Aalto-yliopisto
Perustieteiden
korkeakoulu

Temporal Evolution of the ScientificCollaboration Network in Europe

Jori Jämsä

23.3.2015

Instructor: *Ph.D Raj Kumar Pan*
Supervisor: *Prof. Harri Ehtamo*

Systeemianalyysin
laboratorio

Aim of the thesis

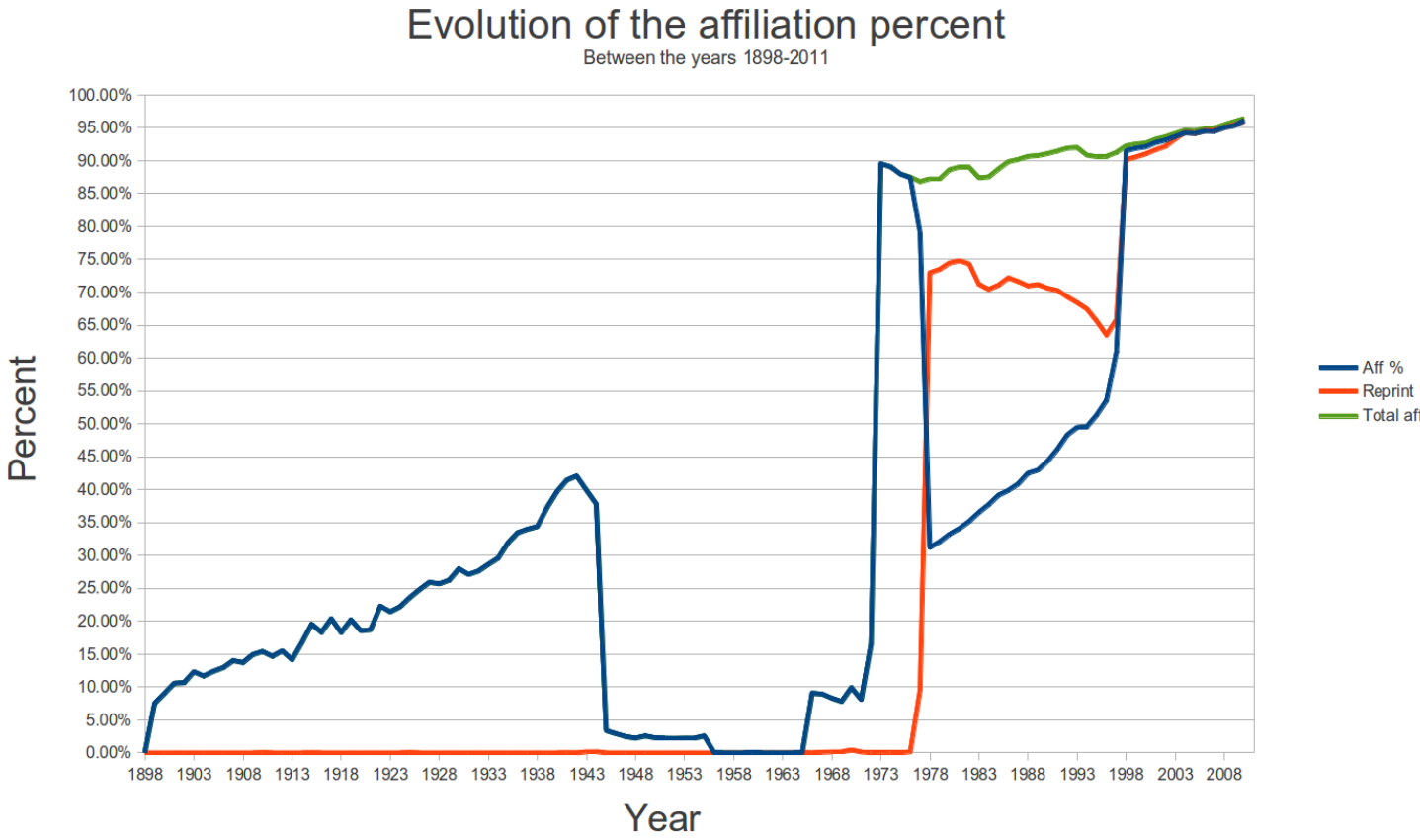
- **How the scientific contribution of different cities have evolved over time?** Contribution is measured by the number of scientific publications.
- **How does the scientific collaboration across different European cities have changed over time?** Collaboration is measure by the number of co-authored papers.

Dataset & Methods

- Dataset:
  - Database of Institute of Scientific Information (ISI) Web of Science
  - Publications from 1900-2010
  - More than 10000 journals and millions of articles
  - For each publication we have over 30 recorded attributes
    - Name of the author, publication year, affiliations, language of the publication...

