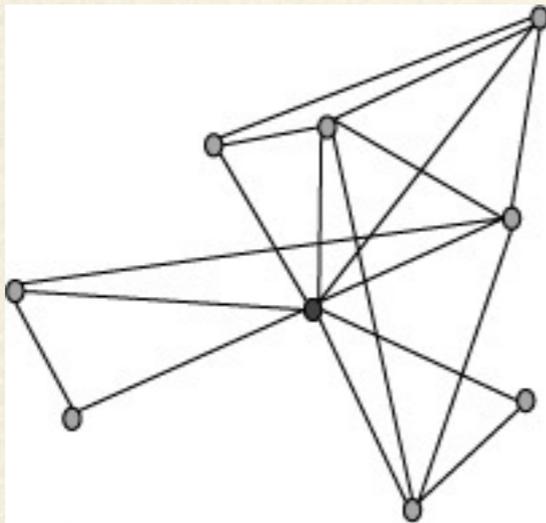
Complete network



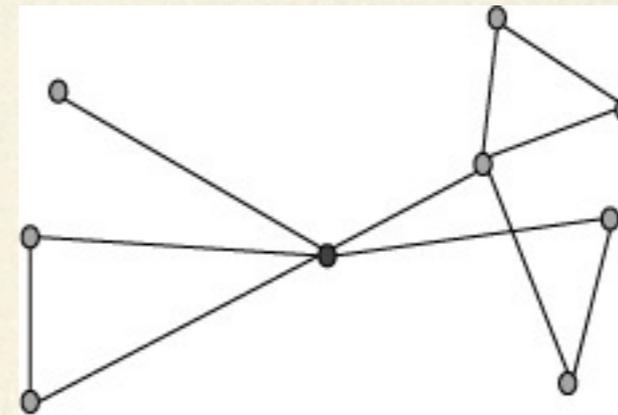
Ego-centered network



Density



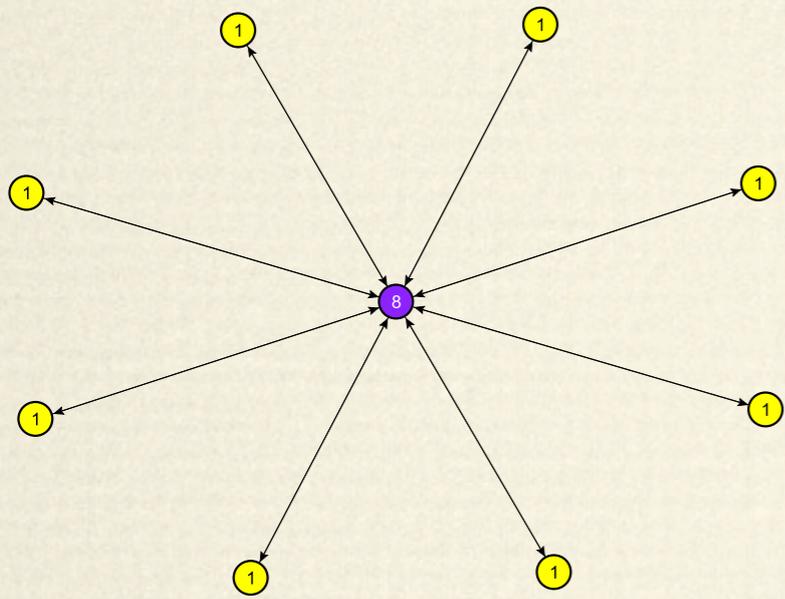
$$\text{density} = 18/36 = 0.5 = 50\%$$



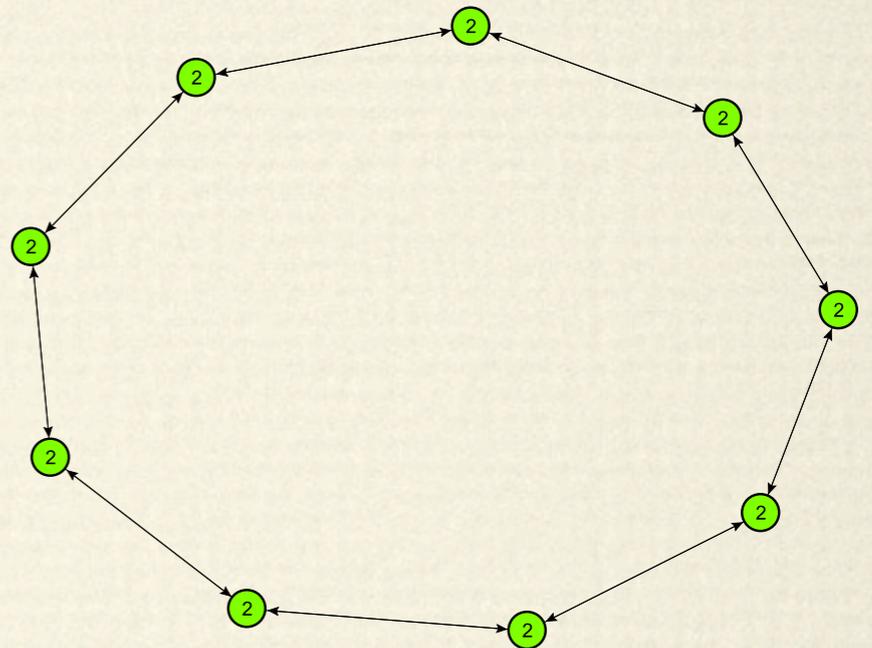
$$\text{density} = 11/36 = 0.3 = 30\%$$



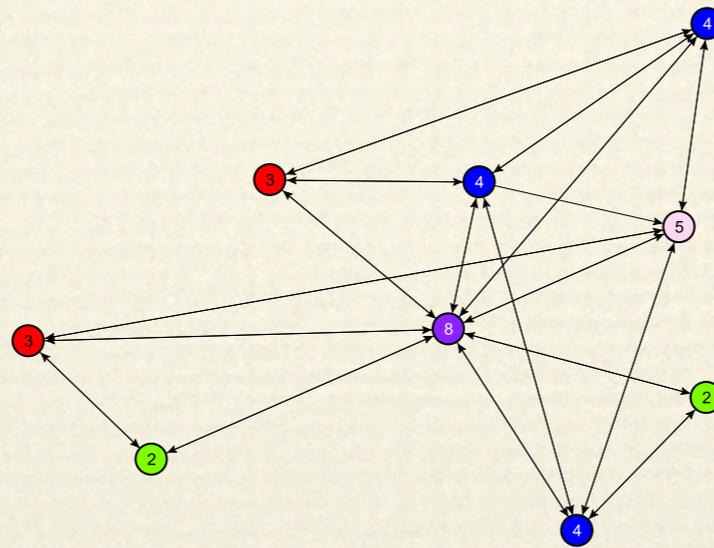
Centrality



100% centrality



0% centrality



64% centrality



Hypothesis test in Network Analysis

Randomly generate thousands of networks keeping the structure of the initial one (size and density)



Calculate the percentage of random networks where the dependent variable considered (average. difference of average. correlation. centrality and so on) is comparable with the one of the real network (test p-value)



If this ratio is small enough (usually 5%). the observed result in the real network is not random but statistically significant



The present study

- ❖ Centre relations are characterised via different metrics:
 - ❖ The theoretical capacity of the network linking them
 - ❖ The actual quantity of data exchanged
 - ❖ Their physical distance
 - ❖ Their geo-political affiliation
 - ❖ The Internet Round Trip Time (RTT)
 - ❖ Their social networking



Our analysis

- ❖ Our study was based on a questionnaire containing four+one questions:
 1. Which centers provide your center with significant help in its work at least once a week
 2. Which centers make it difficult for your center to carry out its job responsibilities
 3. Which centres are in regular e-mail contact with you. at least once a week
 4. With which centers your center would like to have more interaction in its work?
 5. Overall. how do you rate the functioning of your centre?
- ❖ The form was sent to 73 centres and 50 (68%) answered



Questionnaire

*The following question is relative to the way you personally grade the work of your center
(be assured that the answers provided will be kept strictly confidential)*

Overall, how do you rate the functioning of your own center?

