Jori Jämsä

Temporal Evolution of the Scientific Collaboration Network in Europe

School of Science

Bachelor's thesis

Espoo June 1, 2015

Thesis supervisor:

Prof. Harri Ehtamo

Thesis instructor:

PhD Raj Kumar Pan

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
BACHELOR'S THESIS

Author: Jori Jämsä

Title: Temporal Evolution of the Scientific Collaboration Network in Europe

Date: June 1, 2015 Language: English Number of pages: 3+43

Degree programme: Engineering Physics and Mathematics

Supervisor: Prof. Harri Ehtamo

Instructor: PhD Raj Kumar Pan

Studying science from a scientific point of view has been one of the rising fields in the last decades. It will not only improve our understanding of how scientific progress occurs and how scientific ideas diffuse across different geographical areas, but also the knowledge of science will help us improve policy makers' R&D decisions. In this study, we analyse a publication set of approximately 37 million articles from the perspective of network science to study the temporal evolution of the scientific collaboration network. We focus on articles by authors in European institutes during 1973-1977 and 1998-2010. By utilising geocoding, we obtain the information necessary for the visualisation of the collaboration network at the city level. The data set is analysed by using basic statistical measures. In addition, a community detection method is applied to reveal the evolution of the collaboration clusters. Finally, we visualise the network and communities using a network graph software.

The overall evolution is evident. The total number of publications published in Europe has increased significantly. The dominating cities in terms of affiliations are Moscow, Paris and London. In the latter period under study, the cities in southern Europe and Turkey have increased their share of the total publications rapidly, whereas the influence of Russian cities has been decreasing. In addition, the collaboration between cities has grown stronger. When observing the average and the median strength of the collaboration between cities, we can detect that the domestic collaboration is growing at a faster rate and is stronger overall in comparison to the international collaboration. The community analysis reveals that there are currently twelve collaboration clusters in Europe and that language families and established national collaboration relationships seem to serve as the basis of these cliques.

Keywords: Network theory, science of science, geography, geocoding, internationalisation, temporal evolution of scientific collaboration, collaboration cliques in Europe

AALTO-YLIOPISTO
PERUSTIETEIDEN KORKEAKOULU

KANDIDAATINTYÖN
TIIVISTELMÄ

Tekijä: Jori Jämsä

Työn nimi: Tieteellisen tutkimuksen yhteistyöverkoston kehitys Euroopassa

Päivämäärä: June 1, 2015 Kieli: Englanti Sivumäärä: 3+43

Tutkinto-ohjelma: Teknillinen fysiikka ja matematiikka

Vastuuopettaja: Prof. Harri Ehtamo

Ohjaaja: TkT Raj Kumar Pan

Tieteen tutkimus tuottaa tietoa poliittisen päätöksenteon tueksi sekä parantaa käsitystämme tieteestä itsestään. Tässä tutkimuksessa analysoimme verkkoteoriaa hyödyntäen 37 miljoonan artikkelin tietokantaa ja tutkimme tieteellisen yhteistyöverkoston kehitystä Euroopassa. Keskitymme tutkimaan vuosina 1973-1977 ja 1998-2010 kirjoitettuja tieteellisiä artikkeleita, joiden kirjoituksessa on ollut mukana eurooppalaisia tutkijoita. Hyödynnämme tulosten visualisoinnissa saatavilla olevia paikkatietoja sekä verkkograafeja. Keskitymme tutkimaan yhteistyöverkostoa kaupunkien tasolla. Analysoinnissa hyödynnämme erilaisia tunnuslukuja sekä verkkoteorian yhteisöanalyysin työkaluja tutkimusklikkien löytämiseksi ja niiden kehityksen tarkastelemiseksi.

Selkeitä trendejä on havaittavissa tutkimusaineistosta. Julkaisujen määrä on noussut huomattavasti sekä yksittäisissä kaupungeissa että kokonaisuudessaan Euroopassa. Lisäksi kaupunkien välinen yhteistyö on vahvistunut. Toisaalta painoarvoltaan suurimpien kaupunkien suhteellinen osuus kaikista julkaistuista artikkeleista on laskenut painoarvoltaan pienten kaupunkien määrän lisääntyessä. Suurimmat julkaisumäärät ovat Moskovalla, Pariisilla ja Lontoolla. Jälkimmäisellä tarkasteluajanjaksolla Etelä-Euroopan ja Turkin kaupunkien julkaisumäärät ovat kasvaneet huomattavan nopeasti verrattuna muihin Euroopan alueisiin. Lisäksi Venäjän kaupunkien vaikutus on hiipunut. Kaupunkien välistä yhteistyötä tarkasteltaessa korostuu maiden sisäinen yhteistyö verrattuna kansainväliseen yhteistyöhön. Maiden sisäinen yhteistyö myös vahvistuu nopeammin kuin kansainvälinen yhteistyö. Yhteisöanalyysin perusteella Euroopassa on tällä hetkellä kaksitoista yhteistyöklikkiä, jotka vaikuttavat muodostuvan kieliperheiden sekä vakiintuneiden yhteistyösuhteiden ympärille.

Avainsanat: Verkkoteoria, tieteen tutkimus, geografia, geokoodaus, yhteistyö tieteessä, kansainvälistyminen, tieteellisten yhteistyöverkoston ja tutkimusklikkien kehitys Euroopassa

Contents

1 Introduction

The field referred to as the science of science studies scientific research itself and seeks to place science in e.g. social, economic or historical context. For example, one can study the amount of citations for each author and see how they have done over time or how the citations are distributed across authors. Other examples include investigating how different funding methods impact the amount of papers or what kind of effect better equipment has on the total number citations. Furthermore, one can inspect how different concepts, ideas or keywords spread through scientific research via collaboration between authors and how they spread between different disciplines of science. The aim of this field is to understand the nature of science and, moreover, improve our knowledge of science itself.

Gaining this knowledge is of interest per se, but the knowledge can also be utilised as a tool in decision making. Bringing forth the structure and the dynamics regarding science can help decision makers improve their R&D decisions and provide them with information about the effects of their decisions. This should lead to more fact based decision making and management regarding the field of science. In addition, it might help individual scientists to find ways to improve their own impactfulness and, furthermore, enable the science community to develop their field of practice systematically.

One natural way to observe the world of science is to see it as a structure similar to any other social construction as it consists of institutes, research groups, individual people and other social entities. An extremely informative and intuitive way to visualise these kind of structures is via network graphs span between different agents. An approach like this also provides us with powerful mathematical tools for analysing these structures. Therefore it serves as a perfect basis for our research on the evolution of the collaboration in the field of science.

This study focuses on the analysis of scientific collaboration from the perspective of network theory. We study how the scientific contribution of different countries and cities has changed over time, how cities within and between countries collaborate with each other, whether or not the recent ICT developments have affected the international collaboration and how countries with shared cultural background or language collaborate with each other.

We utilise mathematical concepts introduced in network theory and extensive geographic information provided by Google to both analyse and visualise available information about the scientific publications. We will only focus on one of the continents, Europe, to retain the scope of this research and to exploit the already thoroughly internalised information about the history of Europe.

The basis of this study will be a publication database provided by Institute for Scientific Information (ISI). This data set contains millions of articles from different journals giving us an extensive foundation for our research. This database combined with recently developed theories in the field of network theory and geographical information provided by Google serve as the perfect starting point for our research.

In the first chapter of the study, we will introduce some of the fundamental theories and concepts of network science necessary to execute this study. This will include the introductions of some basic key figures and one community detection algorithm. In the second chapter, we will clean the data to fit our purposes and utilise geocoding to build the foundation for our visualisations. In the latter part of the data extraction, we will define the most suitable data set for further analysis.

After generating the means to create the necessary visualisations and extracting only the information both necessary and sufficient for the analysis, in the third chapter we will begin with the analysis of characteristic key figures regarding cities and collaboration between them. This will give us a fair idea of the evolutionary events regarding the collaboration network. We will visualise these results in a very simplistic manner. Finally, in the fourth chapter we will extend our research to the community analysis of the scientific collaboration network to unveil the underlying cliques of the network and the evolution they undergo. In this chapter, we will also create more compelling visualisations to make the results of this section easier to grasp.

2 Network Theory

In this chapter we will introduce some basic definitions and concepts used in network science. Firstly, we will go through some basic definitions needed to interpret network graphs. Secondly, we will introduce the network measures used in this study. Thirdly, we will present the community detection method used by Gephi, our network graph tool in this study.

2.1 Mathematical representation of networks

The basic parts of a network graph are nodes (or vertices) v_i and links (or edges), which are represented by a tuple e_i=(v_i,v_j). Nodes represent the interacting units of the system, whereas links represent the interaction between these interacting units. The linkś weight can be defined in multiple ways. In unweighted networks the weights of links are binary, i.e. $w_{ij} = \{0, 1\}$, depending on the fact if an edge exists or not. In weighted networks the weight represents the strength of the interaction.

There are many extensions to this basic model, e.g. directed networks, multigraphs and so forth. The graphs introduced in this study will be weighted, undirected graphs with no self-loops.

2.2 Basic network measures

Next we introduce some basic network measures used in this study.

Node Degree and Strength

One of the most simplistic network measure for nodes is the *node degree* k_i. It describes how many neighbours a node has, that is, how many other nodes are connected to it. A very similar measure for weighted networks is the *node strength* s_i, which is the sum of the weights of the links connected to it[1]:

$$s_i = \sum_{j}^{N} w_{ij}.$$

Naturally, for unweighted networks $k_i = s_i$.

Modularity

In this study we want to reveal community structures in our data. The goal is to separate and identify communities (or clusters) with densely connected nodes, the nodes in different clusters being sparsely linked. In this study we determine the best partition by using a standard goodness measure, *modularity* Q. The modularity is defined as[2, 3]:

$$Q = \frac{1}{2m} \sum_{i,j} [w_{ij} - \frac{s_i s_j}{2m}] \delta(\sigma_i, \sigma_j),$$

where σ_i represents the community (or cluster) to which the node i is assigned. δ is a standard delta function and m is the total weight of links in the whole network.

The Louvain Method

In this study we will be using Gephi as the basic network software tool for plotting network graphs. The community detection algorithm used by Gephi is known as the Louvain method and it is a so-called greedy algorithm which tries to find the global maximum by choosing the local optimum of each step[4][5]. The value to be optimised in this case is the modularity of the network. The algorithm goes as follows[3]:

Step 1: Each node in the network is assigned to its own community. Then for each node i, the change of modularity is calculated for removing i from its own community and moving it into the community of each of its neighbors. Once modularities have been calculated for all communities i is connected to, i is placed into the community that resulted in the greatest modularity. If no increase is possible, i remains in the original community. This process is repeated until no modularity increase is possible.

Step 2: A new network is spanned so that each new node is a community obtained in step 1. Links within communities are replaced with self loops and links from multiple nodes in the same community to a node in a different community are represented by weighted edges between communities. After the new network is spanned, the first step is reiterated.

Steps 1-2 are repeated until no modularity increase is possible.

The Louvain method is an approximate method and due to its greedy nature, it does not ensure that the global maximum of modularity emerges. However, several tests have confirmed that this algorithm is rather accurate and often provides a decomposition with a modularity close to the true maximum[3].

3 Data Extraction and Data Cleaning

3.1 Data set and extraction of the affiliation information

The publication database for this research was obtained from Institute for Scientific Information (ISI). The database contains 37 million publications from multiple citation indices[6]. Each publication has over 30 recorded attributes about the journal and the issue it was published in and over 70 recorded attributes about the publication itself. These attributes include, for example, the item ID, the article title, the language, the author, the author's full address and so forth. From these attributes we compile a data set that contains only the essential information for the construction of the network.

The most essential part of the information available for each publication is the information regarding the authors of the publication. For each publication we have a list of authors and the addresses given by the authors. This affiliation information will allow us to pinpoint each of these authors on a map and to construct a network for each paper.

To construct the desired network, we need coordinates from all the cities which have recorded publications. This enables us to plot all the cities on an actual map, making the results easier to grasp. In order to do this we construct a list of all of the mentioned cities. By counting the occurrence of each city in the data we obtain a data set with the following notation for each line:

```
LONDON:ENGLAND|555988
HOUSTON:USA:TX|232996
```

Notation:

```
CITY:COUNTRY(:STATE)|Reference number
```

In the future we will refer to the first column as 'the key' since it is a unique search parameter. Second column tells us how many times the key occurred in the database, thus giving us a sense of the importance of the key.

3.2 Extraction of the coordinates

The resulting list is used to determine coordinates of cities by using different application programming interfaces (APIs). We use Google's API to pinpoint the initial coordinates and use Wikipedia's API to confirm these coordinates. The Google API is used as default since it can find even the misspelled locations whereas Wikipedia's API is very sensitive to misspelled location names and can be disturbed even by capitalisation of names.

3.2.1 Google application programming interface (API)

To be more specific, we are using The Google Geocoding API. Geocoding stands for the usage of the street address to pinpoint the exact latitude and longitude of a location. An example of the resulting lines is given below:

```
MOSCOW:RUSSIA|(55.742, 37.615)||Moscow||Moscow|RU|227353
```

Notation:

```
Original key|Coordinates|Google location information|Reference number
```

The first column is the original key from the ISI database and the last is the number of occurrences (or importance) of the key. Coordinates and non-capital names we get as a query result from the Google API.

The country codes given by Google API follow the ISO 3166-1 alpha-2 standard, which is widely used as a country codes top-level domain in internet geocoding services[7]. The Google API is highly effective and the fraction of keys that lead to coordinates is 0.851. Considering the high number of misspelled keys and old names of cities and countries in the oldest publications, this is a very satisfactory result.

3.2.2 Wikipedia application programming interface (API)

The Wikipedia API is based on opening the articles in Wikipedia by giving the API an articles title. It is not very practical for extracting coordinates. Firstly it does not provide location information directly but it needs to be mined from the articles themselves. Secondly the API forces us to be very specific about the searched title. Manually searching 'espoo, finland' will lead you to an article but with API the search is unsuccessful. Basically when searching, we need to to search results with all of the combinations of capitalisation.

Furthermore, extracting location information proved to be difficult. In the end we were able to extract some information from Wikipedia. Examples of the resulting lines are given below:

```
WASHINGTON:USA:DC|(38.895, -77.036)|(city)|[District of Columbia]|
[United States]|228012
MOSCOW:RUSSIA|(55.75, 37.616)|(city)|[Central economic
region|Central]|[state power in Moscow is the Moscow City
Duma.}}</ref>]|[Central Federal District| Central]|
[Federation Council of Russia|Federation Council]|227353
```

Notation:

```
Original key|Coordinates|Wikipedia location information
|Reference number
```

As we can see, the information is very tangled especially in the location information section. The information could only be extracted from a very little number of publications leading to the conclusion that the Wikipedia API is not suitable for coordinate extraction. However the entanglement of the resulting lines does not effect the final results as the Wikipedia coordinates are used to confirm coordinates gained from Google API. The location information only gives us the means to compare results from Google and Wikipedia APIs with the naked eye.

