A note on the collaboration between Chalmers University of Technology and University of Gothenburg during the period 2000-2014.

Bo Jarneving

Digital Services

Gothenburg University Library

Published 2015-11-30

Introduction

Chalmers University of Technology and University of Gothenburg , located in close geographic proximity of each other, have over a longer period of time developed collaboration in many areas and in many ways, which is, amongst other things, reflected by joint departments and research centers as well as joint scientific publishing. In this study special attention is devoted to patterns emanating from such publishing efforts. The methods of choice are bibliometric which implies research questions operationalized by quantitative indicators and descriptive statistics.

Statement of purpose and research questions

The purpose of this study was to arrive at a basic bibliometric mapping of joint research between Chalmers Technical University and University of Gothenburg during the period 2010-2014. This implies a focus on publishing efforts and their subsequent impact on later research. The following research questions were stated:

- Which sources are frequently applied when publishing research?
- What is the pattern of publication growth?
- Which countries and organizations are the main collaborators?
- Which significant collaborative network structures can be found?
- Which subject areas are frequently published in?
- What is the relation between subject areas and collaborating organizations?
- What is the impact of joint research papers?

It was hoped for that this study would provide with facts and figures that may inspire to the testing of usable hypotheses.

Data and Methods

The data set being analyzed comprised 2 978 bibliographic descriptions of papers classified as articles, review articles, proceedings papers and letters, published during the period 2000-2014 and downloaded 2015-10-29 from the Web of Science Core Collection. This data set was imported to the InCites database where additional data was computed for each participating record. For reasons of reading fluency, it will not incessantly be repeated that all data and all reasoning pertain to the universe of *Web of Science Core Collection* only. Hence, the reader should bear in mind this particular delimitation when statements of a general character are put forward in this report. This is particularly important when considering the fact that some sub-fields and specific research themes may be indexed in the Web of Science databases to a lesser extent.

The methods and statistical techniques applied are standard descriptive and bibliometric methods. No special attention is devoted to these methods and they are presented as they occur in the presentation of findings.

Findings

The presentation of empirical findings begins with the elaboration of sources, that is, the channels of communication in the global journal market that have been chosen by authors. The identification of the more important journals may be used for collection management purposes as well as in the information provision context. Next the growth rate of collaborative papers is computed and compared with the growth rate of each university. The geographical dimension of collaboration is elaborated on next, identifying the more frequent collaborators and mapping the pattern of international collaboration. At a lower level of paper aggregation, collaboration between organizations is detailed from different angles in the subsequent section. The dispersion of papers over subject areas as well the relation between organizations and subject areas are dealt with before a mapping of impact ends the study.

Sources

In this section we will focus on the distribution of papers over sources (journals). Elaborations on the distribution of items over sources is particularly satisfying as an item belongs to exactly one source, hence there is no overlap as in the case of multiple assignments to addresses and research fields for a paper. A total of 1 070 distinct journals produced on the average 2,7 papers during the period of observation and the range was 157. The maximum number of papers was 158. A common measure of the equality or evenness of a distribution is the Gini Index. It ranges between 0-1 where 0 implies perfect equality, that is, all sources contribute the same. Consequently, 1 reflects a perfect inequality which implies that a single source is the only contributor. In this case we arrived at a Gini index of 52 percent. This percentage could be graphically illustrated by a so called Lorenz curve (Figure 1), where perfect equality is illustrated by a straight line. The percentage of the Gini coefficient corresponds to the ratio of the area between the line of equality and the Lorenz curve to the area above the straight line. The curve grows steeply at the beginning, indicating a stronger influence of a few sources and when 50 percent of all sources are cumulated, 82 percent of all items (papers) are found and just before that point, the curve flattens out to a straight line indicating marginal sources. In the context of bibliometric distributions references are often made to the principle of Pareto or the 20:80 ratio. In the current context this ratio would imply that 20 percent of sources generated 80 percent of the output. Empirically many distributions pertaining to the use of publications have been found to adhere to this principle. Here, the deviation from this principle indicates the absence of a core of highly used sources (core journals). In Table 1 the 30 most productive sources are listed.

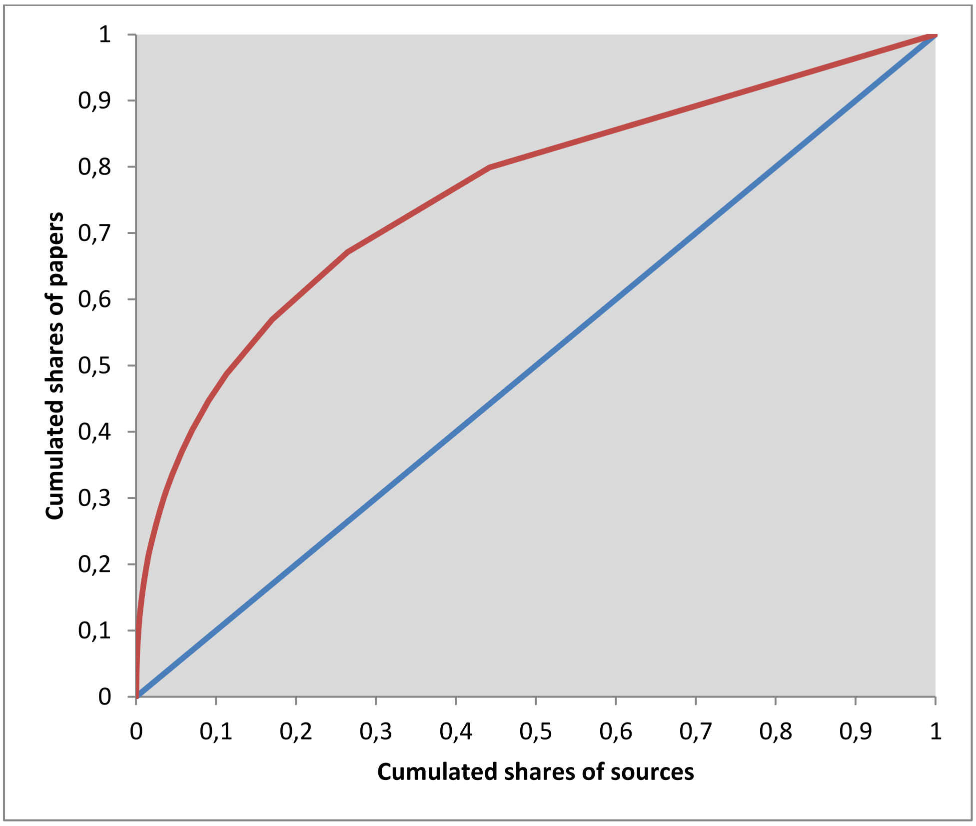

Figure 1. Lorenz curve illustrating the dispersion of papers over sources.