| <i>Outstanding</i> | <i>Very high</i> | <i>High</i> | <i>Average</i> | <i>Low</i> | <i>Very low</i> | <i>I do not know</i> |
|--------------------|------------------|-------------|----------------|------------|-----------------|----------------------|
| | | | | | | |

Please respond to these questions with crosses in the corresponding columns of the table below

| | |
|------------|--|
| Question 1 | Which centers provide your center with significant help in its work at least once a week (you can indicate more than one)? |
| Question 2 | Which centers make it difficult for your center to carry out its job responsibilities (you can indicate more than one)? |
| Question 3 | Which centres are in regular e-mail contact with you, at least once a week (you can indicate more than one)? |
| Question 4 | With which centers your center would like to have more interaction in its work (you can indicate more than one)? |

| # | Centre | Question 1 | Question 2 | Question 3 | Question 4 |
|---|---------|------------|------------|------------|------------|
| 1 | Aalborg | | | | |
| 2 | Athens | | | | |
| 3 | Bari | | | | |



Results

.....

| | Help | E-mail contacts | Wished collaborations | Bandwidth | Data transfers | RTT Capacity |
|---|------|-----------------|-----------------------|-----------|----------------|--------------|
| With CERN | | | | | | |
| Density | 4.3 | 4.9 | 3.6 | 16.2 | 8.4 | 18.2 |
| Betweenness centralisation | 36.2 | 31.3 | 26.3 | 6.2 | 8 | 21 |
| Indegree centralisation (provided) | 72.6 | 59.5 | 42.1 | 34 | 12.9 | 39.9 |
| Outdegree centralisation (providing) | 33.1 | 17.9 | 13 | 29.9 | 31.6 | 71.1 |
| Number of cliques (min. size 3. sym. max) | 24 | 19 | 23 | 38 | 24 | 67 |
| Without CERN | | | | | | |
| Density | 2.2 | 3.4 | 2.5 | 14.9 | 7.1 | 16.4 |
| Betweenness centralisation | 0.5 | 3.3 | 2.4 | 6.9 | 6.9 | 22.8 |
| Indegree centralisation (provided) | 12.7 | 11.5 | 6 | 33.9 | 11.8 | 34 |
| Outdegree centralisation (providing) | 10.5 | 9.3 | 12.4 | 29.6 | 26.7 | 72.3 |
| Number of cliques (min. size 3. sym. max) | 10 | 9 | 6 | 38 | 15 | 65 |

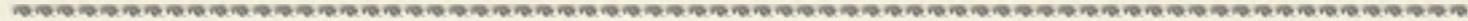


Results

| | Help | E-mail contacts | Wished collaborations | Bandwidth | Data transfers | RTT Capacity |
|-----------------------------|------|-----------------|-----------------------|-----------|----------------|--------------|
| North Europe (with CERN) | 8.1 | 11.1 | 3 | 37.7 | 20.8 | 45.5 |
| North Europe (without CERN) | 3.5 | 8.2 | 1.5 | 35.3 | 16.8 | 41.1 |
| South Europe | 9.7 | 23.6 | 8.3 | 47.3 | 34.9 | 43.9 |
| Asia | 3.3 | 0 | 23.3 | 4.3 | 0 | 11.1 |
| Russia | 23.2 | 26.8 | 16.1 | 17.8 | 10.8 | 52.1 |
| Total (with CERN) | 4.3 | 4.9 | 3.6 | 16.2 | 8.4 | 18.2 |
| Total (without CERN) | 2.2 | 3.4 | 2.5 | 14.9 | 7.1 | 16.4 |



Results



| | Help | E-mail contacts | Wished collaborations | Bandwidth | Data transfers | RTT Capacity |
|--------------|--------|-----------------|-----------------------|-----------|----------------|--------------|
| with CERN | 0.15** | 0.24** | 0.07** | 0.30** | 0.27** | 0.40** |
| without CERN | 0.16** | 0.27** | 0.10** | 0.30** | 0.26** | 0.39** |
| **p<.01 | | | | | | |