3.3 Cleaning the city information

After obtaining the location information from Google and Wikipedia, a cleaning of the gathered information was needed. The remaining unusable keys included publication with e.g. old country names or misspelled locations. Copying information from lines that contained the current name of the country or city information reduced the amount of unusable keys remarkably. In some cases information was copied directly from other keys if the matching of the rows was trivial. In some cases we searched information via Google.

During this process we also assigned countries to locations for which owner of the land was unclear. Such territories were Israel/Palestine, Kosovo and several districts of India. In these cases Google API did not give any country code for the region. For areas around India and Israel we gave their respective ISO country codes (IN and IS) for simplicity. For Kosovo we used the temporary country code XK used by the EU[9]. These areas did not have any impact on the total affiliations of each country as the total affiliation number of these areas was (understandably) negligible.

In the resulting data set the number of unusable keys reduced dramatically from thousands to 14 keys with locations mentioned over 40 times in the publication database. Combined, these keys contain 1400 affiliations, which is less than 0.02 percentage of the affiliations in the data set.

After erasing the unnecessary information, such as Wikipedia information, affiliation count and erasing unusable lines, we obtain a list that enables us to convert keys from the original publication data set to Google API coordinates and current addresses (even for old city/country names). Examples of the lines of the cleaned data set are given below:

```
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
```

Notation:

`Original key from the data|Google API key|Coordinates`

3.4 Choosing the time interval for the study

After we have retrieved coordinates for each city that is mentioned in the database, it is possible to plot these cities on a map. However, the affiliation information is not available for all of the publications. A large number of the publications do not have all the information necessary for the analysis. This is why we need to first assess the data set and determine the appropriate time interval(s) for the analysis.

3.4.1 Evolution of the affiliation percentage

The most simple way to observe the availability of the affiliation data is to plot the evolution of the affiliation percentage. The affiliation percentage of each year represents the relative proportion of the publications with affiliation data compared to the overall amount of publications in the database. As we can notice from figure 1 there is remarkable variation between the years.

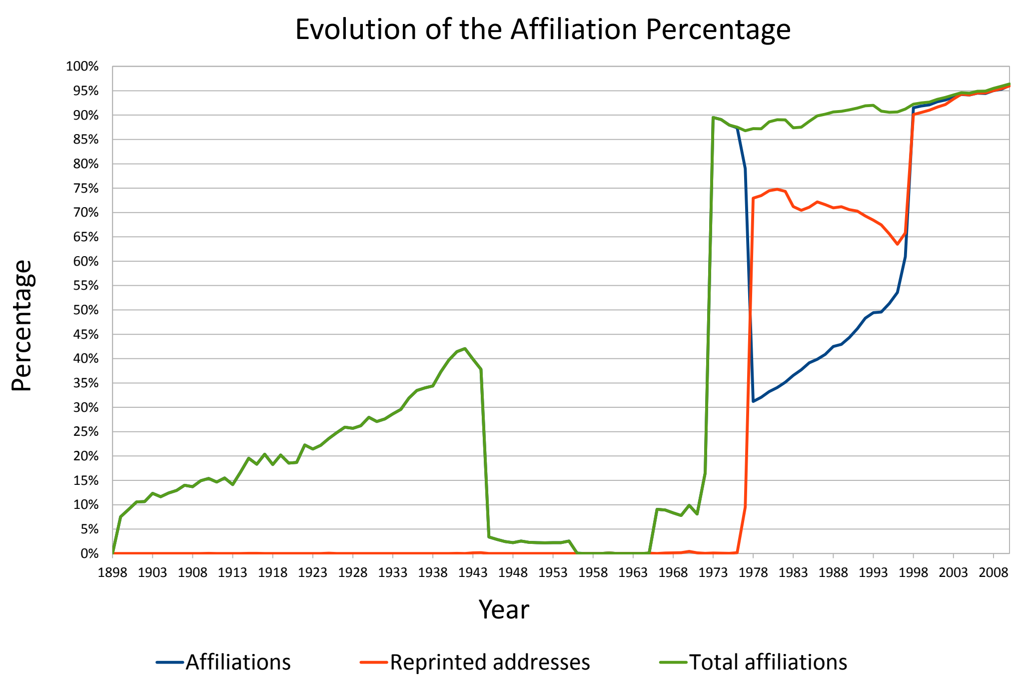


Figure 1: Evolution of the affiliation and reprinted address percentage over time.

In this study we use a threshold of 85 % in terms of affiliation percentage. Above this limit we consider the affiliation information available sufficient for the analysis. The sufficient affiliation information data is available only for the years 1973 to 1977 and between the years 1998 and 2010.

However, there is additional information available on the publications regarding the corresponding author of the paper. This attribute is the "reprinted address" which gives us the address of the corresponding author. Although the reprinted addresses are available for the majority of the papers during the period between years 1978 to 1998, as the affiliation information is available only on the corresponding author, the network analysis can not be carried out. The plotted total affiliations in figure 1 are the relative proportions of the publications which have any affiliation information.

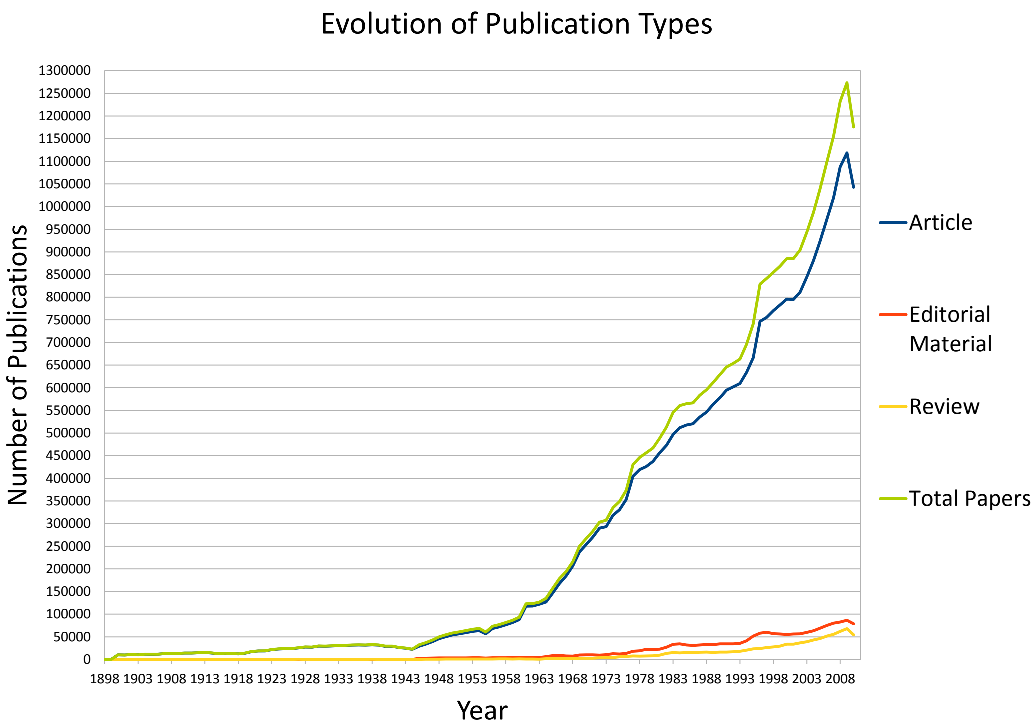


Figure 2: Evolution of the total publications

In figure 2 we can observe the evolution of different publication types in terms of number of publications per year. The dominating publication type is a basic article. Reviews and editorial materials do not play a significant role in the overall picture and they are excluded from the study.

3.4.2 Evolution of the publication languages

Next we study the evolution of the publication languages. We use a proportional scale in figure 3. The total number of articles published in English is tremendous in the new millennium. The evolution of other publication languages would not be visible in the plot constructed from the absolute numbers of the publications as the number of articles for other languages is low.

As shown in the figure 3, the evolution of publication languages does not explain the variation in affiliation percent between years 1978-1997.

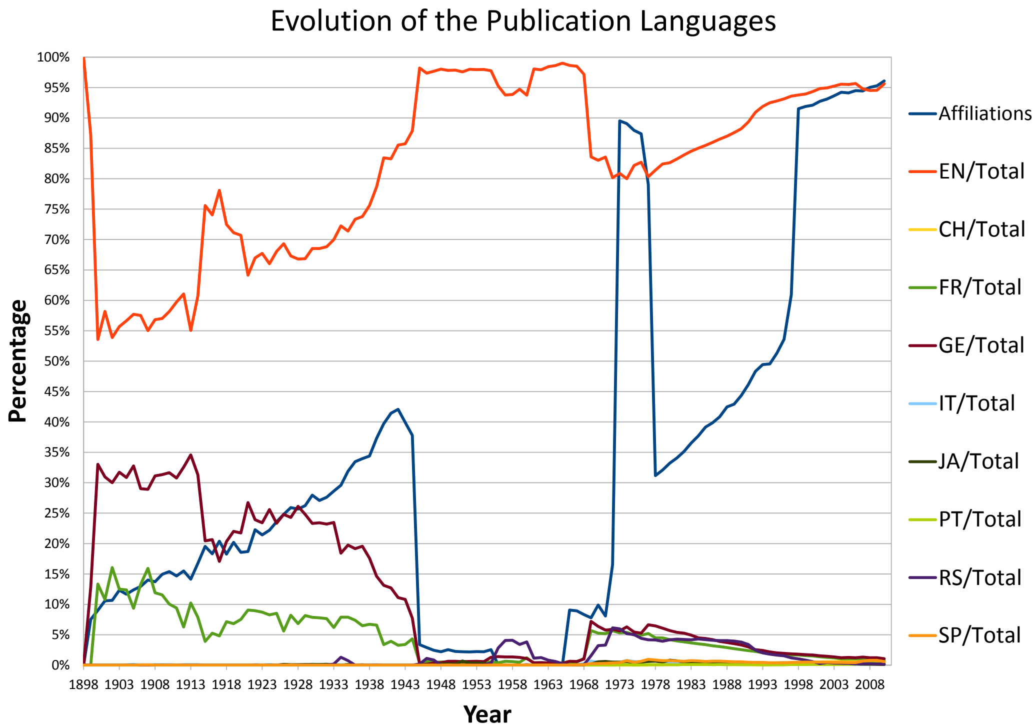


Figure 3: Evolution of the languages in publications

3.5 Constructing the data set for the analysis

Thus the analysis can be carried out only from the year 1973 onwards. Further analysis of the evolution of the collaboration network can be done for periods between the years 1973-1977 and 1998-2010. The fragmentation of the data is not considered to be a problem as we are able to compare the networks between the seventies and the new millennium and the evolution in the new millennium is of interest per se.

3.5.1 Extracting the information on cities and their connections from the chosen data set

For the chosen article types and time periods, we convert the original data to commensurable keys. Next, we construct data sets that reveal us the weight of the cities and links (i.e the strength of collaboration) between them.

The weight of one publication is distributed as follows: if a publication has four authors and two of them are from the same city, one city will obtain a weight of $\frac{1}{2}$ and other two obtain a weight of $\frac{1}{4}$. Weights of the links are derived similarly. If there are 4 authors from different cities, each link weighs $\frac{1}{10}$ since there are 6 links between nodes and 4 self-links. The total weights of the cities and the links between them are simply the sum of the gained weights from all of the articles in the chosen data set.

We will not include self-links in the construction of the network or in the statistics since we are not interested in the collaboration within cities but between them. Furthermore, we exclude weak links that have a weight of less than one from the data and only assess strong collaboration relationships.

To construct a respective network for each year separately, we divide the publications of each year to separate data sets. We regard this resolution high enough for our purposes. The information of the nodes is separated from the information of the links so that these data sets can be analysed independently.

3.5.2 Definition of Europe

In this study we will extend the analysis only for Europe in order to keep the scope of this study confined andand to exploit the already thoroughly internalised information about the history of Europe. The publication and location data are available for every continent so it is be possible to expand this research in the future.

To narrow down the data set to Europe, we first need to define what we mean by Europe in this study and determine which countries to be included in the analysis and if we want to add geographical constraints. We used the member countries of the Council of Europe as the base for the list of countries and added Belarus, Vatican and Kosovo to the list. List of included countries and their ISO 3166-1 country codes can be seen in the appendix A[10][8].

In order to exclude eastern parts of Russia we added a geographical constraint for Russian publications. All cities and links that involved cities with longitude over 63 degrees were excluded from this study. This is approximately the longitude of the Ural Mountains, which is considered to be the boarder between the European and Asian continents[11].

With these constraints we are able to exclude cities outside Europe from the data and construct new data sets for European cities only. In publication with intercontinental collaboration the author located outside Europe are excluded. However, European cities have their proportional share of the intercontinental publication and links between authors in European institutes are included in the study. For example, if there are two American cities and two European cities involved in a study (with one author each), the European cities get a weight of $\frac{1}{4}$ and the link between the two European cities gets a weight of $\frac{1}{10}$, other links are not taken into account in our study.

3.5.3 Networks inside and between the countries

In addition to the data sets that contain all the links in Europe, we constructed data sets that only include links within countries and outside them. This way we can observe the collaboration between cities in different countries or cities within the same country side by side with the collaborative network of the whole Europe. Domestic and international network graphs are naturally subgraphs of the whole collaboration network. This will allow us to have a more detailed view on the evolutionary events and will enable us to see whether the fluctuation in the whole network happens because of the evolution of collaborative networks inside countries, between them or both.

When observing the graphs, we must keep in mind that the cities and countries are converted to current cities and countries regardless of the publication year. This leads to the fact that e.g. areas once part of the Soviet Union are now shown as independent countries and international links that once were actually within countries' borders are now shown as international links.

With these more focused data sets we can finally begin the analysis and see how the collaboration has evolved in Europe.

4 Basic Statistics of the Scientific Collaboration Network in Europe

In the previous section we have constructed the data sets for each year with only European cities with publication from appropriate time intervals. In this section we will focus on basic statistics and observations about the data set and discuss network construction and more profound analysis in the next chapter.

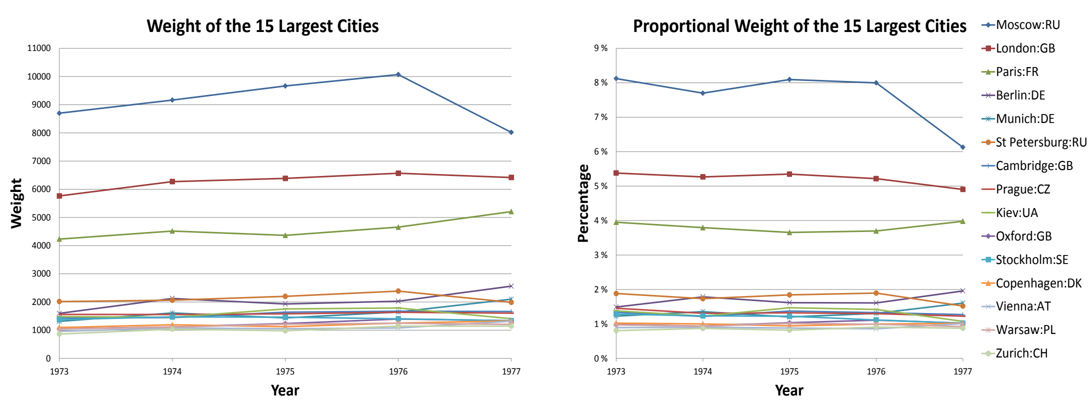
4.1 Overall progress and amount of publications at the city level

In order to have meaningful graphs out of the data we must select a suitable threshold that enables us to eliminate excess information from the graphs. The analysis in this section is made so, that only 15 largest cities and 150 strongest links are included.