Table 1. The distribution of papers over sources.

Sources	# papers
PHYSICAL REVIEW B	158
PHYSICAL REVIEW LETTERS	78
JOURNAL OF CHEMICAL PHYSICS	53
APPLIED PHYSICS LETTERS	44
JOURNAL OF APPLIED PHYSICS	38
JOURNAL OF HIGH ENERGY PHYSICS	30
ANALYTICAL CHEMISTRY	29
SURFACE SCIENCE	27
PLOS ONE	25
JOURNAL OF FUNCTIONAL ANALYSIS	23
PHYSICA C-SUPERCONDUCTIVITY AND ITS APPLICATIONS	22
ACTA CRYSTALLOGRAPHICA SECTION E-STRUChalmers University of TechnologyRE REPORTS ONLINE	21
NUCLEAR PHYSICS A	21
LOW TEMPERATURE PHYSICS	19
APPLIED PHYSICS A-MATERIALS SCIENCE & PROCESSING	19
PHYSICAL REVIEW E	19
PHYSICAL REVIEW A	16
CHEMICAL PHYSICS LETTERS	15
NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH SECTION B-BEAM INTERACTIONS WITH MATERIALS AND ATOMS	14
JOURNAL OF PHYSICAL CHEMISTRY A	14
PHYSICAL CHEMISTRY CHEMICAL PHYSICS	14
JOURNAL OF BIOLOGICAL CHEMISTRY	13
APPLIED SURFACE SCIENCE	13
JOURNAL OF PHYSICS-CONDENSED MATTER	13
ACM SIGPLAN NOTICES	13
ATMOSPHERIC CHEMISTRY AND PHYSICS	13
JOURNAL OF THE ELECTROCHEMICAL SOCIETY	13
ANNALES DE L INSTITUT FOURIER	12
PHYSICA SCRIPTA	12
SOLID STATE IONICS	12

Growth

For a number of years the number of publications generated in collaboration has grown steadily with approximately four percent per year (Figure 2). At the entrance of the new century a total of 3 271 papers had been generated and a half decade later that number had grown to 6 249 papers. These figures may be related to the *growth rate* of the two collaborating universities: The growth rate for Chalmers University of Technology was 5,9 percent and for University of Gothenburg 4,6 percent (Figure 3). This means that the growth rate for each university exceeds the growth rate of their common subset of collaborative papers.

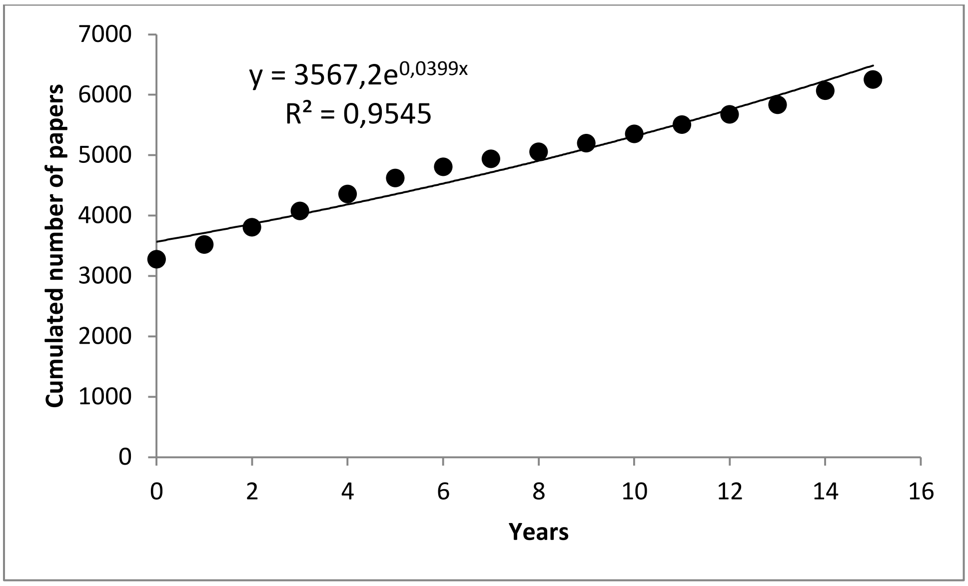

Figure 2. The growth rate of collaborative papers for the period 2000-2015. Papers restricted to articles, reviews, proceedings papers and letters.

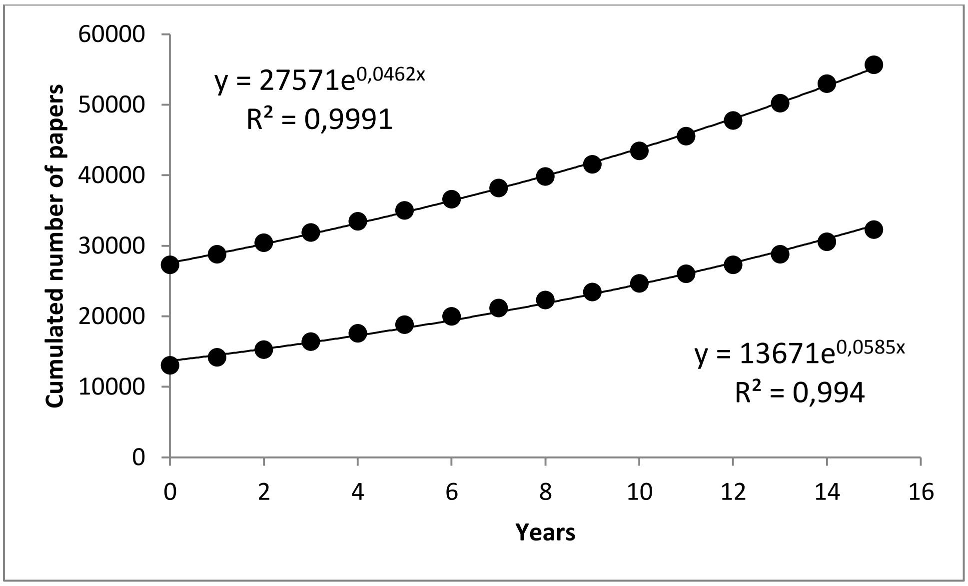

Figure 3. The growth rate of papers for the period 2000-2015 for Chalmers University of Technology (lower curve) respectively University of Gothenburg (upper curve). Papers restricted to articles, reviews, proceedings papers and letters.