- Methods:
  - Python (Good language for processing text files)
  - Geocoding (Google API)
  - Network graphs (Gephi)

Outlines



- Complete analysis can be carried out for years 1973-1977 and 1998-2010
- Analyses is focused on Europe

Stages of Work (1/3)

1. Extraction of the needed information for geocoding
   - Most attributes in the original data are excluded and only the useful information for geocoding is extracted
     - Gigabytes of .txt files -> one file with simple lines (e.g LONDON:ENGLAND|555988|LONDON, ENGLAND|ENGLAND)
       - First column is the key that is used to identify unique locations
       - Total number of affiliations is used to determine the importance of the key

Stages of Work (2/3)

2. Geocoding
   - Wikipedia API and Google API were used.

After couple rounds of cleaning we were able to determine coordinates for each of the original keys:

BOSTON:USA:MA|Boston:MA:US|42.358, -71.059

BERLIN:FED REP GER|Berlin:DE|52.519, 13.406

BERLIN:GERMANY|Berlin:DE|52.519, 13.406

Stages of Work (3/3)

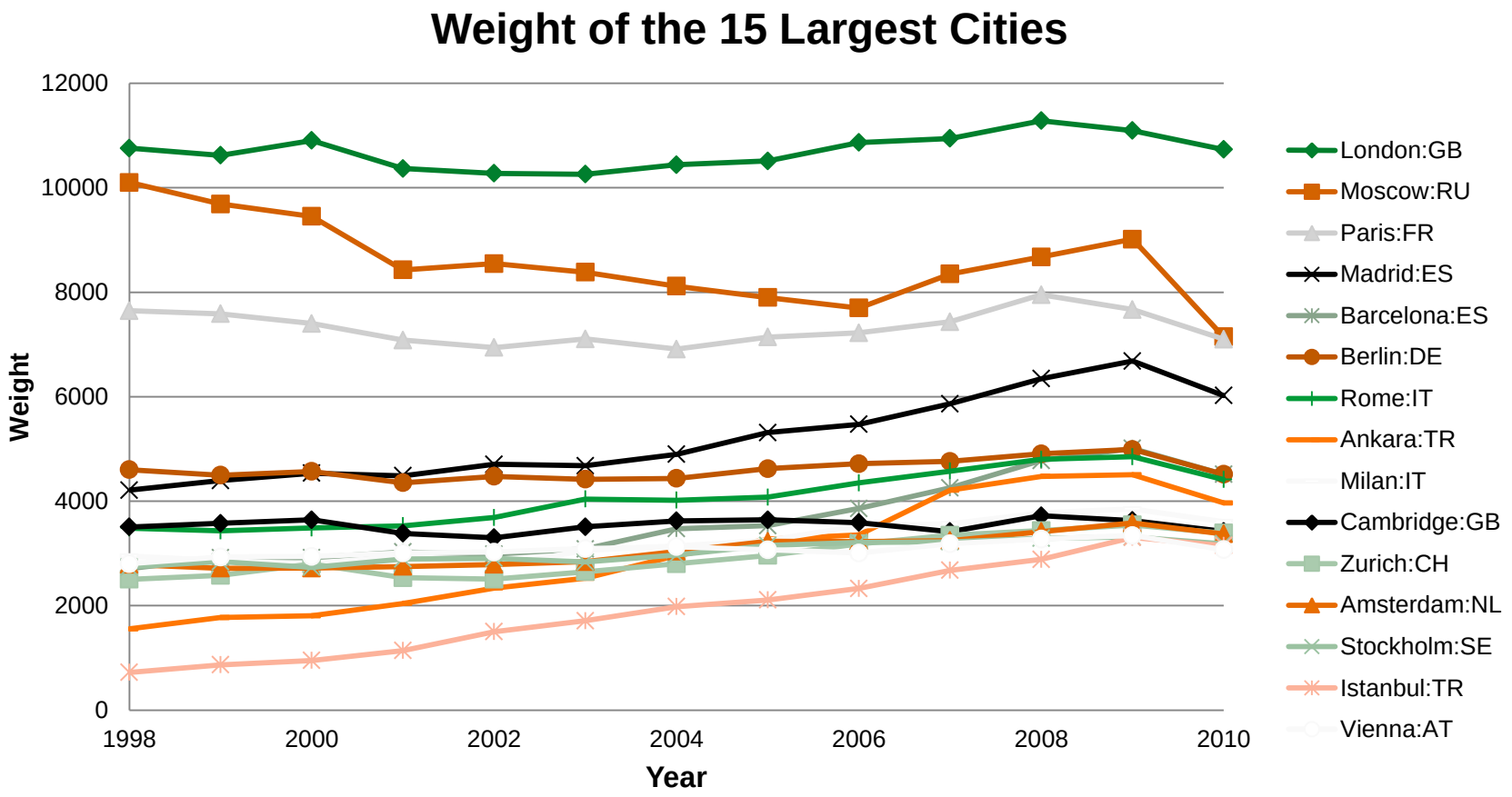
3. Statistical analysis
   - Nodes and edges with relative weights
   - Node represent cities and link represent collaboration