In this subsection, the order of cities in the graphs is based on the last year's weight, e.g. the plot of 1973-1977 is based on the 15 largest cities sorted according to weights of the year 1977. This excludes cities that were amongst the 15 largest cities in the previous years but are not in the last year in the interval of study. Each graph in this subsection is presented both in absolute weights and in proportional scale.

4.1.1 Amount of affiliations in 1973-1977

The weight of affiliations and the percentage according to it are presented in figure 4 for the years between 1973 and 1977.



(a) Absolute weight (b) Proportional weight

Figure 4: The evolution of the weight of the 15 largest cities (1973-1977)

The absolute numbers and gradients for the 15 largest cities can be seen in table 1.

City	1973	1974	1975	1976	1977	Slope
Moscow:RU	8698	9164	9664	10070	8018	-170
London:GB	5761	6271	6385	6568	6417	164
Paris:FR	4232	4517	4363	4656	5207	244
Berlin:DE	1595	2126	1930	2027	2562	242
Munich:DE	1312	1612	1434	1644	2100	197
St Petersburg:RU	2016	2063	2200	2386	1988	-7,06
Cambridge:GB	1435	1448	1638	1671	1662	56,7
Prague:CZ	1558	1557	1580	1634	1605	11,6
Kiev:UA	1479	1460	1755	1788	1406	-18,2
Oxford:GB	1050	1113	1234	1396	1337	71,8
Stockholm:SE	1376	1470	1461	1403	1334	-10,4
Copenhagen:DK	1095	1190	1124	1251	1330	58,7
Vienna:AT	957	1076	1052	1080	1320	90,7
Warsaw:PL	1041	1106	1213	1248	1207	41,5
Zurich:CH	864	1038	978	1142	1145	70,3

Table 1: The 15 largest cities (1973-1977)

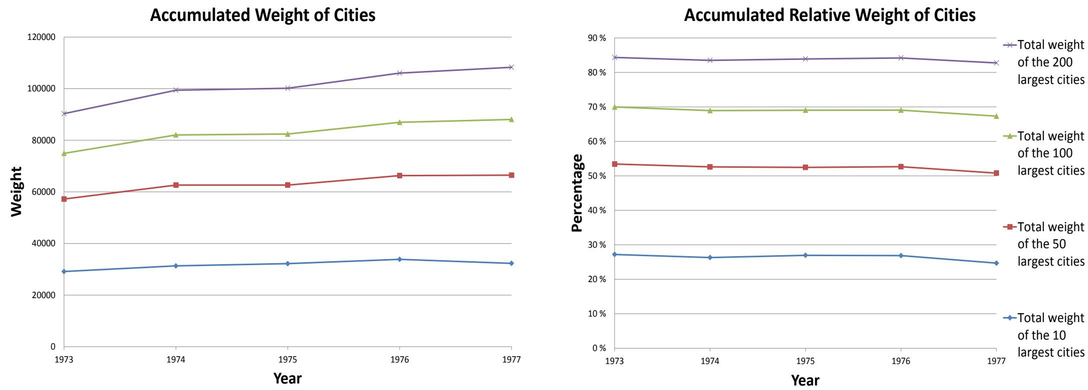
As we can see from the figure 4 and table 1, the dominating cities during this era are Moscow, London and Paris.

In the next vicinity we got Berlin, Munich, St Petersburg, Cambridge and Prague with weight of over 1500 in the year 1977. We can detect from figure 4 that although the total number of affiliations is on the rise for the cities with heaviest weights, their proportional size is decreasing. London, Paris, Berlin, Munich and Vienna have the fastest growth-rate and Moscow has the lowest gradient due to fall in 1977 as can be determined from table 1.

Figure 5 shows that the proportional size of accumulated affiliations of the 200 largest cities is slightly decreasing although the total amount is increasing. This is confirmed by the gradients. All absolute numbers and gradients can be found in table 2.

Year	1973		1974		1975		1976		1977		Gradient	
Total weight of the	Absolute weight	Percentage	Abs.	%	Abs.	%	Abs.	%	Abs.	%	Abs.	%
10 largest cities	29137	27 %	31331	26 %	32184	27 %	33841	27 %	32301	25 %	791	-0,0063%
50 largest cities	57216	53 %	62622	53 %	62641	52 %	66303	53 %	66465	51 %	2312	-0,0066%
100 largest cities	74938	70 %	82065	69 %	82418	69 %	86956	69 %	88066	67 %	3282	-0,0067%
200 largest cities	90313	84 %	99407	84 %	100167	84 %	1060208	84 %	108270	83 %	4489	-0,0039%

Table 2: The accumulated weights of the largest cities (1973-1977)

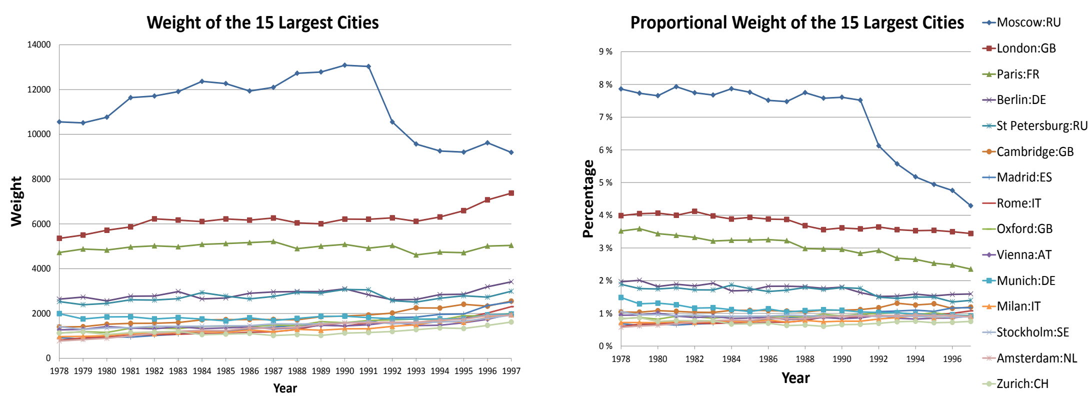


(a) Accumulated absolute weight (b) Proportional accumulated weight

Figure 5: The evolution of the accumulated weight of the cities (1973-1977)

4.1.2 Amount of affiliations in 1978-1997

Since the full information is unavailable, this section will be based on the reprinted addresses mentioned in section 3.4.1. The weight of the 15 largest cities and the percentages according to them are presented in figure 6. The absolute numbers and gradients for the 15 largest cities can be seen in table 3. As there is no collaboration information available, each publication contributes as weight of one for the city.



(a) Absolute weight (b) Proportional weight

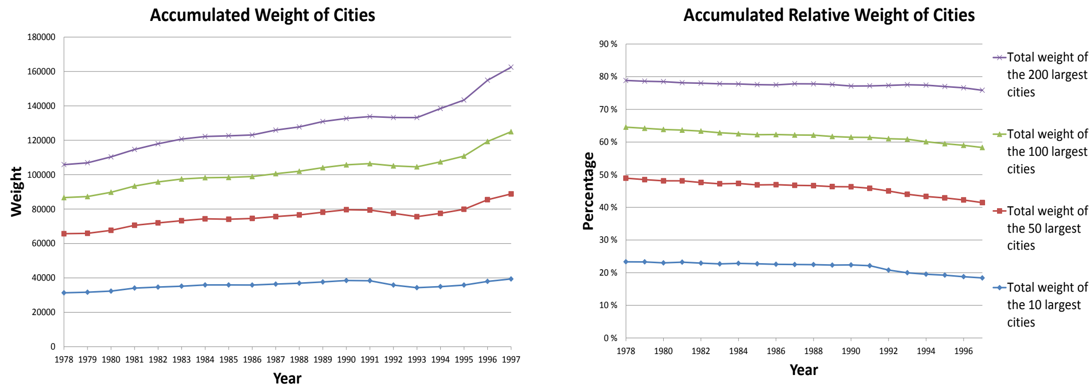
Figure 6: The evolution of the weight of the 15 largest cities (1978-1997)

City	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
Moscow:RU	10556	10511	10764	11637	11710	11907	12364	12267	11934	12096	12727
London:GB	5355	5503	5716	5867	6226	6169	6105	6221	6168	6264	6046
Paris:FR	4724	4878	4832	4968	5022	4981	5086	5123	5168	5215	4897
Berlin:DE	2641	2733	2561	2774	2780	2980	2650	2693	2903	2963	2980
St Petersburg:RU	2531	2391	2452	2613	2600	2662	2932	2775	2653	2762	2941
Cambridge:GB	1398	1409	1521	1559	1564	1599	1716	1727	1728	1732	1717
Madrid:ES	920	858	958	943	1017	1074	1195	1207	1322	1391	1478
Rome:IT	802	909	950	1027	1054	1075	1108	1089	1164	1180	1284
Oxford:GB	1122	1187	1156	1340	1353	1322	1424	1441	1407	1484	1460
Vienna:AT	1268	1290	1417	1348	1316	1410	1327	1355	1398	1311	1365
Munich:DE	1992	1756	1845	1853	1755	1817	1752	1670	1808	1686	1778
Milan:IT	968	978	1008	1101	1092	1159	1191	1154	1269	1185	1290
Stockholm:SE	1414	1289	1355	1363	1435	1427	1434	1447	1458	1536	1530
Amsterdam:NL	772	834	894	1003	1107	1206	1246	1235	1360	1347	1384
Zurich:CH	1131	1178	1062	1143	1176	1200	1055	1079	1110	1015	1058
City		1989	1990	1991	1992	1993	1994	1995	1996	1997	Slope
Moscow:RU		12784	13085	13031	10552	9567	9257	9205	9622	9192	-71,8
London:GB		6005	6214	6208	6272	6115	6310	6591	7072	7375	106
Paris:FR		5005	5086	4917	5033	4614	4743	4716	5011	5041	16,7
Berlin:DE		2980	3101	2835	2613	2625	2844	2859	3194	3416	40,8
St Petersburg:RU		2913	3073	3056	2575	2507	2682	2791	2731	2994	24,4
Cambridge:GB		1857	1884	1933	2019	2250	2244	2409	2327	2553	60,8
Madrid:ES		1487	1573	1619	1817	1839	1963	1978	2356	2516	84
Rome:IT		1501	1438	1577	1612	1577	1696	1792	2024	2306	79,2
Oxford:GB		1636	1587	1690	1714	1755	1713	1901	1889	2006	46,5
Vienna:AT		1470	1443	1495	1624	1461	1476	1586	1741	1983	37,6
Munich:DE		1870	1882	1814	1765	1747	1757	1756	1935	1975	-0,895
Milan:IT		1252	1314	1317	1425	1496	1658	1604	1820	1943	51,3
Stockholm:SE		1542	1576	1632	1544	1575	1619	1655	1787	1925	26,9
Amsterdam:NL		1514	1573	1591	1609	1607	1705	1739	1902	1916	60,2
Zurich:CH		1019	1125	1150	1196	1278	1356	1329	1468	1617	25,6

Table 3: The 15 largest cities (1978-1997)

As we can detect from figure 6, Moscow is dominating the amount of affiliations until the fall of the Soviet Union in the end of 1991. After that the weight of Moscow is rapidly decreasing. On the other hand, London and Paris are keeping their position as the second and third largest cities in terms of affiliations. In the next vicinity there are Berlin, St Petersburg, Cambridge, Madrid and Rome with weight of over 2300. London, Madrid and Rome have the fastest growth-rate and Moscow has the lowest gradient due to the fall of the Soviet Union.

We can see the percentage of publication decreasing for almost all of the cities in the figure 6. Only Berlin and London are maintaining their share of total affiliations. 7 shows that the accumulated affiliations of the 200 largest cities are decreasing although the total amount is increasing. This is confirmed by the gradients in table 4.



(a) Accumulated absolute weight (b) Proportional accumulated weight

Figure 7: The evolution of the accumulated weight of the cities (1978-1997)

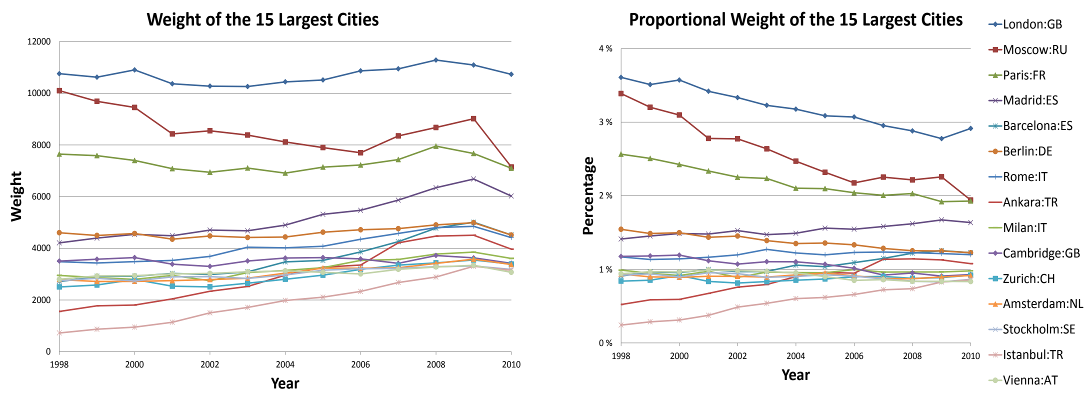
Year		1978	1979	1980	1981	1982	1983	1984	1985	1986	1987
Total weight of the	Absolute weight										
10 largest cities		31317	31669	32327	34076	34642	35179	35907	35898	35845	36398
50 largest cities		65699	65922	67650	70594	71978	73236	74349	74091	74561	75624
100 largest cities		86665	87272	89781	93392	95765	97482	98244	98400	98937	100573
200 largest cities		105856	106884	110358	114652	117952	120728	122204	122602	123090	125920
	Percentage										
10 largest cities		23 %	23 %	23 %	23 %	23 %	23 %	23 %	23 %	23 %	22 %
50 largest cities		49 %	48 %	48 %	48 %	48 %	47 %	47 %	47 %	47 %	47 %
100 largest cities		65 %	64 %	64 %	64 %	63 %	63 %	63 %	62 %	62 %	62 %
200 largest cities		79 %	79 %	78 %	78 %	78 %	78 %	78 %	78 %	77 %	78 %
Year	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	Gradient
Total weight of the	Absolute weight										
10 largest cities	36895	37638	38484	38361	35831	34310	34928	35828	37967	39382	424
50 largest cities	76607	78178	79625	79438	77552	75590	77518	79877	85462	88811	1216
100 largest cities	101990	104079	105723	106423	105137	104547	107472	110784	119305	124963	2016
200 largest cities	127766	130899	132679	133759	133225	133188	138472	143402	154891	162500	2981
	Percentage										
10 largest cities	22 %	22 %	22 %	22 %	21 %	20 %	20 %	19 %	19 %	18 %	-0,0026 %
50 largest cities	47 %	46 %	46 %	46 %	45 %	44 %	43 %	43 %	42 %	41 %	-0,0039 %
100 largest cities	62 %	62 %	61 %	61 %	61 %	61 %	60 %	59 %	59 %	58 %	-0,0033 %
200 largest cities	78 %	78 %	77 %	77 %	77 %	78 %	77 %	77 %	77 %	76 %	-0,0016 %

Table 4: The accumulated weights of the largest cities (1978-1997)

4.1.3 Amount of affiliations in 1998-2010

The relative and absolute amounts of affiliation are presented in the figure 8. The absolute numbers and gradients for the 15 largest cities can be seen in table 5.