Geographical distributions

The distribution of collaborating countries comprised a total of 199 distinct countries. The average number of papers for a country was 1,74, the median 1,0 and the mode 1,0. This implies a positively skewed distribution. Normalizing the standard deviation by the arithmetic mean (σ/μ) the resulting coefficient of variation (CV) was 0,67. This measure, besides facilitating the comparison of distributions on different scales, could also be used as a measure of concentration or inequality. The higher the coefficient the larger inequality of the distribution. As for now, this coefficient is not very informative, but we may use it for later comparisons. In 53 percent of all papers only domestic collaboration takes place, implying that we would find a foreign partner in less than half the set of collaborative papers. Mapping the collaboration on country level, not surprisingly, we find that USA is the most frequent collaborator, followed by Germany, Russia, England, France, Denmark and Norway (Table 2).

Table 2. The distribution of papers over collaborating countries, rank 1 -30.

Countries	# papers
USA	324
GERMANY	233
RUSSIA	178
ENGLAND	149
FRANCE	132
DENMARK	108
NORWAY	90
PEOPLES R CHINA	71
POLAND	65
JAPAN	60
SPAIN	59
ITALY	55
NETHERLANDS	52
UKRAINE	50
SWITZERLAND	44
FINLAND	42
SCOTLAND	38
AUSTRALIA	36
SOUTH KOREA	35
CANADA	33
BELGIUM	31
ISRAEL	28
AUSTRIA	25
CZECH REPUBLIC	22
INDIA	19
SOUTH AFRICA	16
SRI LANKA	13
IRAN	13
SLOVAKIA	12
ROMANIA	12

Comparing this distribution with a corresponding one for Sweden (a total national distribution) would facilitate some interesting comparisons. First, we compare some statistical features. For the national distribution the mean was 1772, the median 74 and the mode 1. Again we see a lopsided distribution with the tail at the right hand side. However, a CV= 2,91 implies a stronger concentration of papers to countries compared with the previous distribution based on the set of collaborative papers. Next, we would like to compare the distributions with regard to correlation. Obviously such a comparison can not take place without some re-ordering of data: first we do not want to include more random occurrences in the comparison, hence we delimit the comparison to the top 30 countries from each distribution. As this may generate two not completely overlapping lists of country names we form the union of these lists. It was found that five countries in total occurred in one list only, hence the total length of the final list of countries was 35 (Table 3). Measuring the rank correlation between distributions we arrive at r_s = 0,768. Though this is a strong positive correlation, there are notable differences. For instance, the rank of *Russia* differs by 12 positions. Also, *Italy*, *Netherlands* and *Finland* have clearly deviating rank positions. Accepting the national Swedish distribution as a baseline we conclude that the collaboration between Chalmers University of Technology and University of Gothenburg implies a set of papers with a quite deviating geographical distribution of collaborating countries.

Table 3. The union of the lists from the total Swedish distribution of papers and the distribution of papers for Chalmers University of Technology /University of Gothenburg.

Countries	Rank Sweden	Rank based on the set of collaborative papers
USA	1	1
GERMANY	2	2
ENGLAND	3	4
FRANCE	4	5
DENMARK	5	6
ITALY	6	12
NETHERLANDS	7	13
FINLAND	8	16
NORWAY	9	7
SPAIN	10	11
CANADA	11	20
PEOPLES R CHINA	12	8
SWITZERLAND	13	15
AUSTRALIA	14	18
RUSSIA	15	3
JAPAN	16	10
BELGIUM	17	21
POLAND	18	9
AUSTRIA	19	23
SCOTLAND	20	17
CZECH REPUBLIC	21	24
GREECE	22	35
HUNGARY	23	40,5
BRAZIL	24	35
PORTUniversity of GothenburgAL	25	42,5
INDIA	26	25
IRELAND	27	38
SOUTH AFRICA	28	26
ISRAEL	29	22
SOUTH KOREA	30	19
SLOVAKIA	30	31
ROMANIA	35	31
IRAN	42	27,5
UKRAINE	45	14
SRI LANKA	78	27,5

In a collaborative network, some countries would constitute hubs connecting partners. Visualizing such a network would facilitate our understanding of the context in which collaboration between Chalmers University of Technology and University of Gothenburg takes place. Multidimensional scaling (MDS) is a technique which generates a spatial representation of such networks, commonly in two dimensions, separating peripheral nodes from central ones and depicting the strength of association between each pair of analyzed objects. The operations used to obtain the spatial representation are complex and there is a multiplicity of computational methods. Though the elaboration of operations and computational procedures are outside the scope, the principle of MDS may still be laid out in a fairly comprehensible way:

Let *A*, *B*, *C*, and *D* be the representations of four objects. Let X be the value of association for *A* and *B*, and Y for *C* and *D*. Ideally, the following conditions should be the rule:

(1) If X = Y, then the distance, between both points in the configuration, representing *A* and *B* should be the same as the distance between the points representing *C* and *D*.

and

(2)) if X < Y, then the distance between both points representing *A* and *B* in the configuration, should be greater than the distance between the points representing *C* and D.

The principle of calculating the strength of collaboration between objects (here countries) also needs to be elaborated on. Given the same number of common papers for pairs of collaborating countries, we consider the collaboration strength inversely related to the frequency of publications for collaborating countries. This means that we consider the collaboration between one country assigned to a small number of papers and another country assigned to a large number of papers more significant than the collaboration between two countries both generating a large number of papers. Without delving into technicalities it suffices to say that we normalize for size when the collaboration strength is calculated.