4. Network construction
   - Domestic and international networks

5. Choosing the filters for the network graphs and visualizing the results for each year
   - Modularity analysis

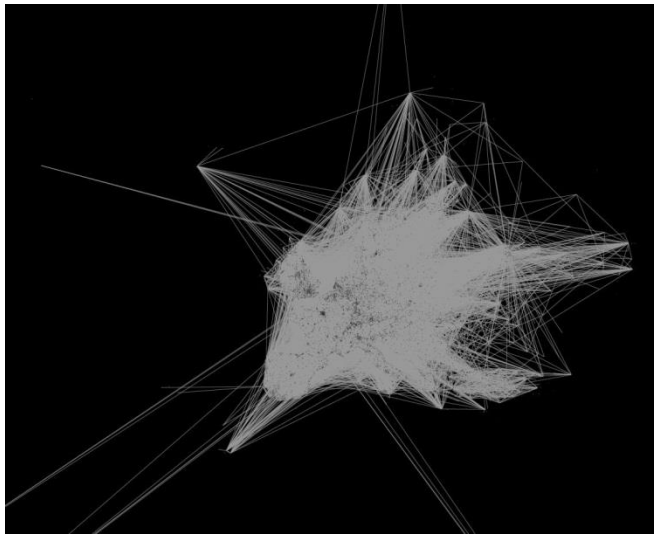
Results

Top 15 cities during the years 1998-2010

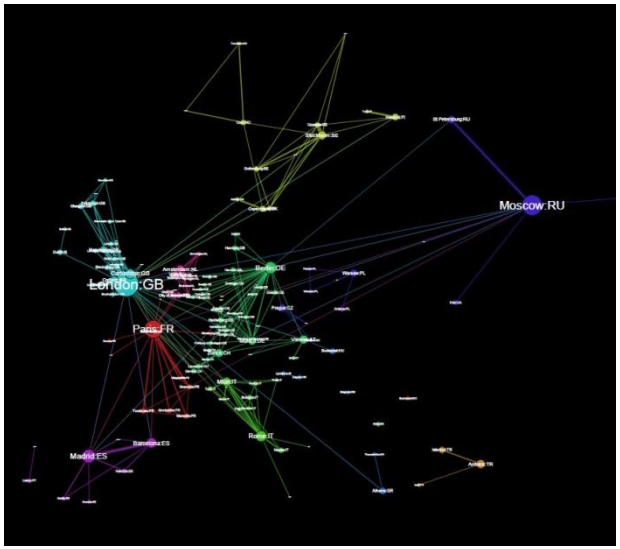


Results

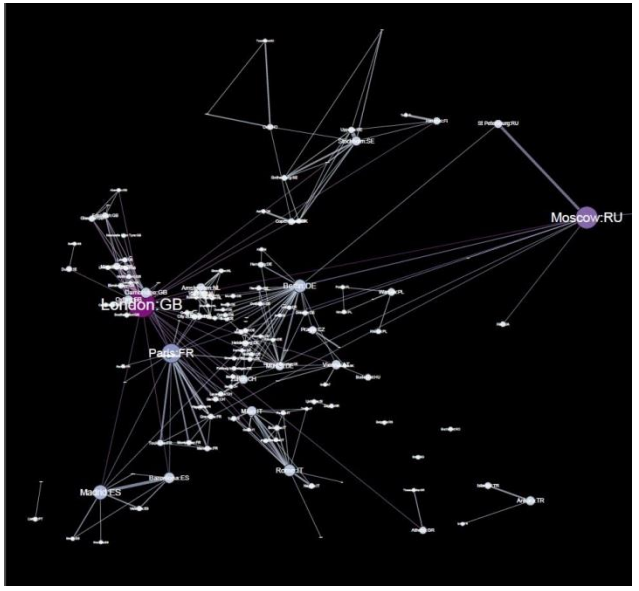
Unfiltered
(years 1998-2010 combined dataset)



Modularity (Louvain method)