As we can observe from the figure 8, the numbers represented in the previous section are biased and in reality the dominating city in terms of affiliations is London. Moscow and Paris are next in terms of weight and Madrid is almost the same size as the two previously mentioned. In the second vicinity, there are Barcelona, Berlin and Rome with weight of over 4000. We can detect the rise of Ankara and Istanbul very clearly by looking into figure 8 and table 5. Other major cities with rapid growth-rate are Madrid and Barcelona. Moscow's number of total affiliations is decreasing rapidly.



(a) Absolute weight (b) Proportional weight

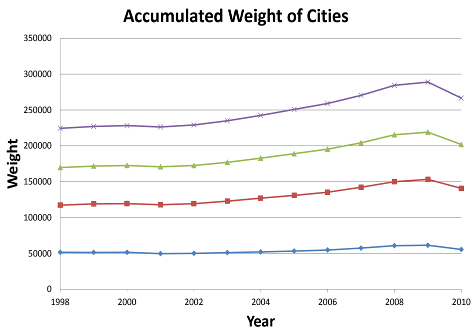
Figure 8: The evolution of the weight of the 15 largest cities (1998-2010)

City	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Slope
London:GB	10758	10622	10905	10368	10276	10260	10442	10514	10868	10945	11285	11094	10733	-2,08
Moscow:RU	10101	9690	9452	8428	8548	8383	8116	7899	7698	8349	8677	9016	7147	-246
Paris:FR	7646	7587	7400	7083	6943	7106	6911	7142	7224	7435	7953	7669	7100	-45,5
Madrid:ES	4210	4395	4538	4487	4706	4678	4898	5313	5471	5865	6345	6682	6022	151
Barcelona:ES	2710	2916	2919	3027	2987	3085	3473	3531	3860	4256	4780	5009	4514	150
Berlin:DE	4602	4497	4571	4353	4476	4418	4437	4623	4718	4761	4907	4990	4513	-7,42
Rome:IT	3483	3432	3486	3530	3688	4042	4019	4077	4351	4576	4801	4851	4414	77,6
Ankara:TR	1556	1773	1805	2042	2336	2523	2961	3269	3353	4212	4474	4506	3966	201
Milan:IT	2950	2864	2797	2952	2746	3060	3143	3250	3530	3567	3776	3853	3606	54,7
Cambridge:GB	3505	3577	3642	3385	3300	3509	3624	3642	3590	3418	3722	3630	3419	-7,19
Zurich:CH	2504	2579	2796	2534	2507	2649	2802	2960	3182	3348	3437	3546	3390	73,8
Amsterdam:NL	2789	2717	2719	2748	2789	2841	3039	3232	3229	3245	3416	3580	3373	48,7
Stockholm:SE	2741	2841	2734	2895	2904	2836	2969	3152	3210	3223	3290	3314	3183	36,8
Istanbul:TR	724	868	950	1139	1502	1712	1980	2108	2330	2677	2886	3307	3151	202
Vienna:AT	2801	2930	2941	3009	3026	3090	3127	3069	3008	3187	3276	3344	3072	22,5

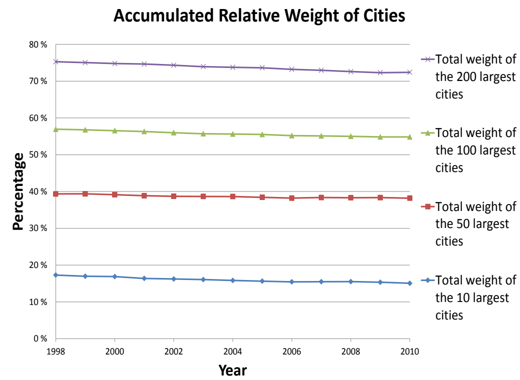
Table 5: The 15 largest cities (1998-2010)

From the figure 8 we can detect that the proportional sizes of the largest cities is decreasing as most of the curves are declining. However, there are couple of exceptions with Madrid, Barcelona, Ankara and Istanbul increasing their share in the top 15.

From the figure 9 we can determine, that the total amount of affiliations gathered by the 10 largest cities has stopped increasing, but that the total affiliations gathered by the 50, 100 and 200 largest cities continue to ascend. At the same time, we can observe from figure 9, that the relative weights are decreasing. This is confirmed by the gradients in table 6.



(a) Accumulated absolute weight



(b) Proportional accumulated weight

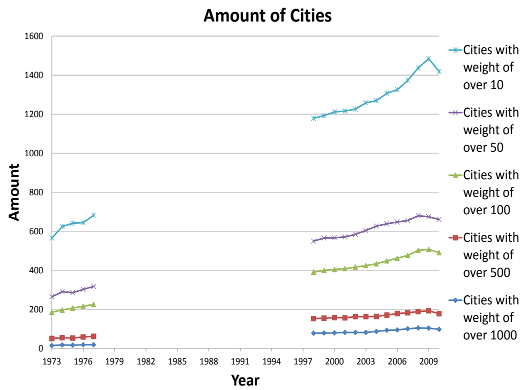
Figure 9: The evolution of the accumulated weight of the cities (1978-1997)

Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	Gradient
Total weight of the	Absolute weight													
10 largest cities	51521	51353	51514	49654	50006	51065	52025	53263	54660	57384	60718	61299	55434	326
50 largest cities	117282	119096	119457	117816	119296	122878	127000	130877	135245	142179	150042	153208	140613	1944
100 largest cities	169697	171741	172577	170720	172536	177017	182820	189044	195322	204233	215445	219076	201848	2679
200 largest cities	224384	227024	228224	226318	229079	234972	242416	250799	259104	270346	284348	288888	266545	3513
	Percentage													
10 largest cities	17 %	17 %	17 %	16 %	16 %	16 %	16 %	16 %	15 %	15 %	16 %	15 %	15 %	-0,19 %
50 largest cities	39 %	39 %	39 %	39 %	39 %	39 %	39 %	38 %	38 %	38 %	38 %	38 %	38 %	-0,10 %
100 largest cities	57 %	57 %	57 %	56 %	56 %	56 %	56 %	56 %	55 %	55 %	55 %	55 %	55 %	-0,18 %
200 largest cities	75 %	75 %	75 %	75 %	74 %	74 %	74 %	74 %	73 %	73 %	73 %	72 %	72 %	-0,24 %

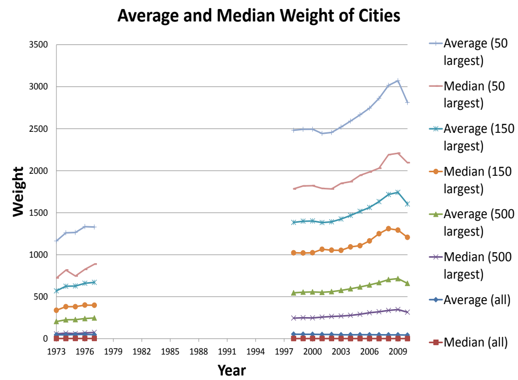
Table 6: The accumulated weights of the largest cities (1998-2010)

4.1.4 Basic statistics

In table 7 and figure 10, we can observe the basic statistics of all of the cities in different years. These statistics are not derived for the years 1978-1997 since only the reprinted addresses are available for the years 1978-1997.



(a) Evolution of the amount of different sized cities.



(b) Evolution of the average and median of the top cities.

Figure 10: The evolution of the characteristics key figures of the cities

Year	1973	1974	1975	1976	1977	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Average and median of largest cities																		
50 average	1163	1260	1264	1334	1329	2481	2493	2493	2443	2455	2519	2591	2665	2744	2861	3016	3071	2812
50 median	723	816	745	826	887	1784	1818	1822	1789	1783	1849	1869	1945	1985	2029	2190	2208	2095
150 average	569	624	627	660	670	1384	1398	1401	1382	1391	1425	1467	1516	1562	1632	1716	1742	1604
150 median	338	381	381	400	398	1023	1020	1023	1064	1054	1052	1093	1107	1164	1250	1310	1292	1207
500 average	204	225	226	239	246	546	554	557	552	560	575	594	616	640	667	703	715	659
500 median	57,1	66,0	60,0	68,4	74,2	245	249	247	257	264	269	278	290	309	321	336	347	315
All average	45,9	45,1	48,8	50,7	46,8	53,5	52,4	52,0	49,6	49,9	46,0	46,2	46,8	46,8	46,1	45,5	45,1	42,6
All median	2,17	2,17	2,50	2,33	2,33	1,50	1,42	1,50	1,25	1,33	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Numbero of Cities with weight of																		
over 1000	14	17	16	18	18	77	78	79	81	81	81	86	92	94	100	104	103	97
over 500	50	54	52	58	61	152	154	157	156	162	162	163	170	178	182	188	192	177
over 100	185	196	206	215	225	390	399	404	408	416	424	433	448	461	476	502	507	490
over 50	264	290	285	302	317	549	565	566	571	584	604	626	638	647	655	679	674	660
over 10	565	624	641	644	682	1179	1192	1211	1216	1226	1258	1269	1307	1325	1373	1437	1484	1418

Table 7: The characteristics key figures of the largest cities (1973-2010)

We can notice that the average and median number of affiliations in the categories of the 50, 100 and 500 largest cities are increasing but the median and average number of all cities is decreasing. In addition we can determine that the number of cities with over 1000 affiliations has not grown remarkably. Moreover, less the number of affiliations required for the category, more the number of cities in that category increases in time and the greater the gradient is. Gradients can be found in table 8.

As we can tell from table 8, the rate of growth is greater during the years 1998-2010 in comparison to 1973-1977 in every category. In addition, we can determine that the growth has been even faster during the years 1978-1997, since the gradient for the years 1998-2010 is smaller than the overall gradient of 1973-2010.

	Gradient			
Average and median of largest cities	1973-1977	1977-1998	1998-2010	Overall
50 average	6,65	54,9	27,6	44,6
50 median	6,55	42,7	25,9	37,1
150 average	4,04	34,0	18,3	28,0
150 median	2,38	29,8	15,3	23,5
500 average	1,69	14,3	9,41	12,3
500 median	0,685	8,12	5,88	6,98
All average	0,0378	0,320	-0,911	-0,0886
All median	0,00667	-0,0397	-0,0417	-0,0315
Cities with weight of				
over 1000	0,16	2,81	1,67	2,24
over 500	0,44	4,33	2,08	3,43
over 100	1,60	7,86	8,33	8,24
over 50	2,12	11,0	9,25	10,7
over 10	4,68	23,7	19,9	23,05

Table 8: The gradients of the characteristics key figures (1973-2010)

4.1.5 Total publications in the cities of Europe

By observing the graphs presented in sections 4.1.1 to 4.1.3, we can determine the temporal evolution on the city level.

Firstly, we can acknowledge the fact that Moscow, London and Paris are the top three cities in terms of publications in the data set. The publication amount of Moscow has fallen from the peak of 10 070 to 7 147, but remains as the second largest city in the new millennia. In the other hand, the weight of London has been constantly over 10 000 during the years 1998-2010. London has surpassed Moscow somewhere in between 1978-1977, but the data of that time is inconclusive. The Paris has been the third largest city in terms of publications throughout the data set. The weight of Paris has increased from 4232 to 7100.

The inspection of the 15 largest cities is more captivating. Firstly, we deduce that the amount of affiliations of the 15th largest city has risen from 864 to 3071. Secondly, we can detect that in the 1970s was dominated by the cities in the Central Europe. However, the South European cities, such as Madrid, Barcelona and Istanbul, have been on the rise during the 2000s. They have increased their amount of affiliations rapidly whereas the Middle European cities have remained relatively constant.

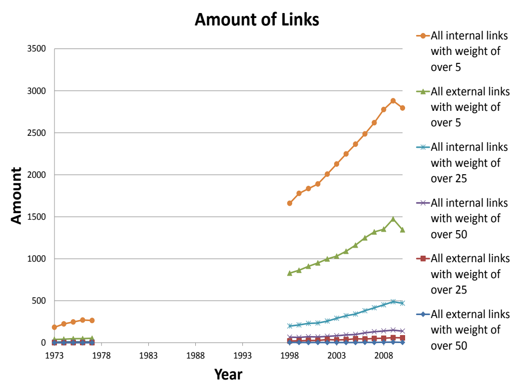
In general, the trend for the cities is mostly in the ascendant, although many of the gradients are only barely positive. In the other hand, there are many gradients that are approximately zero and some are even negative. Overall, there is little variation if we exclude the rise of certain cities and descend of the Russian cities such as Moscow and St. Petersburg.

When observing the graphs in sections 4.1.1 to 4.1.3, we can determine that although the amount of affiliations of the top cities has ascended from the 1970s, the proportional share has been descending in the 2000s. This is confirmed by the figures 4, 6 and 8. The phenomena is explained by the figures 5,7 and 9 as the share of 200 largest cities is decreasing. This implicates, that there is increasing number of cities with smaller weights and that they are increasing their combined share of the total affiliations. This is confirmed by figure 10. The combined weight of the 200 largest cities has decreased from 0.843 to 0.724 and for 10 largest cities the descent is from 0.272 to 0.151 although according to figures 5, 7 and 9 suggest, that the total weights of 10 and 200 largest cities are ascending.