Having roughly sketched some important principles, we may proceed to the generation of an MDS-map depicting the essential parts of the underlying patterns of collaboration. Selecting those countries that occur in the by line of at least five papers during the period of observation the association between 44 countries was analysed applying MDS (Figure 4). In the centre of the map we can identify the more frequent collaborators from Table 1: *USA*, *Germany*, *Russia*, *England* and *France*. USA is a central node with 38 collaborating countries. This applies also to Germany with an equal number of collaborating countries, though the collaborations are generally weaker. England connects with 35 countries and France, Norway and Poland with 33 countries. In total 12 countries collaborate with more than 30 countries. Only Sweden, obviously, is linked with the maximum 43 links as its central position is the result of data collection bias. In the periphery, connected by less frequent collaboration with the more central countries, we identify smaller or less developed countries like Latvia, Sri Lanka and Bulgaria. We conclude that there is a centre of gravity comprising USA, Russia and the largest European economies.

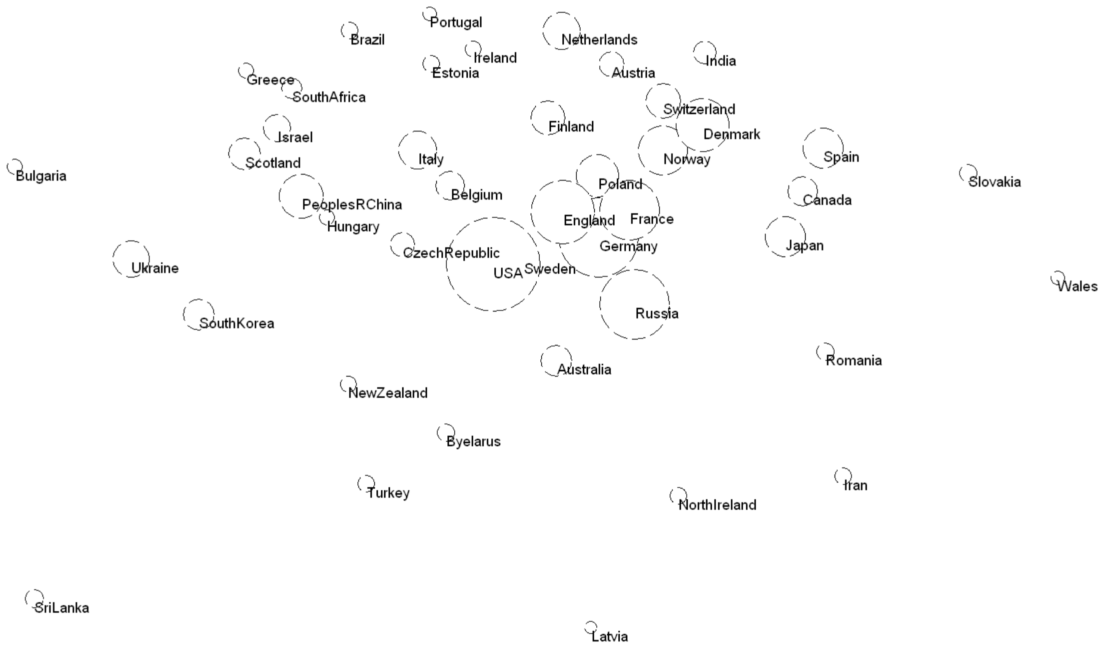

Figure 4. Multidimensional scaling of 44 collaborating countries. Distances on the map are inversely related to the normalized strength of collaboration. The sizes of circles correspond to the number of papers though Sweden is not assigned a circle for reasons of visibility. Kruskal's stress = 17 %.

Collaborating organizations

Having arrived at an understanding of the geographical dimension of collaboration we now focus on collaborating organizations. The methods applied here are essentially the same as in previous section. First we need to know how collaboration is organized in terms of number of produced papers per collaborating organization (Table 4). The first thing that meets the eye is the strong collaboration with Sahlgrenska university hospital, which is explained by the relation between the medical faculty of University of Gothenburg and the university hospital. Also, the strong influence of Russian Academy of Science stands out as an example of a significant foreign collaborator. In the same sense is National Center for Scientific Research (CNRS, France) an influential collaborator. Geographical-cultural influence appears to be an important factor as seven out of ten organizations within ranks 1-10 are Swedish.

Table 4. The distribution of papers over collaborating organizations, rank 1 -30.

Organization	# papers
SAHLGRENSKA UNIVERSITY HOSPITAL	171
RUSSIAN ACADEMY OF SCIENCES	98
UPPSALA UNIVERSITY	93
LUND UNIVERSITY	82
ROYAL INSTITUTE OF TECHNOLOGY	65
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS	65
KAROLINSKA INSTITUTET	63
STOCKHOLM UNIVERSITY	49
LINKOPING UNIVERSITY	47
UNIVERSITY OF OSLO	45
MAX PLANCK SOCIETY	42
AARHUS UNIVERSITY	41
B VERKIN INSTITUTE OF LOW TEMPERATURE PHYSICS	40
UNIVERSITY OF COPENHAGEN	36
UNIVERSITY OF BORAS	34
UNIVERSITY OF CALIFORNIA SYSTEM	33
NATIONAL RESEARCH CENTRE KURCHATOV INSTITUTE	32
IOFFE PHYSICAL TECHNICAL INSTITUTE	32
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS CSIC	32
UMEA UNIVERSITY	31
UNITED STATES DEPARTMENT OF ENERGY DOE	30
CHINESE ACADEMY OF SCIENCES	30
ASTRAZENECA	29
HELMHOLTZ ASSOCIATION	28
POLISH ACADEMY OF SCIENCE	23
NATIONAL ACADEMY OF SCIENCES UKRAINE	21
UNIVERSITY OF TENNESSEE SYSTEM	19
UNIVERSITY OF TENNESSEE KNOXVILLE	19
UNIVERSITY OF NORTH CAROLINA	19
RUTGERS STATE UNIVERSITY	19

Reviewing the frequency distribution of Table 4, we can appreciate that collaboration is not strongly concentrated: A total of 1 176 distinct collaborating organizations were identified. The average number of papers was 3,78, the median 1,0 and the mode 1,0, hence a lopsided distribution with the tail at the right hand. The coefficient of variation was 2,36. Though not immediately comparable, we may still relate this value to the CV for a corresponding national distribution in order to get some point of reference. The corresponding national CV was 23,72, thus we conclude that on the national level, the amalgamation of several larger Swedish universities' collaboration efforts implies more of a concentration of papers to more central producers, as compared with the much smaller selection-biased set of papers under investigation.