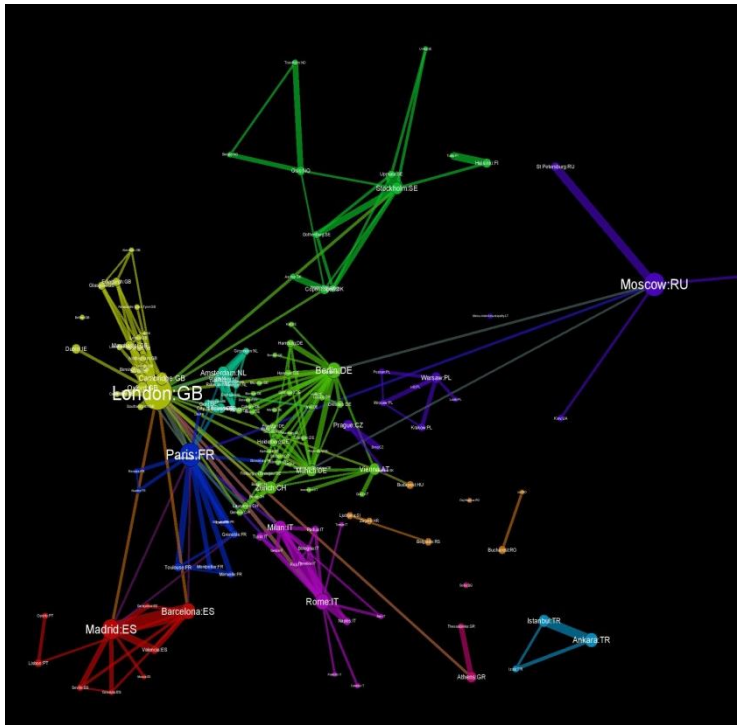


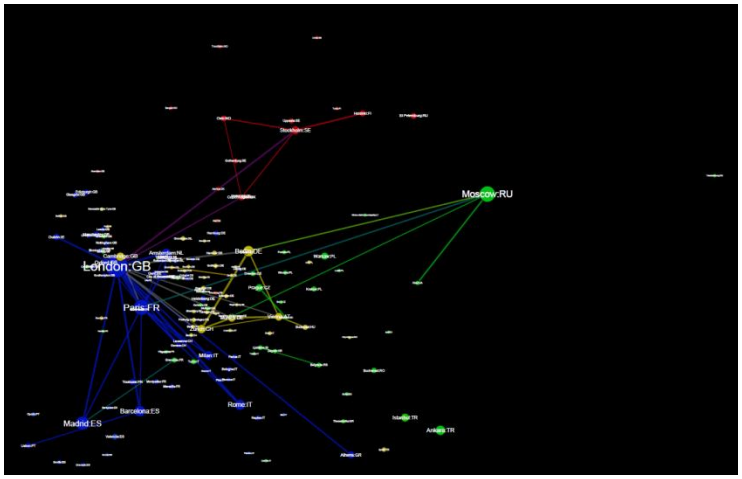
Relative size and colors

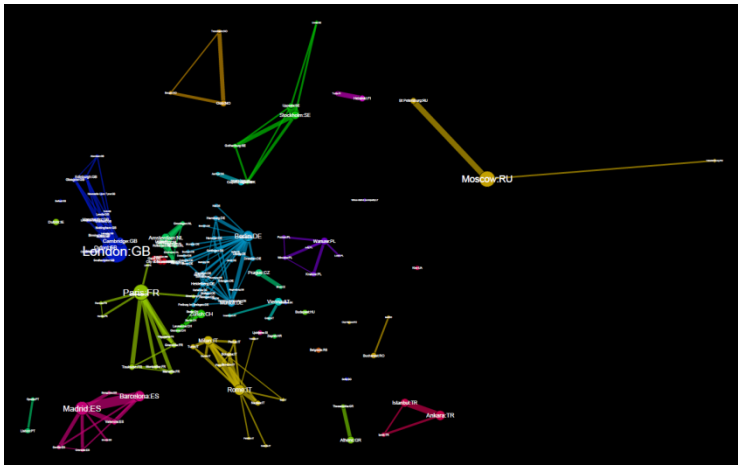


Results

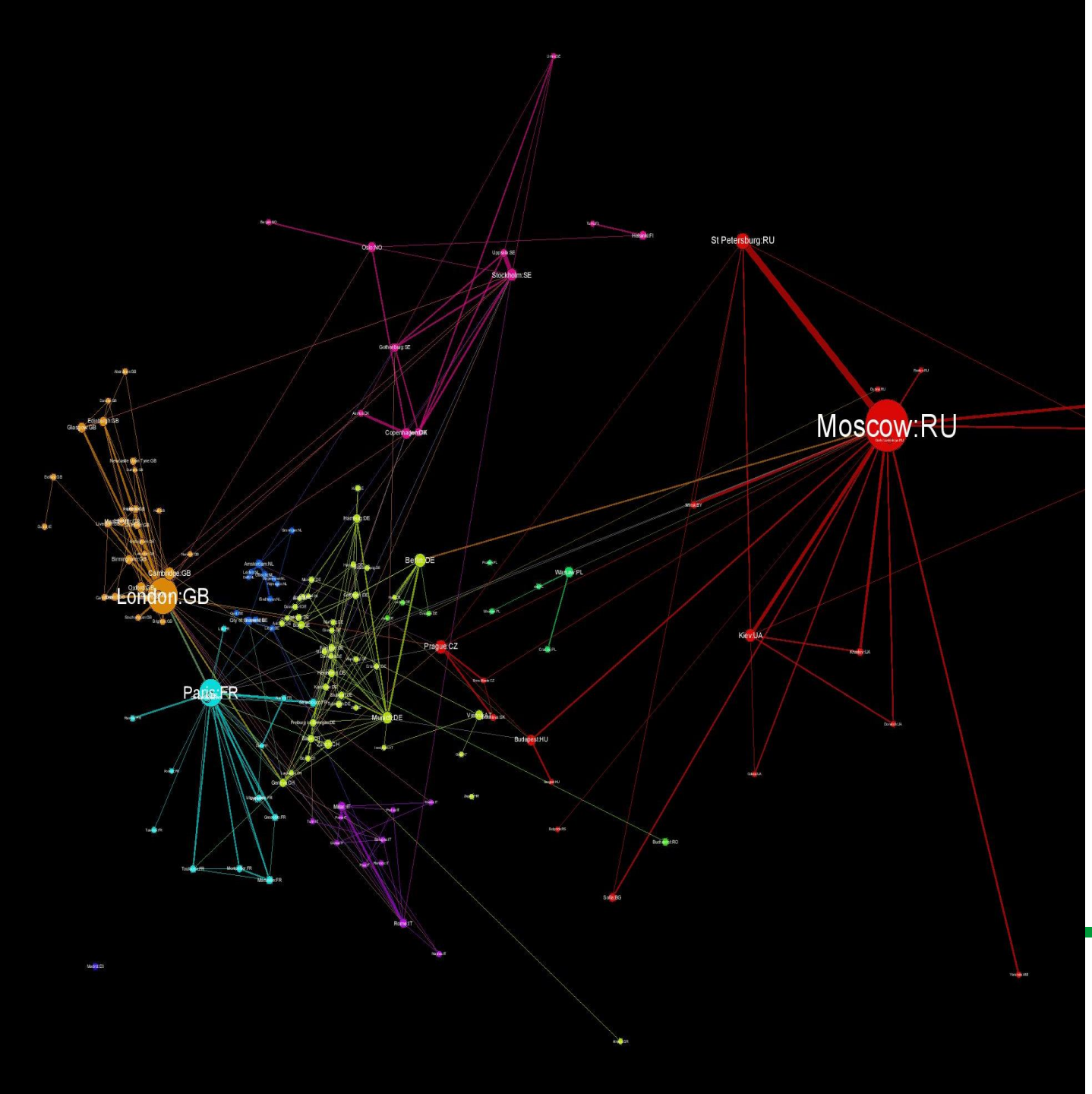