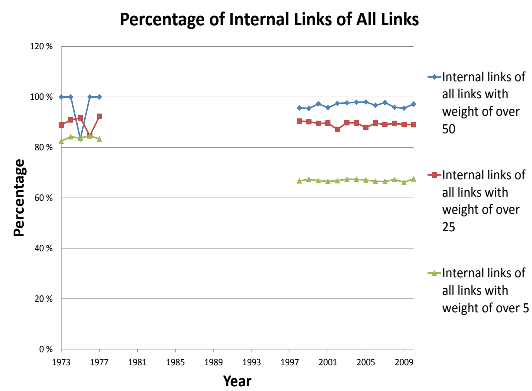
4.2 Statistics of the collaboration between cities

While observing the results of this chapter, one must bear in mind that links with weight of under one are excluded from the data set.

We can observe the number of external and internal links with weights over 5, 25 and 50 in figure 11 and in table 9. The proportion of internal links is presented in figure 11.



(a) Different weighted links



(b) Proportion of internal links

Figure 11: The evolution of the links

Year	1973	1974	1975	1976	1977	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Links with weight of over																		
External links																		
50	0	0	1	0	0	3	3	2	3	2	2	2	2	4	3	6	7	4
25	1	1	1	2	1	21	23	27	27	38	33	37	47	44	51	53	60	58
5	39	42	48	49	53	828	863	910	950	997	1031	1089	1161	1249	1319	1351	1474	1344
Internal links																		
50	2	3	5	4	3	66	63	71	68	76	83	93	98	117	131	140	150	138
25	8	10	11	11	12	198	212	230	234	257	291	321	341	381	416	451	488	470
5	184	224	247	270	265	1660	1778	1835	1890	2007	2128	2248	2364	2486	2620	2777	2881	2795

Table 9: The amount of links (1973-2010)

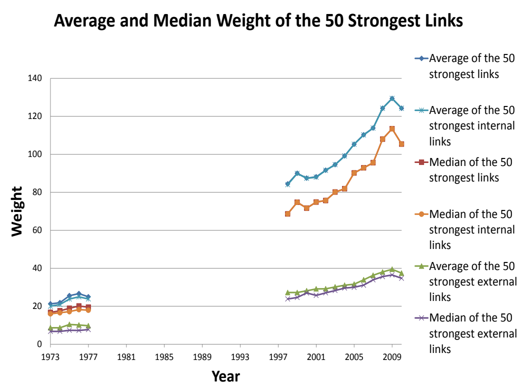
We can determine from the figure 11 that number of internal links is a great deal higher than the number of external links in each category. In addition, the gradient of internal links is greater than gradient of external links, as we can notice in table 10. However, as seen in the figure 11, the proportions of internal and external links stay practically constant during the years 1998-2010. In 1973-1977 the data set is small and minor changes to data can swing the plot. In addition, we can determine

	Gradient			
Links with weight of over	1973-1977	1977-1998	1998-2010	Overall
External links				
50	0,00	0,143	0,0833	0,108
25	0,00	0,95	3,08	1,54
5	3,50	36,90	43,0	35,3
Internal links				
50	0,25	3,00	6,00	3,68
25	1,00	8,86	22,7	12,5
5	20,25	66,4	94,6	70,6

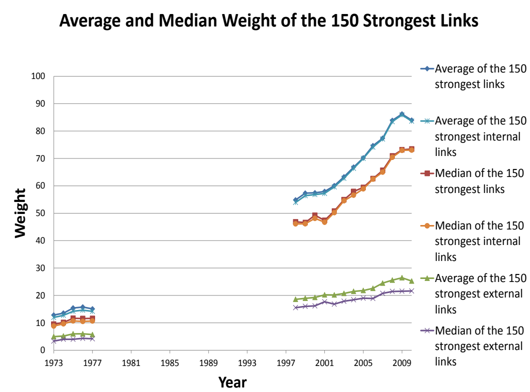
Table 10: The gradients of the number of links (1973-2010)

by observing figure 11, that in 1978 to 1997 there has been significant rise in the amount of external links with weight of over 5.

To see how the weight of 50 and 150 strongest links have evolved, we have plotted curves in figure 12.



(a) The 50 strongest links



(b) The 150 strongest links

Figure 12: The evolution of the characteristics key figures of the links.

Year	1973	1974	1975	1976	1977	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
The 150 strongest links																		
All links																		
Average	12,8	13,5	15,4	15,8	15,1	54,9	57,3	57,5	57,9	60,0	63,3	66,8	70,2	74,6	77,4	83,9	86,2	84,0
Median	9,50	10,2	11,7	11,6	11,6	46,9	46,6	49,3	47,4	50,8	54,9	57,9	59,4	62,6	65,7	70,9	73,2	73,5
External links																		
Average	4,98	5,19	5,96	5,99	5,71	18,6	18,9	19,3	20,2	20,2	20,7	21,5	21,8	22,6	24,5	25,6	26,5	25,3
Median	3,33	4,00	4,00	4,33	4,17	15,5	16,0	16,2	17,7	16,8	17,9	18,4	19,0	18,9	20,7	21,4	21,5	21,6
Internal links																		
Average	12,0	12,8	14,2	14,6	14,2	53,9	56,4	56,8	57,2	59,6	62,8	66,4	70,0	74,2	77,1	83,4	85,9	83,6
Median	8,83	9,63	10,7	10,5	10,7	46,1	46,2	48,1	46,7	50,2	54,6	56,6	58,9	62,4	65,0	70,4	73,0	73,1
The 50 strongest links																		
All links																		
Average links	21,2	21,9	25,6	26,7	25,0	84,4	90,0	87,5	88,1	91,6	94,6	99,2	105	110	114	124	129	124
Median	16,8	17,6	18,9	20,1	19,5	68,6	74,7	71,6	74,9	75,7	80,1	81,9	90,2	92,9	95,6	108	113	105
External links																		
Average	8,66	8,67	10,4	10,1	9,7	27,2	27,2	28,1	29,1	29,2	30,2	31,1	31,6	33,9	36,3	38,1	39,5	37,4
Median	6,75	6,77	7,36	7,25	7,71	23,8	24,6	27,0	25,7	27,1	28,3	29,6	30,0	31,1	33,9	35,6	36,5	34,7
Internal links																		
Average	20,0	20,9	23,8	25,0	23,8	84,0	89,9	87,4	88,0	91,5	94,6	99,1	105	110	114	124	129	124
Median	16,0	16,6	17,2	18,3	17,9	68,6	74,7	71,6	74,9	75,7	80,1	81,9	90,2	92,9	95,6	108	113	105

Table 11: The average and median weights of the strongest links (1973-2010)

	Gradient			
Year	1973-1977	1977-1998	1998-2010	Overall
The 150 strongest links				
All links				
Average	0,563	1,89	2,42	1,92
Median	0,529	1,68	2,22	1,73
External links				
Average	0,184	0,612	0,557	0,548
Median	0,208	0,542	0,506	0,494
Internal links				
Average	0,555	1,89	2,47	1,94
Median	0,458	1,69	2,25	1,74
The 50 strongest links				
All links				
Average	0,930	2,83	3,32	2,78
Median	0,688	2,34	3,06	2,40
External links				
Average	0,267	0,833	0,849	0,777
Median	0,240	0,766	0,911	0,756
Internal links				
Average	0,951	2,87	3,35	2,82
Median	0,472	2,42	3,06	2,42

Table 12: The gradients of the characteristics key figures of the strongest links (1973-2010)

We can determine from the figure 12, that basically all of 50 strongest links are within countries and that the internal links are getting stronger in a greater growth-rate than the 50 strongest links between countries. By observing figure 12, we can determine that this is the case for 150 strongest links as well. Gradients can be seen in table 12. However, there is only a small difference between the 150 largest internal links and 150 largest links of all since there are only a few international links amongst the top 150 strongest links. By looking at appendix B of the 150 strongest links in 2010, we find out that the total amount of external links in the 150 strongest links is only 5.

5 Community Analysis of the Collaboration Network in Europe

In this section we will be inspecting the temporal evolution of the collaboration network in total and in domestic and international components separately.

Since it is not very practical to display three different network for each year, we focus our study on years 1973, 1977, 1998 and 2010. For these years we have three network graphs available: The whole collaboration network, the international collaboration network and the domestic collaboration network. However, as can be seen in figure 13, the domestic network graphs bring no new information and they are thus excluded for the years 1977, 1998 and 2010. All other useful information regarding the evolution during these years can be found from section 4. The networks are constructed so, that only 150 largest cities and 500 strongest links between them are visible. List of thresholds and the corresponding smallest included cities and links is provided in appendix C.

When observing the graphs in this section one must bear in mind, that the cities and countries are those of 2012. Thus links within former Soviet Union areas are viewed as international links even in the 1970's.

5.1 Methods

For network visualisation we use Gephi, which is a well known open-source network analysis and visualisation software. In particular we used a third party plug-in GeoLayout, which allows us to place each key to its place according to its coordinates. Neither the plug-in nor Gephi provide map on the background of the network, but a human mind will visualise Europe behind the network graph on its own due to high density and correct proportional location of cities. Although there are multiple projections available in GeoLayout, we will be using one of the most commonly used projection, the Mercator projection, to guarantee the visualisation effect.

5.2 Collaboration network in 1973-1977

As we can detect in the figure 13a, the network is very focused in central Europe in 1973. The most dense networks are in the area of United Kingdom, France, Germany and Italy. There are many cluster visible, clearest of which are French, English, Italian, Soviet Union, Nordic and Germany clusters. The connections between former nations of Soviet Union are clearly visible. We can observe that that the collaboration between countries is rather common deducing from figure 13b. We can observe network within countries in figure 13c. As it provides no new information in comparison with the whole network (due to low number of international links), we do not provide domestic network graphs for other years of this chapter.

In the year 1977 London and Paris have grown their influence, as we can be observed from figure 14a. We can also see division between West and East Germany. Belgium has risen to the map and can be seen as part of the Benelux cluster. The Nordic countries have also merged into one clear cluster. There is a strong Germany speaking cluster around West Germany. International links are mainly revolving around London, Paris and Moscow. There are fewer international links visible in figure 14b than in figure 13b.

5.3 Collaboration network in 1998-2010

Figure 15a represents the collaboration network in the year 1998. After the gap in the data in 1977-1997, there are few major changes. The most significant changes is the break of the Soviet Union bringing Berlin, Moscow and other cities of the Soviet Union to the same cluster with the middle Europe. In addition, Spain and Turkey have emerged and are now clearly visible. Moreover, Switzerland has formed a new cluster and Ireland has become part of the cluster of England. There are eight major clusters visible: Nordic, United Kingdom, Benelux, Russian&Germany, France, Spain, Italy and Turkey.

International network is represented in figure 15b. We can determine that the western countries collaborate strongly whereas the Middle Europe is more scattered. Germany and Russia form their own collaboration cluster with many Middle European cities and some Nordic countries. Sweden, Finland and Denmark form their own collaboration cluster. By comparing the thickness of the links we can determine, that domestic links are stronger than international links. This claim is supported by results in section 4.2.

The whole collaboration network and international collaboration network for the year 2010 can be seen in figures 16a and 16b. There are few major changes in the overall network. The Russian-German collaboration cluster has disintegrated into many clusters. New clusters are the Middle European cluster (with Czech, Austria, Hungary, Croatia, Slovenia and Serbia) and Greece cluster. We can detect, that Romania is also visible on the map. Furthermore, Portuguese cities have emerged and integrated into the Spanish cluster. Also cluster of Switzerland has integrated into the German cluster. When observing the international network we can detect that the German cluster has detached from Russian cluster and formed its own collaboration cluster. In addition, the Scandinavian cluster has been unified.

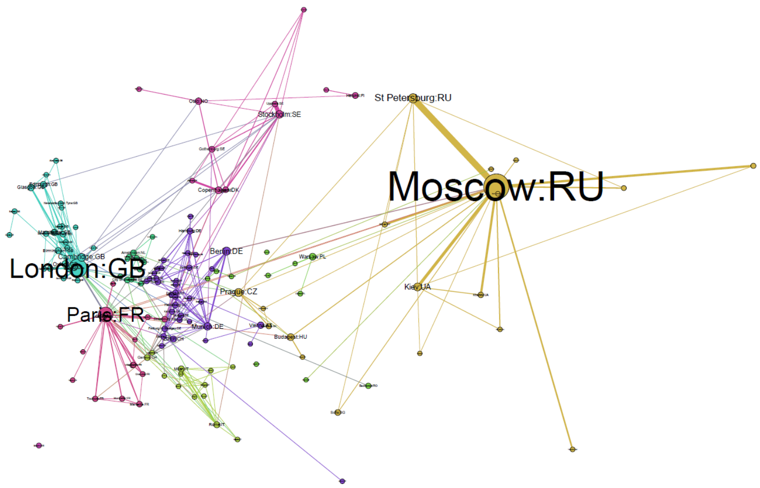
5.4 Overall temporal evolution of the collaboration network

During the 1970s there is no much fluctuation between the years, as we can see in the figures 13a and 14a, but we can determine nodes that belong to different clusters. We can observe that there are eight clusters: The Soviet Union Cluster, the Scandinavian Cluster, the Italian Cluster, the British Cluster, the French Cluster, the Benelux Cluster, the German Cluster and the Polish Cluster. During the 1970s there is little change in the clusters, but in the year 1977 the city of Berlin is absorbed by the Soviet Union Cluster. In the end of this period the boarders of the Soviet Union are clearly visible in the collaboration network.

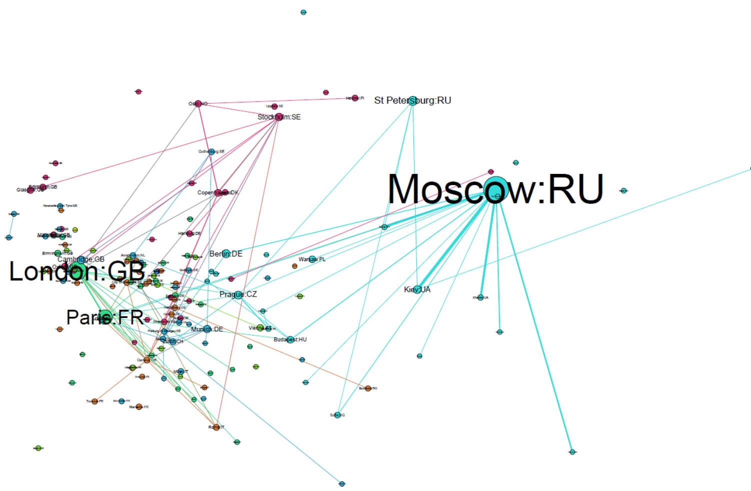
When comparing the collaboration networks in 1970s and in the end of 1990s, the collaboration network seems rather different. Clusters of the Soviet Union and Germany have merged and Switzerland has become its own cluster as seen in figure 15a. There are multiple new clusters in the southern part of Europe. The Spanish-Portuguese cluster is evident, whereas Greece and Turkey have their respective small clusters.