Turning our attention to the over-view of collaborative relations in the MDS-map in Figure 5, we can appreciate a center-periphery pattern with a number of major collaborators in the center and more random collaborations mirrored by corresponding nodes in the periphery. The configuration of Chalmers University of Technology, University of Gothenburg, Lund University and Sahlgrenska University Hospital in the center may be expected, but the strong proximity of University of Copenhagen with Chalmers University of Technology and University of Gothenburg less so, confirming that new information is added when normalizing the collaboration strength. Besides an interpretation in terms of center-periphery, the map may also be interpreted in terms of different dimensions in the two-dimensional plane, for instance a geographical-cultural one. On the right side of the map there is an accumulation of Swedish universities (e.g. Umeå University, Stockholm University; Royal Institution of Technology; Uppsala University), whereas the more sparse left side of the map depicts a network of actors predominantly belonging to the eastern hemisphere with organizations from Russia, Ukraine, Korea, Japan and China.

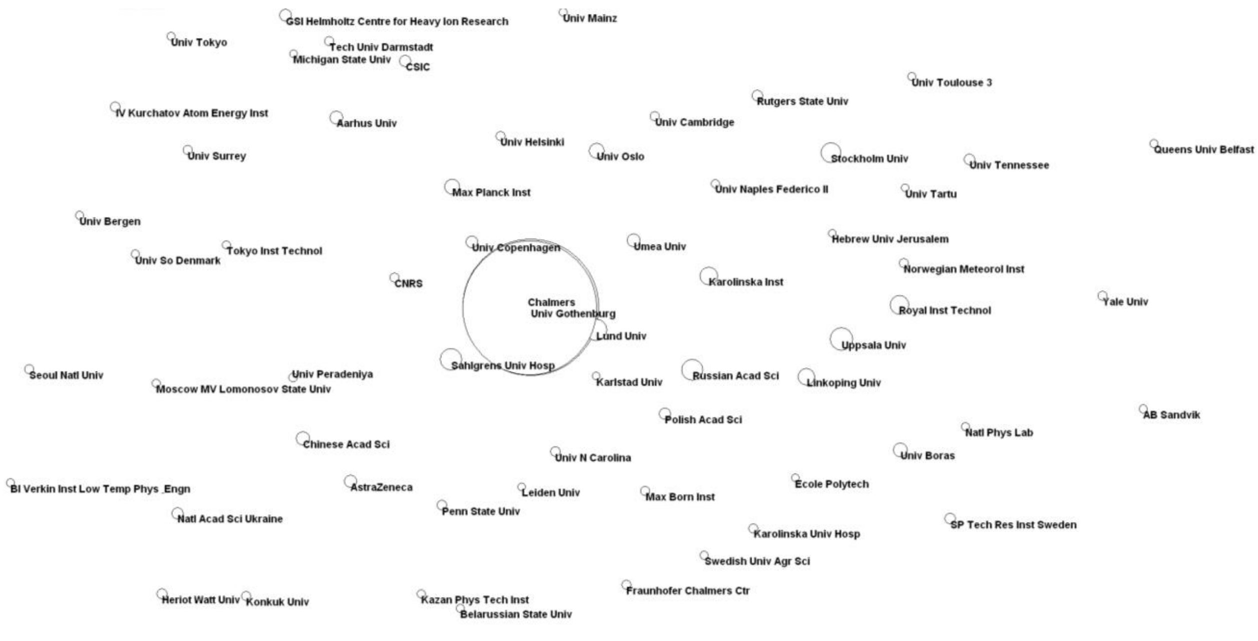

Figure 5. Multidimensional scaling of 63 collaborating organizations occurring at least ten times during the period of observation. Distances on the map are inversely related to the normalized strength of collaboration. The sizes of circles correspond to the number of collaborative papers. Kruskal's stress = 12 %.

In order to zoom in on various aspects of the underlying collaborative network we may complement with some other methods from the field of social network analysis (SNA). First we remove all weaker links of collaboration that have normalized collaboration strength at or below the median. This operation implies a much more sparse network where more random collaboration is filtered out and this has some interesting effects. First of all three nodes (organizations) disappear as they occur only in connection with weaker links of collaboration. For the remaining 60 organizations we compute the number of nodes (neighbors) connected to them. From Table 5 we can conclude that the range of associated neighbors is 1-7 and that almost half of the organizations are assigned to 1-3 links.

Table 5. The distribution of number of collaborative links for 60 organizations.

Number of links	Number of organizations
1	6
2	6
3	17
4	14
5	8
6	1
7	8

From this network of links and nodes we may identify dense (cohesive) sub-networks by extracting so called *k*-cores from the total network where *k* is the number of connected neighbors for a node. A *k*-core associates a node with the *highest* *k*-core in which it appears. From Table 4 we can conclude that eight nodes are assigned to a 7-core, one node to a 6-core, eight nodes to a 5-core and so on. Basically, the identification of *k*-cores implies a possibility of deleting low *k*-cores in order to find densities in the network. Note, however, that a k-core need not be connected as it may occur at different spots in the network, hence a *k*-core need not be a cohesive sub-group itself. [1]

Exploring the network we start by extracting the 7-core from the network and we arrive at a *maximal complete* sub-network, meaning that each node is connected with every other node in the sub-group (Figure 6). This is an important finding. Note that in theory the maximal number of links for a network of size eight is $\frac{N!}{n!(N-n)!} = \frac{8\cdot 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{2(8-2)!} = \frac{40320}{2\cdot 720} = 28$. Based on this we may express the density (*D*) for a network as the quotient of the observed number of links and the maximal number of links. Hence, a maximal complete network has a density *D* = 1. We would consider such a collaboration network significant as it indicates a persevering structure of cooperation during the

[1] V. Batagelj, A. Mrvar: Pajek – Program for Large Network Analysis.
Home page: http://vlado.fmf.uni-lj.si/pub/networks/pajek/
De Nooy et al (2005) Exploratory Social Network Analysis with Pajek.

period of observation.