International







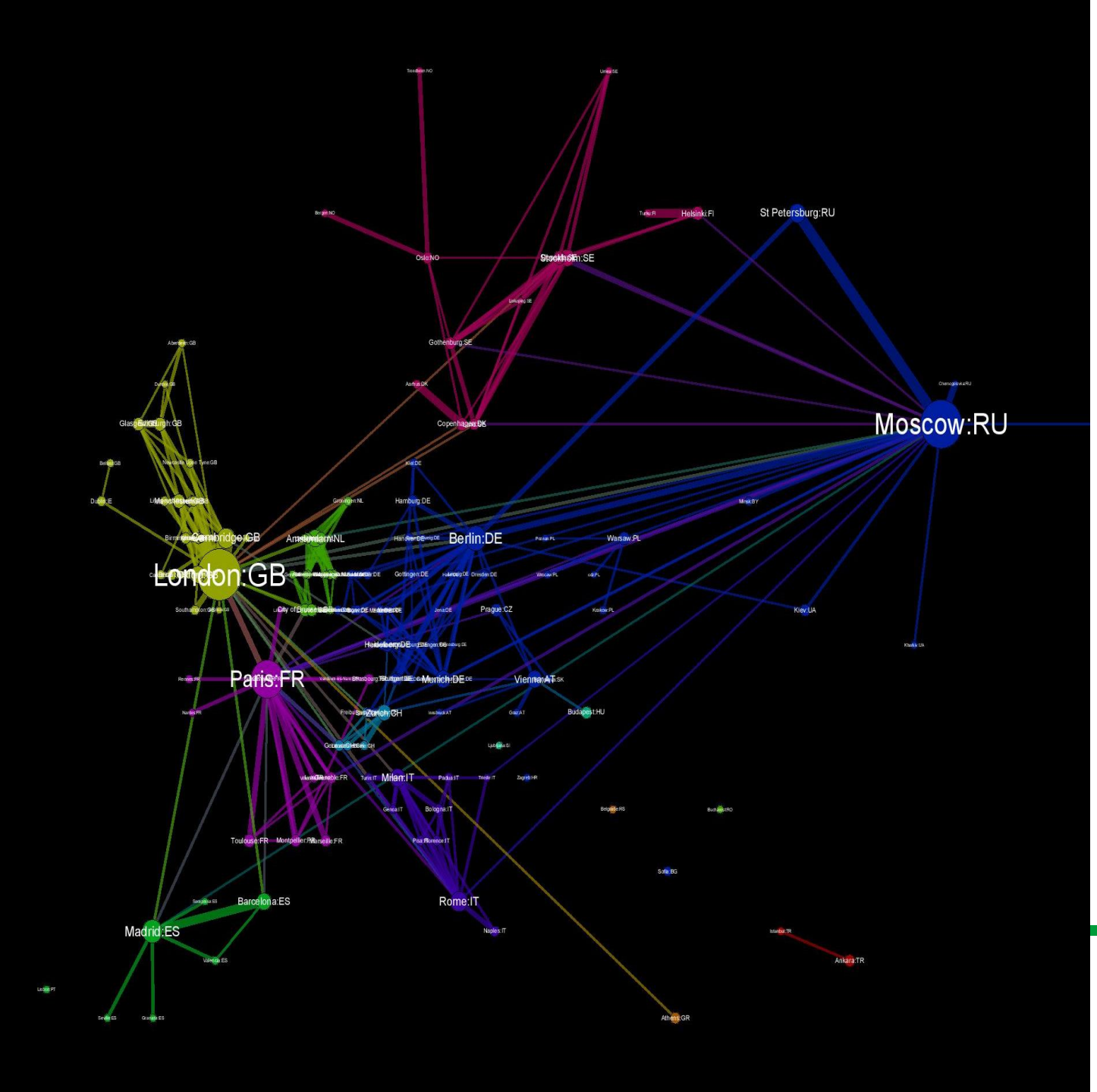
Domestic



Results

EU years 1973
150 biggest cities,
500 strongest links

Systeemianalyysin laboratorio



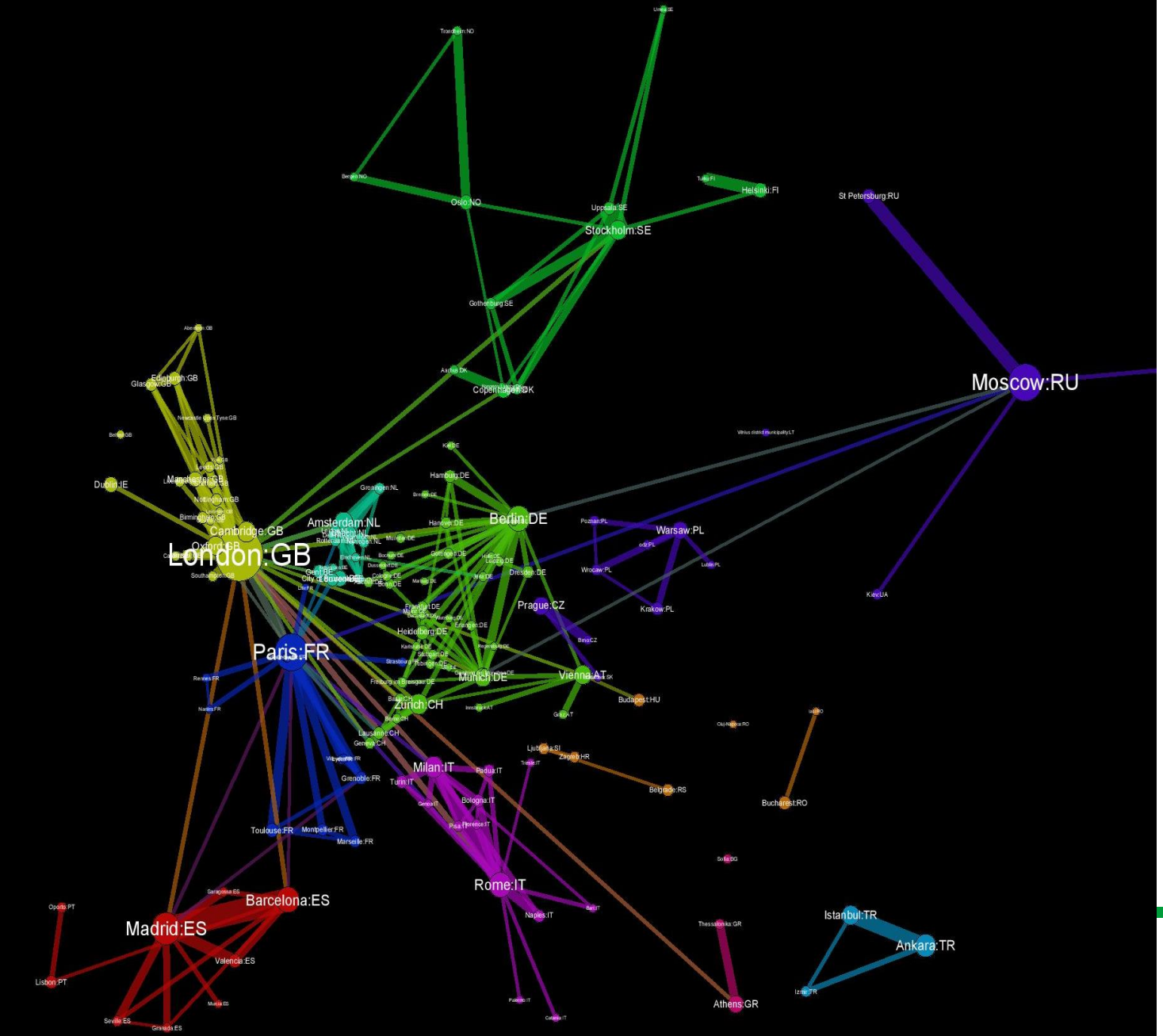