During the 2000s there is some evolution in the overall collaboration network. The most significant changes are the increasing weight of the nodes in the southern Europe (Spain, Greece, Turkey) and decreasing visibility of Russian cities and their collaboration. In addition, there is fluctuation in clusters in the Middle Europe. Austria, Switzerland, Poland and Czech Republic alternate between clusters, sometimes belonging to the Cluster of Germany and sometimes having their own respective clusters or form independent clusters with each other. Eastern countries also merge to Russian cluster at times. Most common combination is that the German speaking nations form their own cluster and Czech Republic forms a cluster with Slovakia, Poland and Russia. Poland is mostly associated with Russian cluster. The fluctuation could be result of the nature of the Louvain method as the partitions may vary.

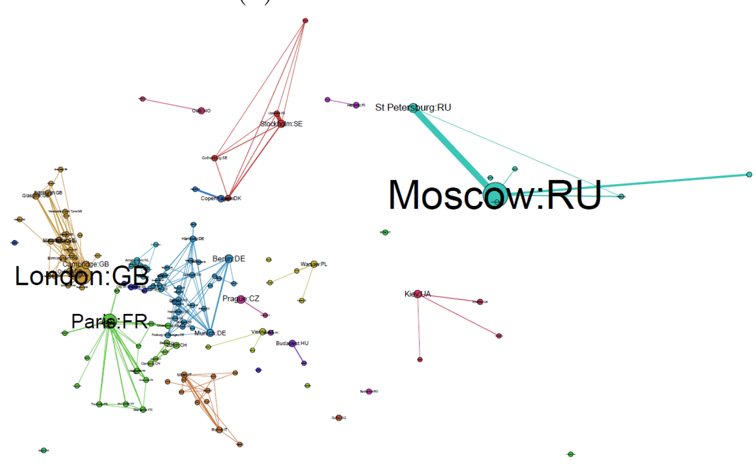
In 2000s there is also an emerging new cluster that includes capitals of nations in the Eastern Europe. The cities are Ljubljana, Zagreb, Budapest, Belgrade, Vienna and Prague. In the year 2010 there are twelve clusters: The British Cluster, the Spanish-Portuguese Cluster, the French Cluster, the Benelux Cluster, the German Cluster, the Scandinavian Cluster, the Russian Cluster (including Poland), the Eastern Europe (as described above), the Greece Cluster (including Sofia), the Italian Cluster, the Romanian Cluster and the Turkish Cluster. These are visible in the figure 16a. We can deduce, that there might be correlation between the nations language and the cluster it belongs to as the clusters seem to be formed mostly around language families.



(a) All Links

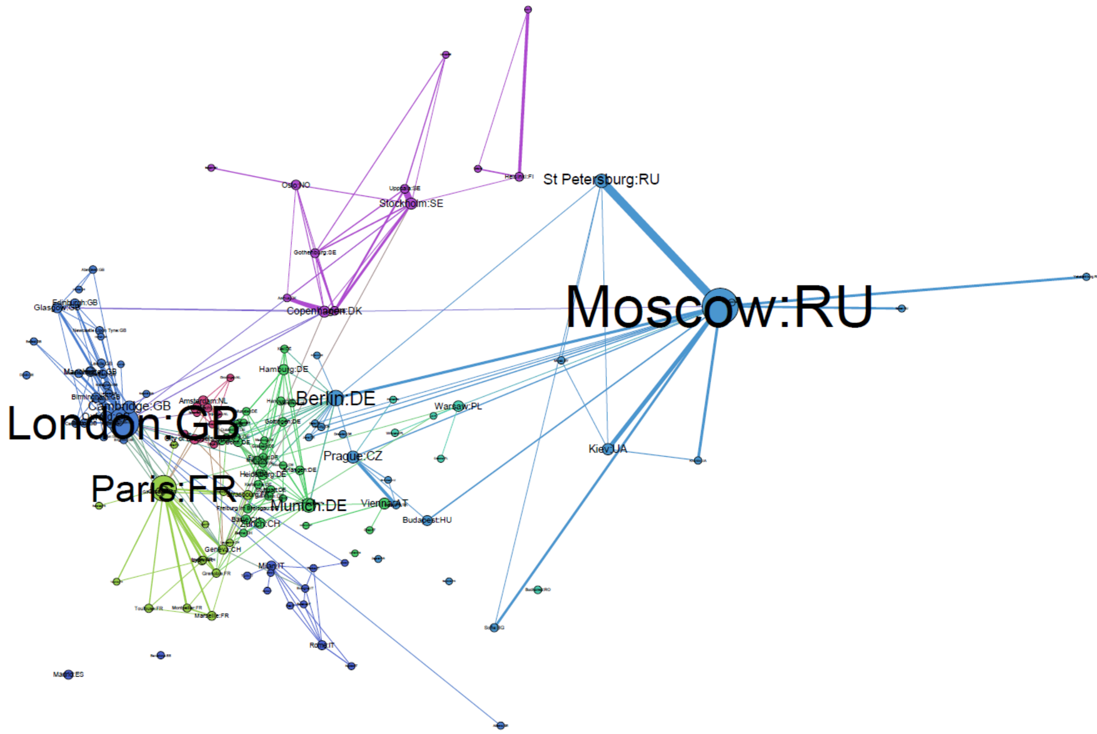


(b) International Links

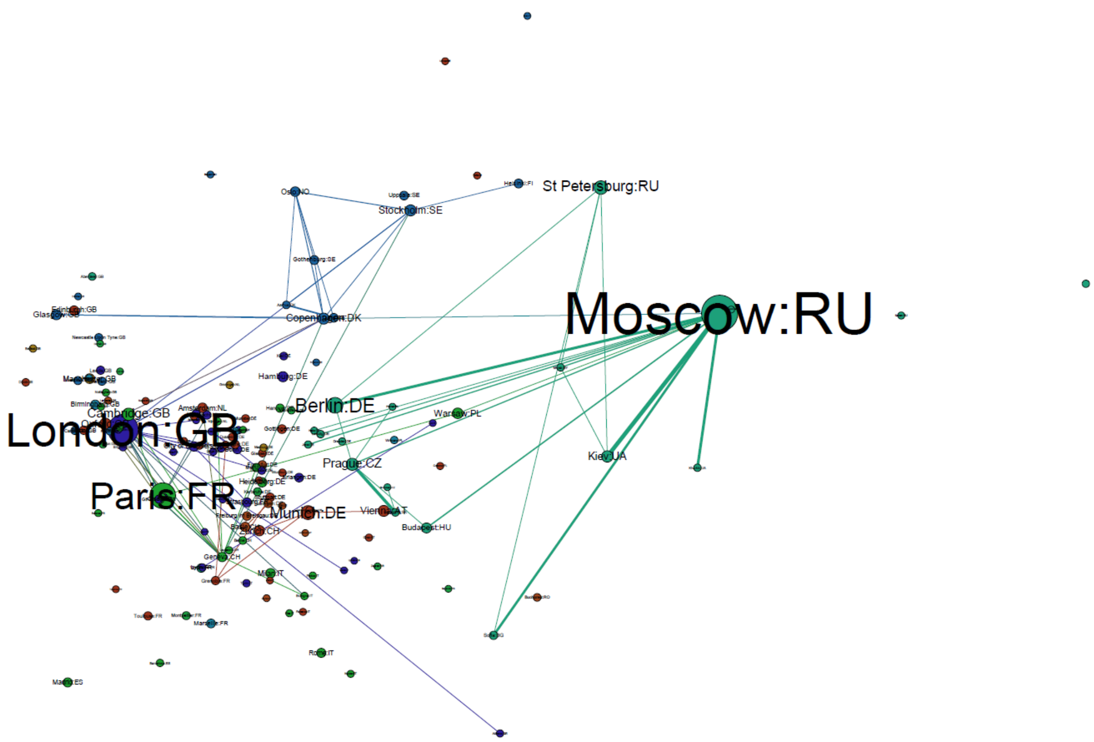


(c) Domestic Links

Figure 13: 150 Biggest Cities and 500 Strongest Links in the Year 1973

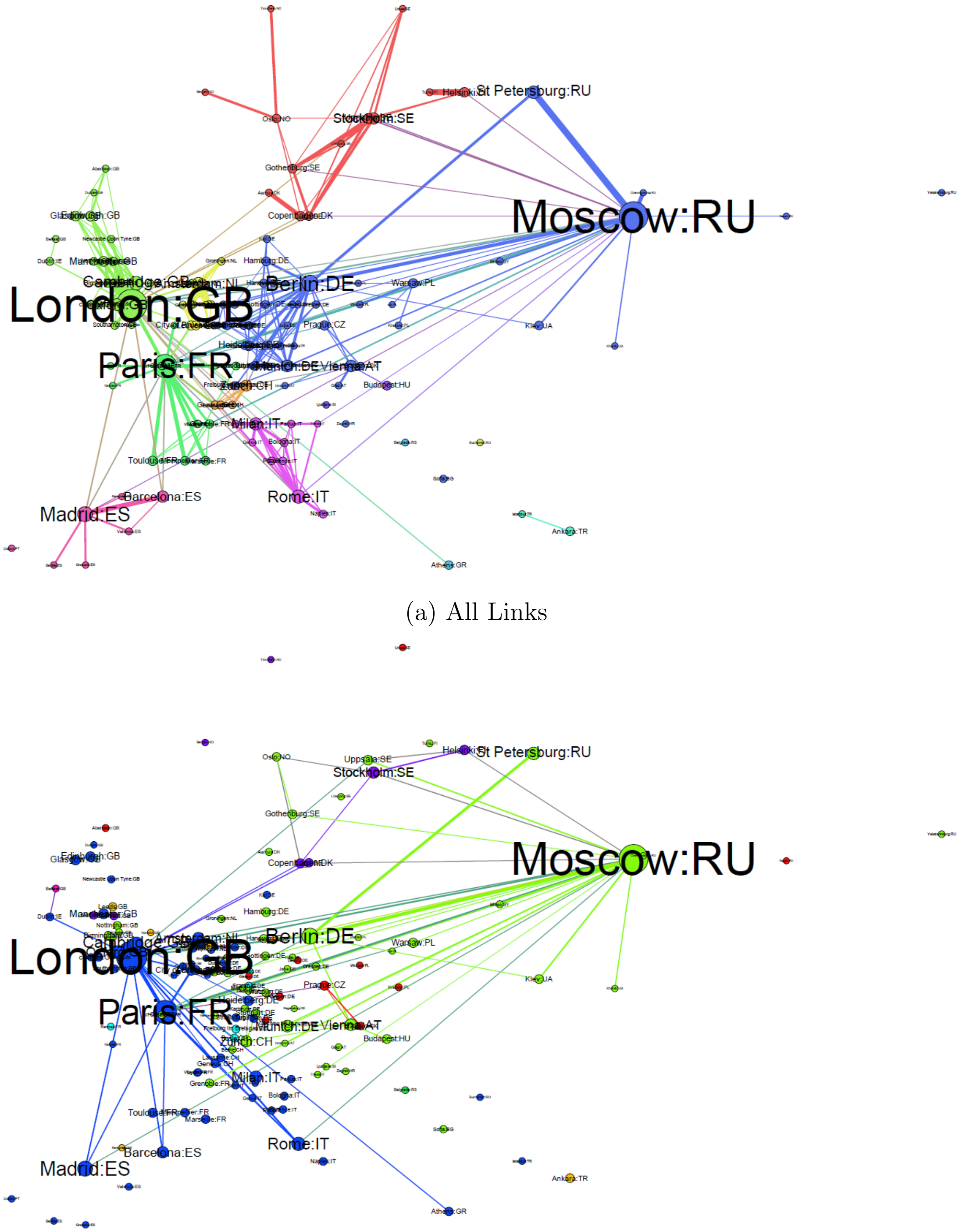


(a) All Links



(b) International Links

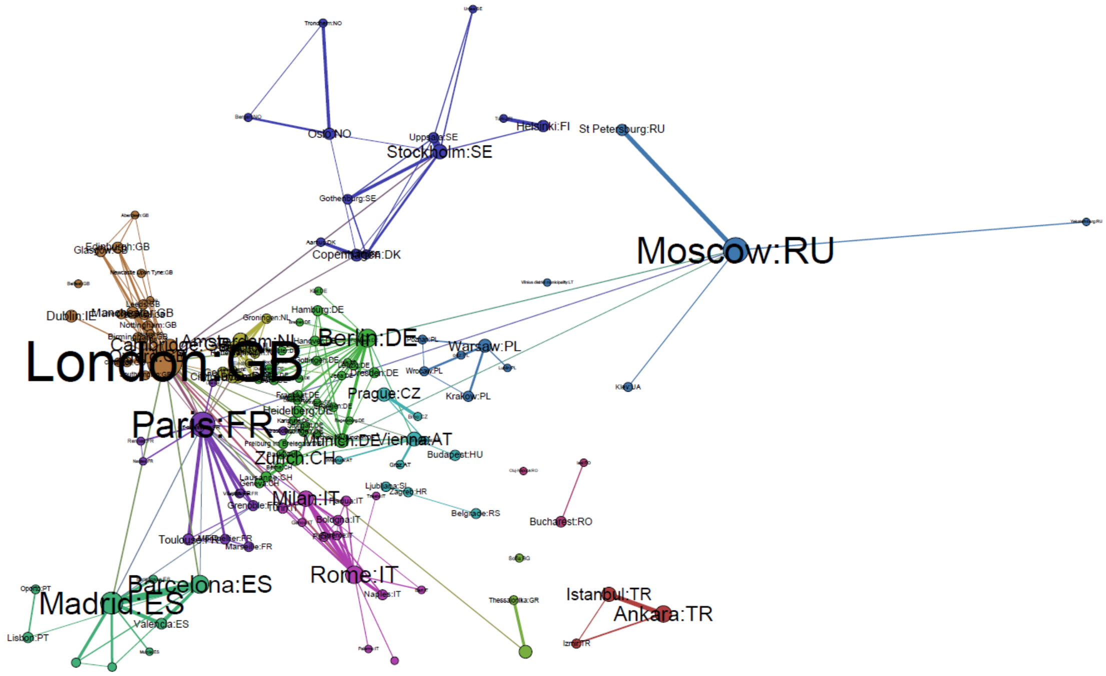
Figure 14: 150 Biggest Cities and 500 Strongest Links in the Year 1977



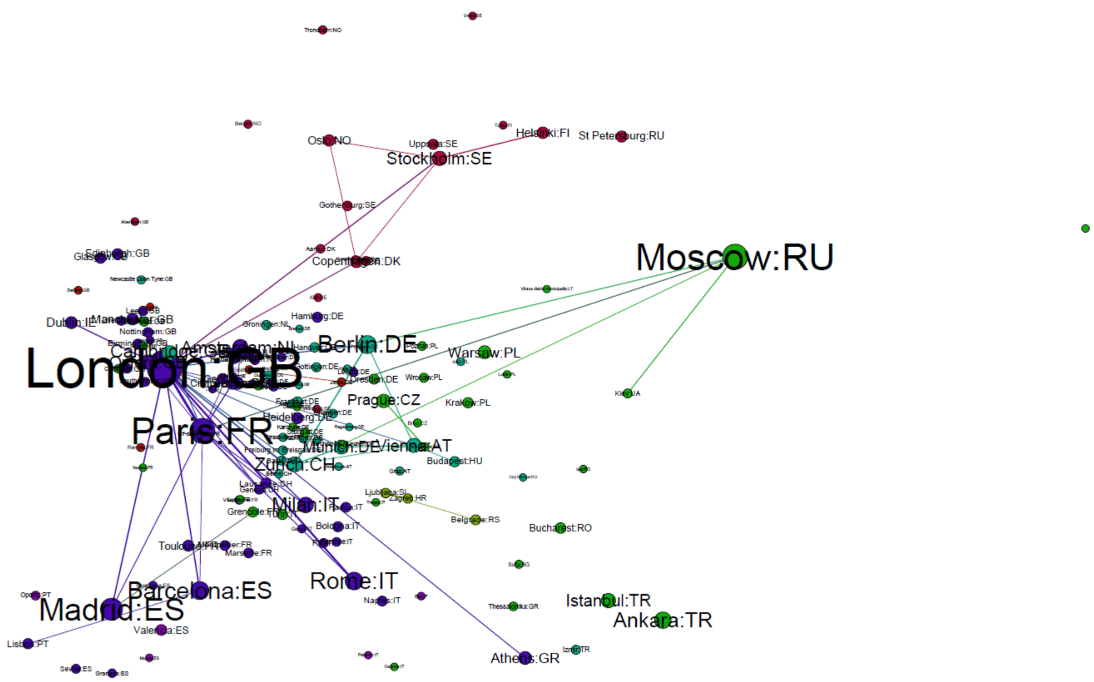
(a) All Links

(b) International Links

Figure 15: 150 Biggest Cities and 500 Strongest Links in the Year 1998



(a) All Links



(b) International Links

Figure 16: 150 Biggest Cities and 500 Strongest Links in the Year 2010

5.5 Temporal evolution of the domestic and international collaboration network

Now that we have the complete picture of the evolution of collaboration at the level of Europe, we can proceed further and inspect collaboration networks within and between countries.