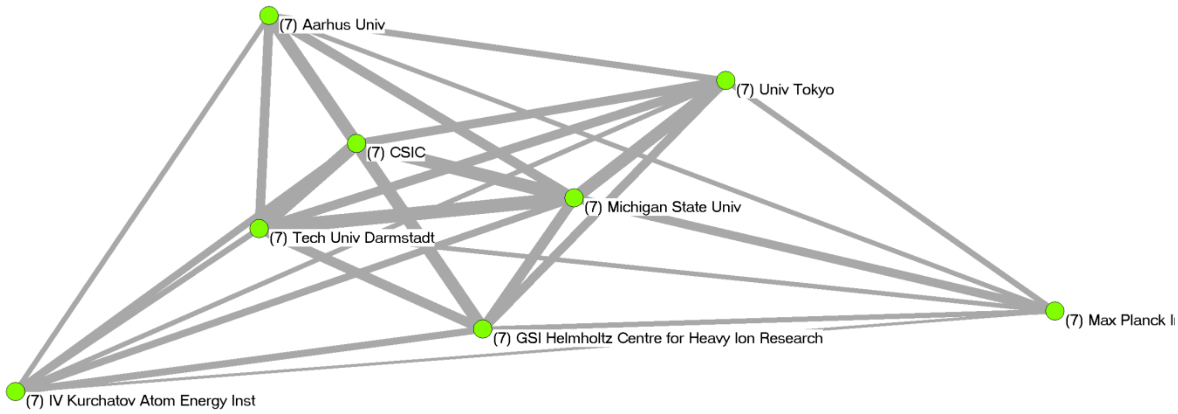

Figure 6. The 7-core extracted from the network: Aarhus University, CSIC (Spanish National Research Council), I. V. Kurchatov Institute of Atomic Energy (Kurchatov Institute), Max Planck Institute, Michigan State University, Technische Universität Darmstadtand GSI Helmholtz Centre for Heavy Ion Research, University of Tokyo. *D* = 1.

Extracting lower k-cores we include both Chalmers University of Technology and University of Gothenburg, and in Figure 7, k-cores with the value of five to six are extracted. In this sub-group we count 16 links for seven nodes, excluding University of Mainz and University of Surrey which connect with nodes elsewhere in the network. This renders a density where D = 16/21 = 0,76 and this sub-group should be considered well connected. Other extractions tried, for instance lower k-cores, were found to provide with little additional analytical value.

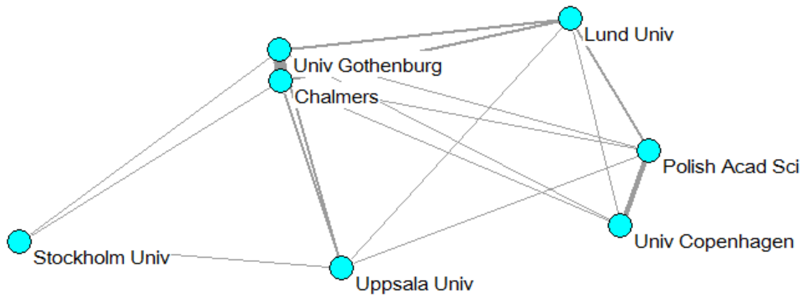

Figure 7. 5-cores and 6-cores extracted from the network: Stockholm University, Uppsala University, Chalmers University of Technology, University of Gothenburg, University of Copenhagen, Polish Academy of Sciences and Lund University forma a cohesive sub-group. *D* = 0,76.

The methods for the exploration of densities in the network are not yet exhausted. Moving the threshold of normalized collaboration strength upwards to the third quartile, we may extract the strongest links over the period of observation. This operation reduces the network to 46 nodes. Furthermore, we require each node to be part of at least one *clique*, that is, a maximal complete sub-network with three nodes. This implies a further reduction to a final 30 nodes. Combining requirements with regard to both the intensity of collaboration and interconnectivity we may thus extract some meaningful structures reflecting the most significant collaboration for the period of observation. Reviewing Figure 8 we can discern three connected graphs. The hub of Chalmers University of Technology/University of Gothenburg connects with nine collaborating universities plus Stockholm University, which in turn form a cohesive sub-group with three other universities. Next, we see a cohesive sub-group of 13 universities and finally at the far right a triad of collaborating universities.

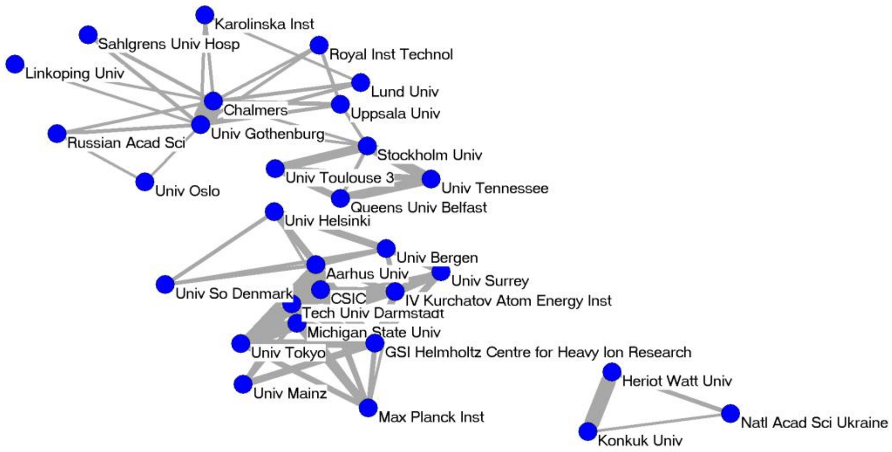

Figure 8. Extraction of sub-network at the third quartile of collaboration strength: in addition, all nodes are required to belong to at least one clique of size 3.

Summing up, the overview of the collaboration network accomplished by multidimensional scaling identified a geographical dimension and a center-periphery structure. Most of the information contained in the underlying data matrix was mirrored in this map. The analytical value, however, was limited and in order to increase our understanding of collaboration patterns, network analytical tools were applied and provided with an elaboration of cohesive sub-networks by varying collaboration strength and interconnectivity. This way, by drilling down into the structure of collaboration, structures otherwise concealed become visible and durable and significant collaboration revealed.