Results

EU year 1998
150 biggest cities,
500 strongest links

Systeemianalyysin laboratorio

Results

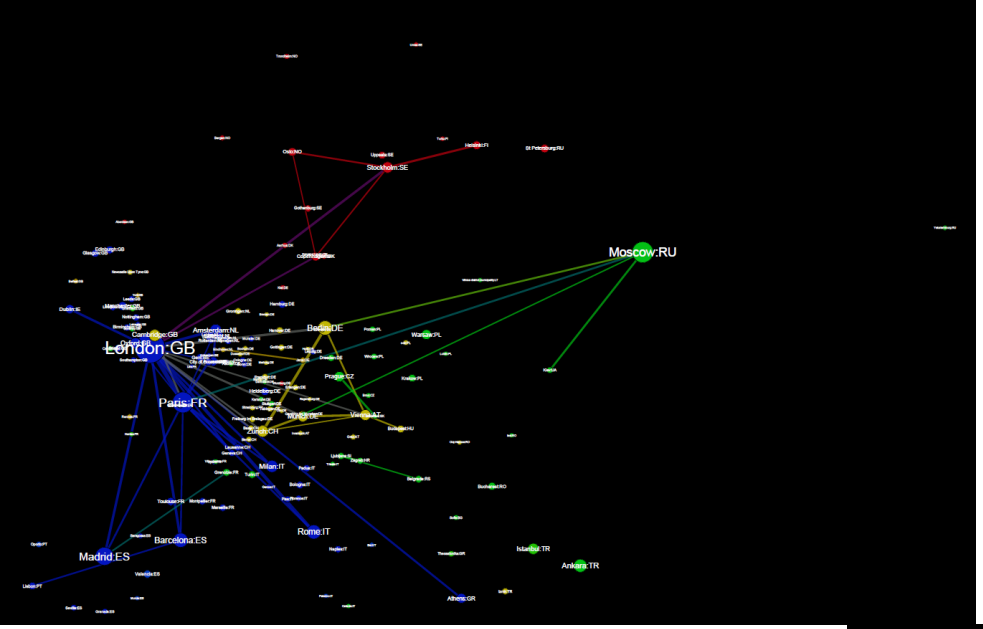
EU year 2010
150 biggest cities,