5.5.1 Collaboration networks between countries

In the figures 13b, 14b, 15b and 16b we can see the international networks of countries in Europe. The domestic links have been excluded. These are the cities that are within 150 largest cities and links that are within 500 strongest links in Europe for their respective years. The figures are subgraphs of the full collaboration networks represented in chapters 5.2 and 5.3. The countries are represented in the current formation i.e the post Soviet Union countries are represented as their individual countries even in the 1970s.

We can detect that the collaboration between the Soviet Union countries was tight in the 1970s. This can be explained by the fact, that these links were domestic before the collapse of the Soviet Union. The only other clearly visible collaboration clusters are between London and Paris and in the Scandinavia. Other parts of the network seem ambiguous. In the year 1978 the Soviet Union and Scandinavian clusters remain, but London and Paris form now their own independent collaboration clusters. In addition, we can detect a cluster forming around West Germany.

In the late 1990s, the former Soviet Union countries remain to collaborate firmly with Moscow. The collaboration clusters are the Russian collaboration cluster, the Scandinavian collaboration cluster and the Western collaboration cluster. In addition, we can observe a third cluster that forms around cities in central Europe and that the countries of southern Europe are joining the western collaboration cluster. In the year 2006 Berlin is merged with the cluster of the Western Europe whereas the Russian cluster is diminishing. In the year 2010 the German speaking countries have formed their own cluster. Collaboration between Western and Southern part of Europe has remained strong. The collaboration in western Europe constitutes around Paris and London whereas the eastern part of Europe seems to collaborate mostly with Moscow and other countries in Eastern Europe. The cluster of Scandinavia remains intact throughout the latter period of inspection.

5.5.2 Collaboration networks inside countries

In the figures 13a, 14a, 15a and 16a we can see the inner networks of countries in Europe. We can determine, that there is little variation inside countries, but the countries visible are changing. The clusters are separate since there are no connecting links due to domestic nature of the links. Once again the most notable change is the ascending of southern Europe in the 1990s.

5.5.3 Comparison of the evolutions of collaboration networks within and between countries

By comparing overall and international networks, we can determine that there are vast number of domestic links compared to international ones. Therefore we can deduce that there is little collaboration between countries in comparison to domestic collaboration. This can be confirmed by seeing the statistics for the links in table 9. The gradients and absolute values of internal links vastly surpass those of external links. The proportional amount of internal links and external links compared to all links stay virtually the same through the years as can be observed from figure 11. The only exception is the rise of external links with weight of over 5. This leads to a conclusion that the minor collaboration between countries amplified somewhere between 1978 and 1997 but stays now constant in comparison to internal links.

However, the average and median weight of the 50 and 150 strongest links are almost identical to the average and median of the 50 and 150 strongest internal links as we can see in figure 10. There seems to be little growth in 150 strongest international links.

Moreover, in the year 2010 the number of links with weight of over 50 is less than 10 for the international links, whereas for domestic links the respective number is over 100. This is confirmed by table 9. In fact, the year 2010 the strongest international link is the one between London and Paris but it is merely the 59th strongest link in total. In conclusion, the internal links are growing stronger whereas there is only slight growth in the strength of the international links. Nevertheless, the proportion of internal and external links in categories remains virtually constant through the 2000s.

To conclude, the domestic collaboration has remained stronger than the collaboration within Europe. Even the growth rate of domestic collaboration is higher than that of international collaboration. This leads to deduction, that internationalisation is yet to truly shake the science world, although the number of international connections and publications has been growing steadily.

6 Summary

In this study we assessed temporal evolution of the scientific collaboration in Europe. We extracted location information from publication in the ISI database, converted the obtained location information into coordinates throughout the data utilising Google API and chose the appropriate time interval for this study. In the end of the data processing we constructed data set for overall, domestic and international collaboration networks in Europe.

By plotting simple curves of the basic network measures and spanning network graphs, we were able to determine some characteristic features for the collaboration network. Firstly, the overall number of affiliations in the top 15 cities are increasing and the difference between the 1970's and 2000's is substantial. The top 15 cities are growing in terms of publications, but their overall share is decreasing due to immensely ascending number of small cities. Secondly, the dominating cities stay practically the same. London, Paris and Moscow dominate when comparing the number of publications. However, there is fluctuation in the second vicinity. In the 1970's none of the top 15 cities were from Southern Europe, whereas in the 2000's cities from Italy, Spain and Turkey have emerged to the top 15 cities. Thirdly, domestic collaboration vastly surpasses the international collaboration both in strength and in growth. This is confirmed by the network graphs. However, the number of small international links is increasing stadily. Finally, we observed network graphs that indicated clear clusters that seemed to form around language families and established national collaboration relationships. International collaboration clusters seem to revolve around Paris&London, Berlin, Moscow and the Scandinavian capitals. The rise of the southern countries and diminishing influence of Russia was also evident in all of the graphs. This could be linked to disintegration of the Soviet Union.

References

[1] Alain Barrat, Marc Barthelemy, Romualdo Pastor-Satorras, Alessandro Vespignani.
The architecture of complex weighted networks
Proc. Natl. Acad. Sci. USA 101, 3747, 2004.

[2] Aaron Clauset, M. E. J. Newman, Cristopher Moore.
Finding community structure in very large networks.
Phys. Rev. E 70, 066111, 2004.

[3] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre.
Fast unfolding of communities in large networks.
J. Stat. Mech. P10008, 2008.

[4] Community Detection Method in Gephi
https://forum.gephi.org/viewtopic.php?t=2856
9.3.2015

[5] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, Clifford Stein.
Greedy Algorithms.
Introduction to Algorithms, Chapter 16, 2001.

[6] ISI Indices http://wokinfo.com/media/pdf/cit-connect-flyer.pdf
10.3.2015

[7] ISO, ISO 3166-1 Alpha-2
http://www.iso.org/iso/country_codes.htm
10.3.2015

[8] ISO, ISO 3166-1 Alpha-2 Country Codes
https://www.iso.org/obp/ui/#search
9.3.2015

[9] EU, Country Code of Kosovo
https://geonames.wordpress.com/2010/03/08/xk-country-code-for-kosovo/
http://ec.europa.eu/budget/contracts_grants/info_contracts/inforeuro/inforeuro_en.cfm
9.3.2015

[10] Member States of the Council of Europe
http://www.coe.int/en/web/about-us/our-member-states 10.3.2015

[11] Klement Tockner, Urs Uehlinger, Christopher T. Robinson.
Ural River Basin.
Rivers of Europe, Chapter 18, Academic Press, ISBN 9780123694492.

A Countries Included in the Study

Country	ISO-code
Albania	AL
Andorra	AD
Armenia	AM
Austria	AT
Azerbaijan	AZ
Belarus	BY
Belgium	BE
Bosnia and Herzegovina	BA
Bulgaria	BG
Croatia	HR
Cyprus	CY
Czech Republic	CZ
Denmark	DK
Estonia	EE
Finland	FI
France	FR
Georgia	GE
Germany	DE
Greece	GR
Hungary	HU
Iceland	IS
Ireland	IE
Italy	IT
Kosovo	XK
Latvia	LV
Liechtenstein	LI
Lithuania	LT
Luxembourg	LU
Macedonia	MK
Malta	MT
Moldova	MD
Monaco	MC
Montenegro	ME
Netherlands	NL
Norway	NO
Poland	PL
Portugal	PT
Romania	RO
Russia	RU
San Marino	SM
Serbia	RS
Slovakia	SK
Slovenia	SI
Spain	ES
Sweden	SE
Switzerland	CH
Turkey	TR
Ukraine	UA
United Kingdom	GB
Vatican City	VT

Table 13: Countries included in this study and their ISO-codes[8]

B The 150 Strongest Links in Europe in the Year 2010

City 1	City 2	Weight	City 1	City 2	Weight
Cambridge:GB	London:GB	276,3296524	Creteil:FR	Paris:FR	73,41875876
Amsterdam:NL	Utrecht:NL	261,030929	Lund:SE	Stockholm:SE	72,69697036
Barcelona:ES	Madrid:ES	254,0795824	DE:DE	Munich:DE	72,39224679
Stockholm:SE	Uppsala:SE	250,3986703	Edinburgh:GB	London:GB	72,22271014
London:GB	Oxford:GB	227,9095319	Gif-sur-Yvette:FR	Orsay:FR	71,90710671
City of Brussels:BE	Leuven:BE	159,8611331	Krakow:PL	Warsaw:PL	71,78167961
Amsterdam:NL	Rotterdam:NL	156,3300144	Paris:FR	Villejuif:FR	71,0908439
Ankara:TR	Istanbul:TR	156,2000492	Graz:AT	Vienna:AT	70,39156344
Amsterdam:NL	Leiden:NL	155,7107671	Ljubljana:SI	Maribor:SI	70,37390778
Milan:IT	Rome:IT	155,3735516	Berlin:DE	Leipzig:DE	69,94972859
Moscow:RU	St Petersburg:RU	146,4291197	Helsinki:FI	Oulu:FI	69,14591713
Orsay:FR	Paris:FR	136,103957	Groningen:NL	Utrecht:NL	69,07906681
City of Brussels:BE	Gent:BE	135,4827756	Copenhagen:DK	Kongens Lyngby:DK	68,53753025
Lyon:FR	Villeurbanne:FR	132,3669708	Warsaw:PL	Wrocaw:PL	67,77885827
Helsinki:FI	Turku:FI	127,7456241	Brescia:IT	Milan:IT	67,37064447
Madrid:ES	Valencia:ES	125,4650229	Helsinki:FI	Kuopio:FI	66,02180982
Helsinki:FI	Tampere:FI	118,0033814	Bologna:IT	Milan:IT	64,64568353
Amsterdam:NL	Groningen:NL	114,9759845	Genoa:IT	Milan:IT	64,43585222
Paris:FR	Toulouse:FR	114,2072845	Budapest:HU	Szeged:HU	64,08940781
Gif-sur-Yvette:FR	Paris:FR	113,4205755	Aveiro:PT	Oporto:PT	63,70634921
Gent:BE	Leuven:BE	111,0890855	**London:GB**	**Rome:IT**	**62,50709889**
Oslo:NO	Trondheim:NO	107,1278822	Split:HR	Zagreb:HR	61,62444136
Berlin:DE	Potsdam:DE	106,4161378	Delft:NL	Eindhoven:NL	60,77222222
Brno:CZ	Prague:CZ	106,1924021	Antwerpen:BE	Leuven:BE	60,64249638
Aarhus:DK	Copenhagen:DK	105,9069747	Athens:GR	Patras:GR	60,56577106
Barcelona:ES	Tarragona:ES	104,8553166	Moscow:RU	Nizhny Novgorod:RU	60,49477794
Birmingham:GB	London:GB	103,8978303	Linkoping:SE	Stockholm:SE	59,88240207
Gothenburg:SE	Stockholm:SE	103,4060547	Helsinki:FI	Jyvaskyla:FI	59,84545518
Naples:IT	Rome:IT	102,294332	Lausanne:CH	Zurich:CH	59,71954299
Bristol:GB	London:GB	100,9823227	Almada:PT	Lisbon:PT	58,30769231
London:GB	Manchester:GB	99,72017016	Prague:CZ	Rez:CZ	58,29674567
Berlin:DE	Munich:DE	99,65662735	Lisbon:PT	Oeiras Municipality:PT	58,2879805
London:GB	Southampton:GB	96,57009147	Greater London:GB	London:GB	57,77044586
Grenoble:FR	Paris:FR	95,75285305	Milan:IT	Padua:IT	57,50810658
Barcelona:ES	Valencia:ES	95,10814207	Bron:FR	Lyon:FR	57,32840984
Geneva:CH	Lausanne:CH	94,77046663	Paris:FR	Rennes:FR	57,31217452
Athens:GR	Thessalonika:GR	94,6114558	Berlin:DE	Heidelberg:DE	56,84195277
Marseille:FR	Paris:FR	94,52486676	Antwerpen:BE	City of Brussels:BE	56,83409629
Madrid:ES	Seville:ES	94,48747959	Plzen:CZ	Prague:CZ	56,60396825
Chernogolovka:RU	Moscow:RU	92,18484848	Bilthoven:NL	Utrecht:NL	56,37196845
Amsterdam:NL	Nijmegen:NL	92,03324811	Villigen:CH	Zurich:CH	55,95991119
London:GB	Nottingham:GB	91,28673927	Leeds:GB	London:GB	55,55259699
Berne:CH	Zurich:CH	90,79415334	Utrecht:NL	Wageningen:NL	55,53996623
Copenhagen:DK	Odense:DK	88,27755879	Cork:IE	Dublin:IE	55,53509116
Dolgoprudny:RU	Moscow:RU	88,06666667	Lisbon:PT	Oporto:PT	55,52496977
Basle:CH	Zurich:CH	87,92417265	Berlin:DE	Dresden:DE	55,35113861
Barcelona:ES	Cerdanyola del Valles:ES	87,29854726	Berlin:DE	Hanover:DE	55,29015398
Madrid:ES	Saragossa:ES	86,79979762	Pisa:IT	Rome:IT	55,07361545
Leiden:NL	Utrecht:NL	86,61828986	Ankara:TR	Izmir:TR	55,02462909
Belgrade:RS	Novi Sad:RS	86,0179089	Lille:FR	Villeneuve-d'Ascq:FR	54,32686203
Rotterdam:NL	Utrecht:NL	84,61612008	Grenoble:FR	Saint-Martin-d'Heres:FR	53,55305539
Antwerpen:BE	Gent:BE	84,38741264	Coventry:GB	London:GB	53,36881056
Cambridge:GB	Oxford:GB	84,21489946	Tallinn:EE	Tartu:EE	52,76390812
Dubendorf:CH	Zurich:CH	83,62619048	Leicester:GB	London:GB	52,40984534
Espoo:FI	Helsinki:FI	83,40585467	Bucharest:RO	Magurele:RO	52,16666667
Milan:IT	Pavia:IT	83,34316935	As:NO	Oslo:NO	52,06071429
Leiden:NL	Rotterdam:NL	83,02699016	**Berlin:DE**	**Zurich:CH**	51,82539146
Berlin:DE	Hamburg:DE	83,01307531	Amsterdam:NL	Delft:NL	51,64675334
London:GB	**Paris:FR**	**82,78286144**	Nantes:FR	Paris:FR	51,50526808
Bologna:IT	Rome:IT	82,05394697	Padua:IT	Rome:IT	51,45044968
Montpellier:FR	Paris:FR	81,9248082	Athens:GR	Heraklion:GR	51,31395922
Nijmegen:NL	Utrecht:NL	81,71887647	Milan:IT	Naples:IT	51,24841701
Barcelona:ES	Girona:ES	80,84114032	Lille:FR	Paris:FR	51,08692723
Palaiseau:FR	Paris:FR	79,83860553	Paris:FR	Strasbourg:FR	50,94564581
Granada:ES	Madrid:ES	79,5526789	**Amsterdam:NL**	**London:GB**	50,94034774
Lyon:FR	Paris:FR	78,57850758	Perugia:IT	Rome:IT	50,90929797
Amsterdam:NL	Maastricht:NL	76,62008839	Brighton:GB	London:GB	50,546399
Garching bei Munchen:DE	Munich:DE	75,7979294	Cardiff:GB	London:GB	49,9043702
Florence:IT	Rome:IT	75,61921846	Copenhagen:DK	Frederiksberg:DK	49,82222222
Milan:IT	Turin:IT	75,55143254	**Barcelona:ES**	**London:GB**	49,73173119
Glasgow:GB	London:GB	75,3914505	Santiago de Compostela:ES	Vigo:ES	49,72754103
Lund:SE	Malmo:SE	75,09732476	Oslo:NO	Tromso:NO	49,64294944
Edinburgh:GB	Glasgow:GB	75,01338097	Como:IT	Milan:IT	49,60883321
Bergen:NO	Oslo:NO	73,65067091	Poznan:PL	Warsaw:PL	49,55144435
L'Aquila:IT	Rome:IT	73,49461056	Aalborg:DK	Aarhus:DK	49,38877131