Research areas

The classification of research papers is a neuralgic point as mirrored by several much deviating classification systems in circulation. For the purpose of assigning classes to papers, in this study it comes natural the use the Web of Science categories. It is a quite granular system comprising 249 categories of which 189 were identified in the set of papers being studied. Notably, journals are assigned to categories, not papers. Thus, a paper published in a journal assigned to a particular category automatically obtains the same category. In addition, a journal, hence a paper, may be attributed to more than one category. In Table 6 the 30 most frequent categories are listed along with their frequencies of occurrence. Not surprisingly, mathematical and physical sciences dominate, but we also see environmental sciences and biochemistry in the upper half of the table. Considering the rank ordered listing as whole, we may measure the degree of concentration of papers to categories using the coefficient of variation. Here, we arrive at CV = 1,95 which can be compared with corresponding values for University of Gothenburg and Chalmers University of Technology which were 1,40 respectively 2,04. This indicates that the variation resembles that of Chalmers University of Technology more, which makes sense considering that the intersection of common subject interests should be more influenced by Chalmers University of Technology being the smaller and more specialized university.

Table 6. The distribution of papers over WoS-categories, rank 1-30.

Web of Science Categories	# papers
PHYSICS CONDENSED MATTER	345
MATHEMATICS	299
PHYSICS APPLIED	274
CHEMISTRY PHYSICAL	227
MATERIALS SCIENCE MULTIDISCIPLINARY	216
PHYSICS ATOMIC MOLECULAR CHEMICAL	177
PHYSICS MULTIDISCIPLINARY	161
MATHEMATICS APPLIED	147
BIOCHEMISTRY MOLECULAR BIOLOGY	142
ENVIRONMENTAL SCIENCES	93
STATISTICS PROBABILITY	88
PHYSICS MATHEMATICAL	82
PHYSICS NUCLEAR	80
CHEMISTRY MULTIDISCIPLINARY	80
PHYSICS PARTICLES FIELDS	78
NANOSCIENCE NANOTECHNOLOGY	66
BIOTECHNOLOGY APPLIED MICROBIOLOGY	65
COMPUTER SCIENCE SOFTWARE ENGINEERING	62
MULTIDISCIPLINARY SCIENCES	61
MATHEMATICAL COMPUTATIONAL BIOLOGY	54
NEUROSCIENCES	52
CRYSTALLOGRAPHY	52
CHEMISTRY ANALYTICAL	52
ENGINEERING ELECTRICAL ELECTRONIC	49
ENGINEERING BIOMEDICAL	48
ASTRONOMY ASTROPHYSICS	48
MATERIALS SCIENCE COATINGS FILMS	47
GENETICS HEREDITY	46
COMPUTER SCIENCE THEORY METHODS	46
OPTICS	45

Exploring the connections between research fields (categories) and organizations we aim at the identification of characteristic relations. Selecting those organizations that on the average have at least one paper each year of the period of observation, a total of the 32 more influential organizations were analyzed with regard to their influence on categories. In order to generate a comprehensible presentation we also need to exclude low frequency categories, and only those categories assigned to at least 15 papers produced by the selected 32 organizations were included. Such thresholds are of course quite arbitrary but still needed as there is no theory to base selection decisions on, the aim being to filter out noise in order to discern the signal. A quite dramatic effect of threshold settings was seen: the number of organizations was reduced from 32 to six and the number of categories from 189 to five (Table 7).

Table 7. The distribution of papers over selected organizations and categories.

Organization / category	# papers
Russian Acad Sci	**52**
Physics, Applied	29
Physics, Condensed Matter	23
Uppsala Univ	**27**
Physics, Condensed Matter	27
GSI Helmholtz Centre for Heavy Ion Research	**19**
Physics, Nuclear	19
Stockholm Univ	**17**
Physics, Atomic, Molecular & Chemical	17
Royal Inst Technol	**16**
Physics, Condensed Matter	16
Sahlgrens Univ Hosp	**16**
Neurosciences	16

Impact

The final issue to explore is how papers in the selected set influence later research in terms of being cited. The raw citation count for a paper is insufficient for comparative analyses, thus the raw or observed citation count needs to be normalized. It has become the praxis to normalize with regard to three attributes: (1) subject category, (2) document type and (3) publication year or publication date. The normalization of a paper's citation impact is thus calculated as the quotient of the observed number citations divided by the expected number of citations. When this measure attains the value of 1, the observed citation rate is on a level with the world average.

We begin the analysis of impact by computing the average category normalized citation frequency for each of the publication years (Figure 9). For the whole period of observation, the average category normalized citation frequency was 1,14. Most of the period, the impact is well above the average and at its lowest in 2007. The oscillating curve seen in Figure 9 is characteristic for distributions based on citation averages which are strongly influenced by accidental extreme scores. Thus, the arithmetic mean is in a sense not a good representative of skewed distributions. For this reason another impact indicator was applied – the average percentile. Put simply, the location of a paper's citation score in a global frequency distribution, derived from the proper category, publication year and document type, and sorted descending is identified. Once the set of papers being studied is exhausted, the average is computed for the chosen unit of analysis. Reviewing Figure 9 we can appreciate that the average percentile distribution indeed provides with complementary information as the two curves are not congruent. For instance, the drop of the average category normalized citation frequency in 2007 has no corresponding drop for percentiles. During the period of observation, most of the time the average normalized percentiles are below the median (50 %).

