



Aalto-yliopisto
Perustieteiden
korkeakoulu

Temporal Evolution of Citation and Collaboration Network (Introduction)

Jori Jämsä

17.09.2012

Instructor : Ph.D Raj Kumar Pan

Supervisor: Prof. Harri Ehtamo

Work is allowed to be saved and published in Aalto University's public WWW-sites. All other rights reserved.

Background

- Collaboration and citations in scientific research can be presented as networks.
- These networks evolve over time.
 - Author, paper, institute and country level.
- Citation in scientific fields is closely related to the collaboration
- Many variables are known to affect collaboration between research groups
 - Geographical proximity, economic constrains..

Goal of this thesis is to study..

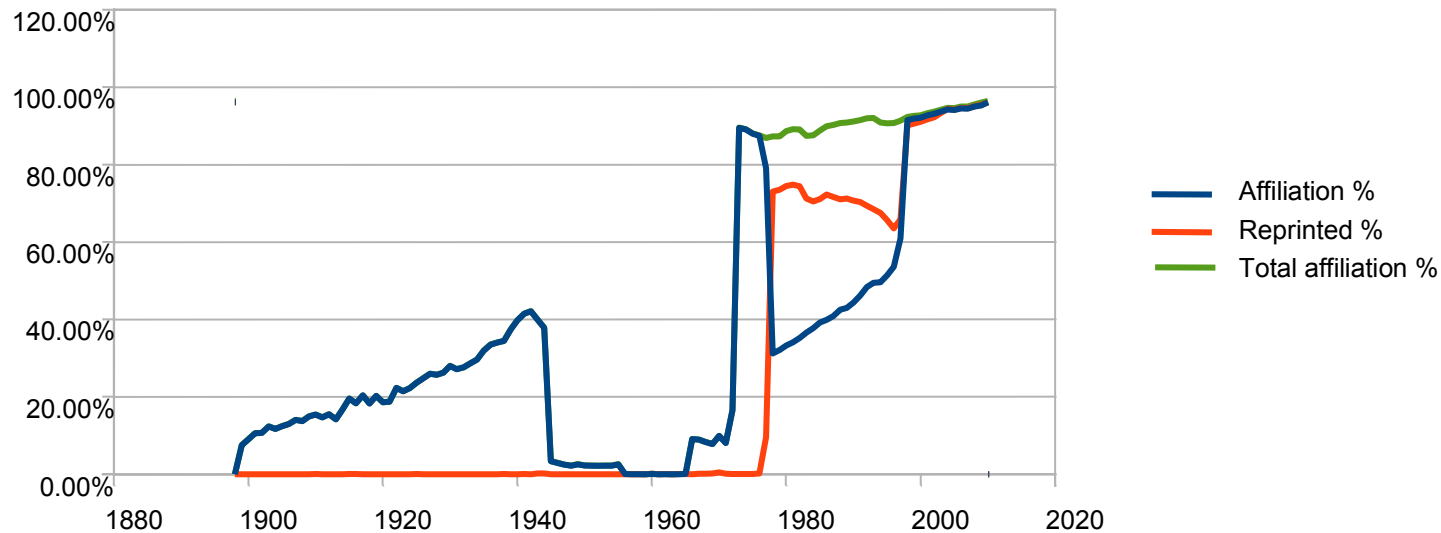
- How different countries have done over the years in terms of publications and citations.
- How geographical proximity affects the collaboration between groups.
- How the collaboration network has evolved in Europe since 1970's.
- Whether the advancements in communication and transportation technologies have affected the evolution.
- Have socio-political aspects effect on development.

Database

- Publications in the databases of Institute for Scientific Information (ISI) Web of Science
 - Science Citation Index Expanded (from over 6650 scientific and technical journals)
 - Social Sciences Citation Index (from 2474 journals)
 - Arts & Humanities Citation Index (from 1395 journals)
- First publications are from the year 1898
- For each publication we have over 30 recorded attributes
 - Name of the author, publication year..
 - The information set is incomplete for some publications

Outlines (1)

- It seems that the data lacks most of the affiliation information before 1970's.
- Some publications don't have the affiliation information or have only reprinted address even after 1970's.
 - Reprinted address includes only the affiliation of the corresponding author.



Outlines (2)

- Complete analysis can be carried out only for years 1973-1977 and 1998-2010.
 - For the years 1978-1997 the analysis can give us a clue of the evolutionary events during that time.
- The analysis is focused only on Europe
- Only cities with enough publications will be included

Disambiguation of the Data

- Some of the affiliations are misspelled:
 - Eespo → Espoo
- Data includes old locations
 - e.g old TKK can be interpreted to a city or other regional area

Methods

- Geocoding:
 - Determining geographical location for each publication from the database
 - Using Wikipedia API and Google API
 - Location (lat,lng) for each affiliation
 - Merging misspelled affiliations
 - Lots and lots of Python..
- Network and graph construction from the reference information
- Complex network analysis and conclusions

Timetable

- Introduction of the thesis 17.09.
- Both analysis and graphs complete 27.10.
- First version of the thesis ready 19.11.
- Thesis complete 22.12.
- Introduction of the results 21.1.