Table 14: The 150 Strongest Links in Europe (2010) (International Links Are Bolded)

C List of Thresholds for Each Year

Year	City	Weight	Rank	City 1	City 2	Weight	Rank
1973	Belgrade:RS	148	150	Aachen:DE	Frankfurt:DE	3,00	582
1974	Malmo:SE	167	151	London:GB	Messina:IT	3,33	511
1975	Barcelona:ES	168	151	Hamburg:DE	Mainz:DE	3,63	500
1976	Belgrade:RS	182	150	Basle:CH	Freiburg im Breisgau:DE	4,00	513
1977	Dijon:FR	194	150	Kiev:UA	Simferopol':UA	3,83	504
1998	Vanduvre-les-Nancy:FR	511	150	Besancon:FR	Paris:FR	15,8	500
1999	Vanduvre-les-Nancy:FR	526	150	Berlin:DE	Kiev:UA	16,3	501
2000	Santiago de Compostela:ES	538	150	Bari:IT	Naples:IT	16,9	500
2001	Loughborough:GB	526	150	London:GB	Zurich:CH	17,3	500
2002	Izmir:TR	533	150	Kuopio:FI	Oulu:FI	17,6	500
2003	Saragossa:ES	557	150	Nottingham:GB	Oxford:GB	19,0	500
2004	Linkoping:SE	589	150	Cagliari:IT	Rome:IT	20,1	500
2005	Minsk:BY	598	150	Madrid:ES	Toledo:ES	21,3	500
2006	Santiago de Compostela:ES	619	150	Almeria:ES	Granada:ES	22,2	500
2007	Catania:IT	635	150	Naples:IT	Turin:IT	24,1	500
2008	Murcia:ES	664	150	Katowice:PL	Krakow:PL	25,1	500
2009	Pavia:IT	663	150	Bristol:GB	Glasgow:GB	26,5	500
2010	Cluj-Napoca:RO	627	150	Orsay:FR	Palaiseau:FR	25,8	500

Table 15: Thresholds for the 150 largest cities and 500 strongest links (1973-2010)

D Kandidaatintyön suomenkielinen tiivistelmä

Tieteen tutkimuksella tarkoitetaan alaa, jossa tutkitaan tiedettä sen omia keinoja hyödyntäen. Tarkastelun kohteena voi olla esimerkiksi tieteen taloudelliset vaikutukset kansantalouteen, sen historia tai tutkimusyhteisön sosiaalisten käyttäytymismallien analysointi sosiologian keinoin. Tavoitteena on lisätä ymmärrystä tieteestä ja sen tekemisestä. Uutta tietämystä voidaan hyödyntää monella tavalla. Tutkija voi esimerkiksi analysoida oman työnsä vaikuttavuutta tai arvioida, kuinka hyödyntää olemassa olevaa yhteistyöverkostoaan parhaiten. Päätöksentekijöille ja poliitikoille tämä tieto luo työkaluja päätösten vaikutusten arviointiin mahdollistaen tietopohjaisen päätöksenteon.

Eräs luonteva tapa kuvata tiedettä on ajatella sitä sosiaalisena verkostona, joka koostuu eri toimijoista ja heidän välisistä yhteistyösuhteistaan. Tällaisten rakenteiden visualisoinnissa apukeinona käytetään verkkograafeja, joista verkoston rakenteen hahmottaminen on helppoa. Tällainen lähestymistapa mahdollistaa myös verkkoteoriassa kehitettyjen matemaattisten työkalujen hyödyntämisen.

Tässä kandidaatintyössä visualisoitiin ja analysoitiin tieteessä olevia yhteistyösuhteita verkkoteorian keinoin. Työssä tutkittiin, kuinka eri maiden ja kaupunkien väliset verkostot ja vaikuttavuus ovat kehittyneet, miten kaupunkien välinen yhteistyö on muodostunut maiden sisällä ja eri maiden kaupunkien välillä, ovatko viime aikoina kehittyneet IT-ratkaisut edistäneet kansainvälistymistä ja kuinka yhteistyöverkossa esiintyvät keskittymät ovat muodostuneet. Lähtötietoina käytettiin julkaisutietokantaa, jota ylläpitää Institute for Scientific Information (ISI). Työssä hyödynnettiin myös verkkoteorian tarjoamia matemaattisia työkaluja sekä Googlen kattavaa paikkatietokantaa sekä sen tarjoamaa ohjelmointirajapintaa. Analyysissä keskityttiin vain tieteellisiin artikkeleihin, joissa on ollut mukana kirjoittajia eurooppalaisista tutkimuslaitoksista. Tällainen lähestymistapa mahdollisti Euroopasta ja sen historiasta sekä kulttuurista jo valmiiksi omaksutun tiedon hyödyntämisen. Nämä alkutiedot ja rajaukset tarjosivat hyvät lähtökohdat analyysille tieteellisen yhteistyöverkoston kehityksestä Euroopassa.

Työn tutkimusaineistona käytettiin Institute for Scientific Informationin julkaisemaa noin 37 miljoonan tieteellisen julkaisun tietovarantoa. Jokaisesta kirjoituksesta on haettavissa esimerkiksi tunnistenumero, tyyppi, julkaisuaika ja sivujen määrä. Tämän työn kannalta olennaisimmat tiedot liittyivät tutkijoihin, joista olennaisimpia olivat kirjoittajien osoitteet. Niitä hyödyntäen julkaisu voitiin sijoittaa kartalle ja osoittaa tietyssä kaupungissa laadituksi. Koska vanhimmat artikkelit on kirjoitettu maissa, joita ei enää ole ja tietokannassa olevissa osoitetiedoissa oli kirjoitusvirheitä, kaikkia sijainteja ei saatu yhdenmukaistettua ja muutettua koordinaateiksi Googlen ohjelmointirajapinnan avulla. Tunnistamattomat osoitteet yhtenäistettiin manuaalisesti ja lopputuloksena yli 99 prosenttia sijainneista saatiin harmonisoitua. Tutkimalla julkaisujen paikkatietojen saatavuutta huomattiin, ettei kaikilta vuosilta ole saatavilla riittävän monen kirjoituksen sijainteja. Ilman osoitetietoja julkaisua on mahdotonta määrittää jossakin kaupungissa tehdyksi ja näin tilastot vääristyvät. Lisäksi artikkeleiden suhteellinen osuus tieteellisistä kirjoituksista on erittäin suuri. Näin työ rajattiin koskemaan ainoastaan vuosina 1973 - 1977 ja 1998 - 2010 julkaistuja artikkeleita, joiden kirjoittamisessa on ollut mukana eurooppalaisia tutkimuslaitoksia. Näinä vuosina kirjoittajien osoitetiedoista oli saatavilla yli 85 prosenttissa julkaisuista.

Kaupunkien saamat painoarvot laskettiin seuraavasti: jos artikkelin laati neljä kirjoittajaa neljästä eri kaupungista, sai jokainen sijainti painon $\frac{1}{4}$. Yhteistyösuhteiden painot lasketaan samalla periaatteella, ottaen huomioon sisäiset linkit. Näin esimerkkitapauksessa jokainen linkki sai painoarvon $\frac{1}{10}$. Eri mantereilla sijaitsevien kaupunkien yhteisartikkeleiden paino jaettiin samaa periaatetta noudattaen, mutta mannerten väliset linkit ja toisilla mantereilla olevat kohteet jätettiin pois analyysistä.

Tutkimus paljasti, että Pariisi, Moskova ja Lontoo ovat painoarvoiltaan suurimmat kaupungit molemmilla tarkasteluajanjaksoilla. Vuosina 1973 - 1977 viidentoista painoltaan suurimman kohteen joukossa oli paljon keskieurooppalaisia kaupunkeja. Jälkimmäisen tarkasteluajanjakson aikana puolestaan monet eteläeurooppalaiset kohteet nousivat viidentoista suurimman joukkoon. Venäläisten kaupunkien paino puolestaan laski. Vaikka eri kohteiden painoarvot keskimäärin nousivatkin, laski suurimpien kaupunkien osuus koko verkoston painosta. Tämä kertoi siitä, että vähäistä tutkimustyötä tekevien kaupunkien lukumäärä on lisääntynyt merkittävästi.

Kaupunkien välistä yhteistyötä tutkittaessa ilmeni, että kaikista vahvimmat yhteistyösuhteet ovat samassa maassa olevien kohteiden välillä. Eri maissa olevien kaupunkien välisten linkkien määrä oli pieni ja yhteistyösuhteet heikkoja. Tästä kertoo se, että 150 vahvimmasta yhteistyösuhteesta vain viisi oli kansainvälisiä vuonna 2010. Lisäksi maiden rajojen sisällä olevat yhteydet voimistuivat nopeammin ja niiden määrä kasvoi voimakkaammin kuin kansainvälisten yhteyksien. Näin voidaan päätellä, ettei IT-teknologian kehityksellä ole ollut ainakaan vielä merkittävää vaikutusta kansainvälisen yhteistyöhön.

Lopuksi tutkittiin Euroopan tieteellisen yhteistyöverkoston keskittymiä ja niiden kehitystä. Tieheymät tunnistettiin Louvainin menetelmällä. 1970-luvulla havaittiin kahdeksan yhteistyökeskittymää. Tiivistymissä oli selvästi havaittavissa Neuvostoliiton vaikutus. Tultaessa 2000-luvulle muutos 1970-lukuun verrattuna oli huomattava ja suurimpana erona oli Neuvostoliiton yhteistyötiheymän yhdistyminen Keski-Euroopan keskittymään. Myös Etelä-Euroopan kehittyminen näkyi selvästi verkkograafeissa. Vuonna 2010 yhteistyökeskittymiä oli 12. Tiiviit yhteistyösuhteet vaikuttivat muodostuvan ennen kaikkea vakiintuneen yhteistoiminnan ja kieliryhmien myötävaikutuksesta. Tästä kertovat esimerkiksi yhteistyökeskittymät Skandinaviassa, Benelux-maissa sekä saksankielisissä maissa.

Tässä työssä tutkittiin tieteellisen yhteistyöverkoston kehittymistä Euroopassa. Lähtöaineistona käytettiin Institute for Scientific Informationin julkaisemaa noin 37 miljoonan tieteellisen julkaisun tietovarantoa. Tietokannasta kerätyt osoitteet yhdenmukaistettiin ja ajantasaistettiin hyödyntämällä Googlen ohjelmointirajapintaa.

Samalla kerättiin kaikkien sijaintien koordinaatit karttavisualisointia varten. Tutkimusaineiston osoitetietojen puutteiden vuoksi analyysi kohdennettiin vain vuosille 1973 - 1977 ja 1998 - 2010. Painoarvoiltaan suurimmat kaupungit olivat Pariisi, Moskova ja Lontoo koko tarkasteluajanjakson ajan. Keski-Euroopan kaupungit hallitsivat 1970-luvun tilastoja, mutta joutuivat väistymään Etelä-Euroopan kaupunkien kasvaessa 2000-luvulla. Suurimpien kaupunkien painoarvot kasvoivat merkittävästi tarkasteluajanjakson aikana. Samaan aikaan merkittävimpien sijaintien suhteellinen osuus kaikkien artikkeleiden yhteenlasketusta painoista pieneni vähäistä tutkimusta tekevien kaupunkien lukumäärän lisääntyessä. Yhteistyösuhteet kaupunkien välillä painottuivat yhä voimakkaasti maiden rajojen sisälle. Nämä linkit myös vahvistuivat nopeammin ja niiden määrä kasvoi voimakkaammin kuin kansainvälisten yhteistyösuhteiden. Tarkasteluajanjakson lopuksi Euroopassa oli kaksitoista yhteistyöklikkiä, jotka vaikuttivat muodostuneen ennen kaikkea vakiintuneiden maiden välisten suhteiden ja kieliryhmien myötävaikutuksesta.