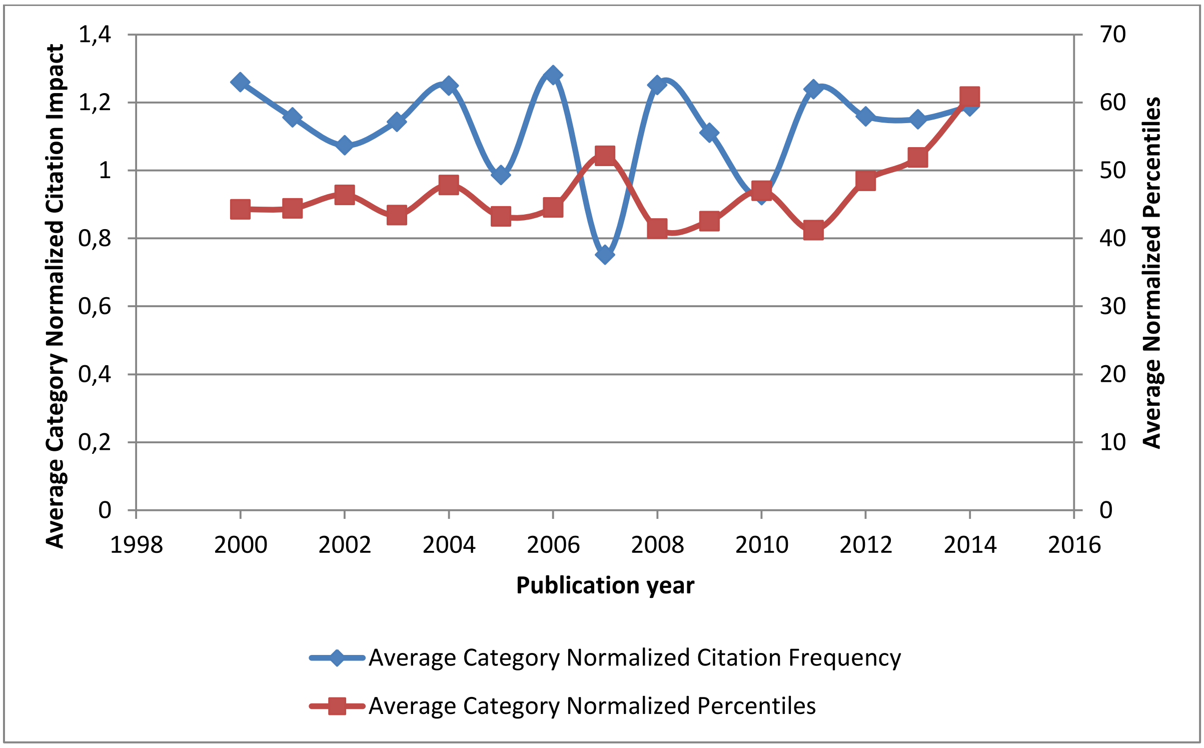

Figure 9. The Average Category Normalized Citation Frequency and Category Normalized Percentiles.

Top-papers

Identifying those papers that have a category normalized citation impact of at least five times the expected, we find 120 such papers, which is four percent of the analyzed set of papers. Mapping the longitudinal development (Figure 10), we can identify a clear downward trend between 2000 and 2007 followed by a recovery.

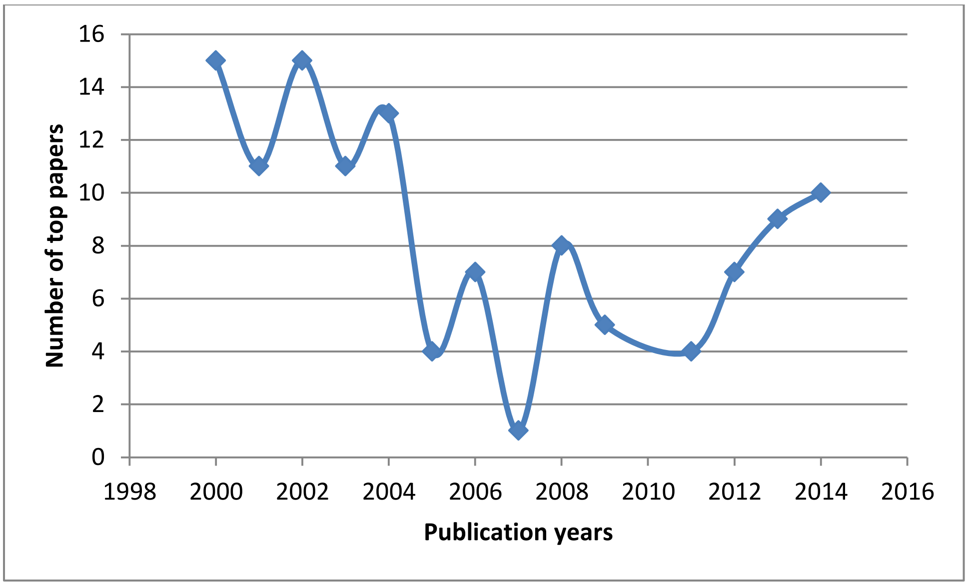

Figure 10. The distribution of top-papers over the period of observation.

Investigating the distribution of papers over subject categories provided with little information as the dispersion of papers over categories was very even. Hence, no particular subject was associated with top-papers.

Summary of findings

About 80 percent of all papers were produced by 50 percent of the journals, and no apparent core of preferred journals could be identified. The growth rate of papers was four percent per year, which is lower than the corresponding growth rate for each of the two universities.

Less than half of the papers are involved in international collaboration. In comparison with a corresponding national distribution, collaboration is more spread though more intense with regard to Russia. The five most frequent collaborating countries were USA, Germany, Russia, England and France. The most important hub of collaboration was USA followed by Germany.

The most significant collaborating organization was Sahlgrenska university hospital which reflects the strong connection between University of Gothenburg's medical faculty and the university hospital. The top-collaborators were mostly Swedish organizations though two foreign universities were ranked within the top-ten organizations: *Russian Academy of Sciences* and *Centre National de la Recherche Scientifique*. Otherwise, collaboration was rather evenly distributed over organizations.

Exploring relations within the network of collaborating organizations, cohesive sub-groups were identified at various levels of collaborative strength and interconnectivity. A significant and dense network of eight foreign organization could be extracted. Another significant sub-network involved University of Gothenburg and Chalmers University of Technology as a common hub connecting several Swedish universities along with *Russian Academy of Sciences* at a high level of collaboration strength.

The center of gravity of research areas comprised mathematical and physical sciences. It was found that the concentration of papers to categories was in line with the corresponding distribution of Chalmers University of Technology but higher than for University of Gothenburg. Applying severe thresholds of paper- and category frequency, six collaborating organization with their most typical research areas were identified.

It was found that the average impact computed as the category normalized citation frequency for the total period of observation was somewhat above the global average. Over the period of observation, impact fluctuated considerably. Measuring impact in terms of a normalized percentile distribution, thus avoiding the influence of outliers, a more gloomy pattern is seen with lower than median points during most of the period of observation. However, a positive trend is seen at the end of the period. Identifying top-papers, i.e. papers with a citation frequency at least five times the expected, a notable decline of the number of papers is seen from the start year to the middle of the period, followed by a recovery.