500 strongest links

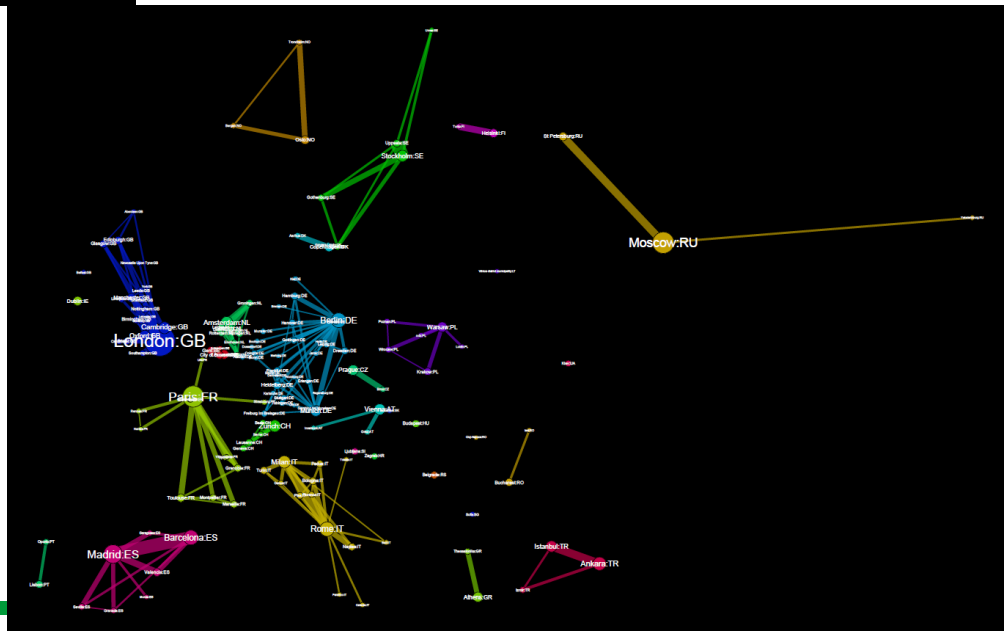


Systeemianalyysin laboratorio

Results

EU year 2010
150 biggest cities,
500 strongest links
(international and domestic networks)