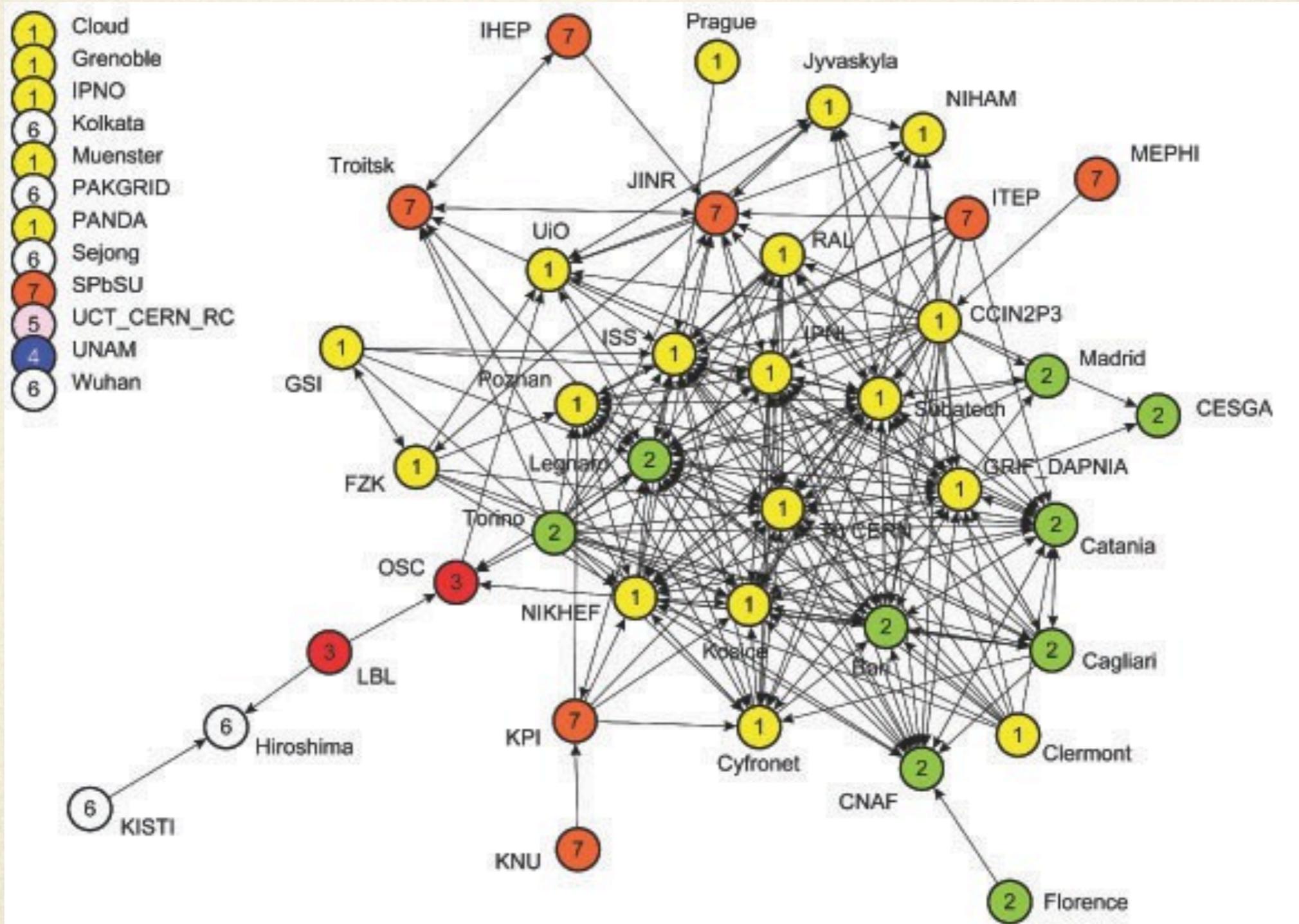


Regions

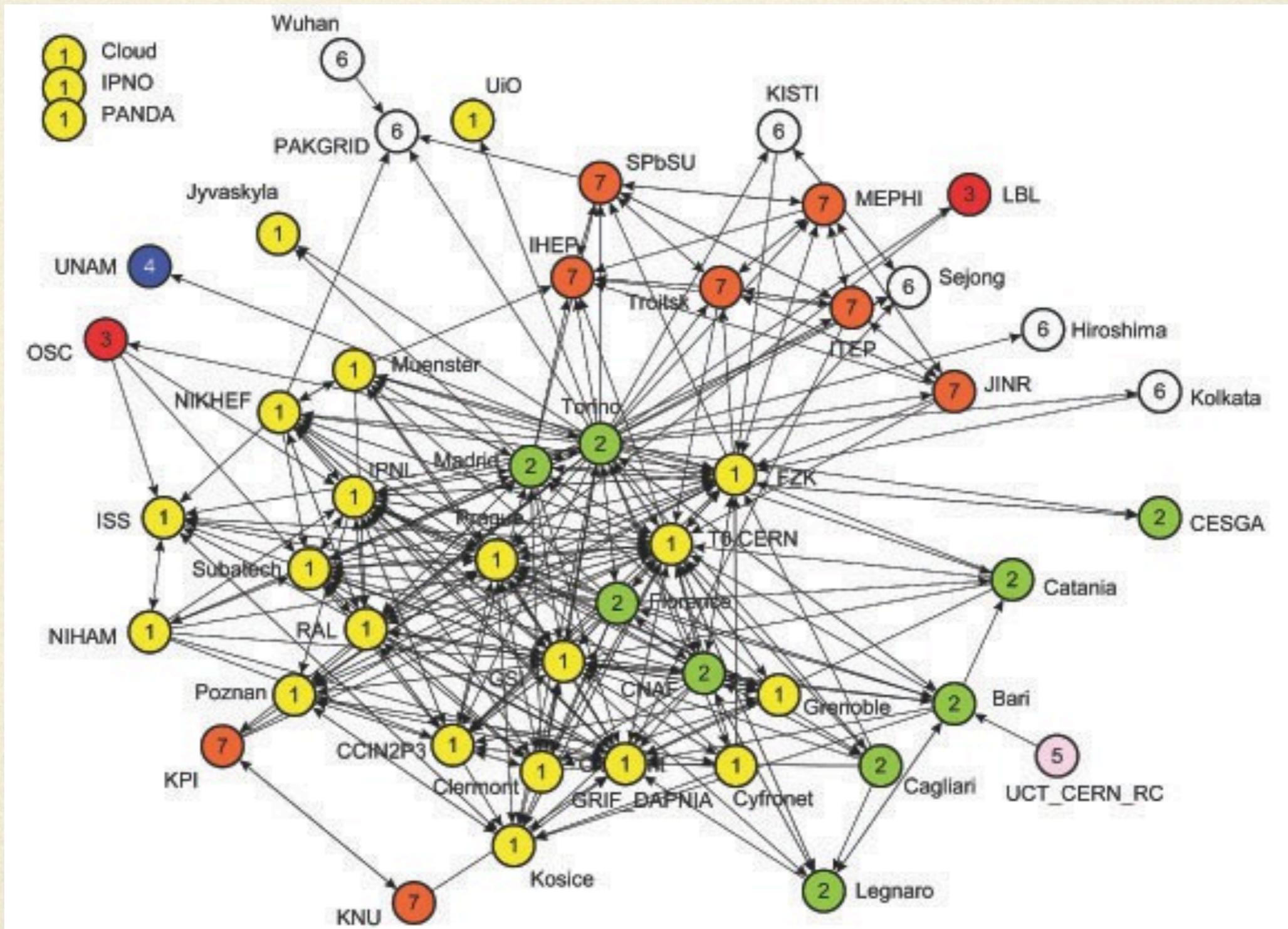
- ❖ Northern Europe (yellow, n=23)
- ❖ Southern Europe (green, n=9)
- ❖ Asia (white, n=6)
- ❖ Russia (orange, n=8)
- ❖ Africa (pink, n=1)
- ❖ South-America (blue, n=1)
- ❖ North America (red, n=2)



Bandwidth by regions



RTT Capacity by regions



Conclusions

- ❖ ALICE Computing Centres are strongly organised in regions
- ❖ This organisation is much stronger than the T1-T2 structure
- ❖ As expected, the network is highly non-random and strongly centered at CERN
- ❖ Collaboration seems to depend also on geographical distance
- ❖ See for instance Asian region
- ❖ Wish for collaboration seems to reinforce the regional character of the Grid