Results

Cities:

- London, Moscow and Paris are dominating the scene
- Influence of Southern Europe is in the ascendant
- Russia cities' influence is decreasing

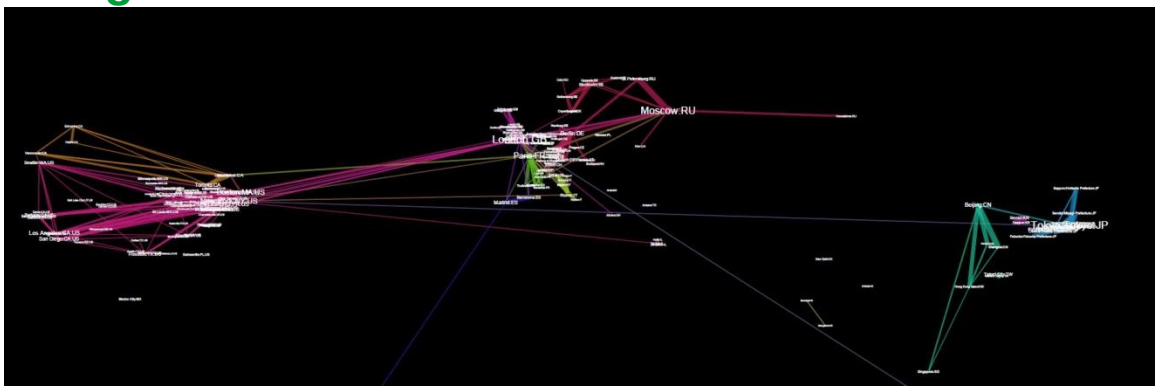
Communities & network:

- E.g. England, France, Spain, Italy, Belgium&Netherlands, Scandinavia, Turkey, Russia&Poland, Germany&Austria&Swizerland
- Collaboration outside countries and language groups relatively weak. Strongest connections inside countries.
- Regional connections are growing even stronger
- However, the amount of connections between countries is increasing (although the growth on domestic collaboration is more rapid)

Future Studies

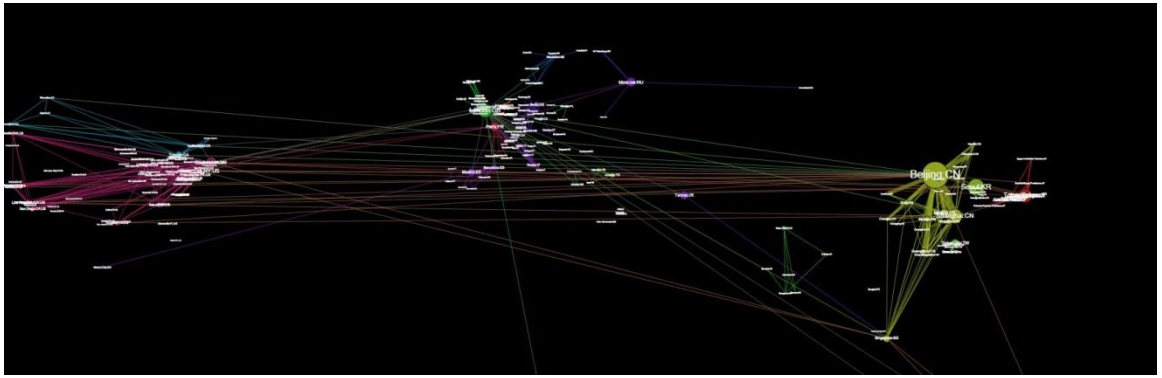
World years 1998, 2010

Cities with over 1000 relative weight and links with over 20 or 30 relative weight



1998

Stronger collaboration between continents can be seen in 2010!



2010

-Rise of China and Bejing (also Japan and South Korea)
-Huge collaboration between USA and China
- Country and language barriers still remain
-Brazil and India getting started

Summary

- Regional collaboration going strong, language and cultural barriers are still in place
- Regional links are getting even stronger
- Europe is becoming a bit more international in the field of science, but the phase is slow (at least in Europe)
- Southern Europe has grown it's influence during the years
- London, Moscow and Paris (+Madrid) are dominating in terms of publications
- One reason behind the increasing international collaboration is the rise of ICT technologies. However, the regional links are also getting stronger, reflecting the importance of language, culture and history of the region.