Ulf Sandström (www.forskningspolitik.se) **15/November/2005**
usandstrom@tele2.se **Work in Progress!** Language corrections to be added!

Toolbox for BIBLIOMETRIC ANALYSIS AND VISUALIZATION of RESEARCH GROUPS

Introduction

This paper consists of a proposed toolbox for bibliometric analysis of research groups. There are three parts of the toolbox:

- Firstly, a method for bibliometric identification of papers by members of the research group;
- Secondly, a method for bibliometric analysis of papers using normalized field indicators;
- Thirdly, a method for visualization of collaborations and intellectual spaces.

Except for the first method these methods rely on work done by others. The contribution in this paper is putting different methods together for the specific objective of analyzing research groups (or departments at universities and research institutes).

The **advantages** of this toolbox are several:

1. It offers a secure method for finding all articles written by the research group
2. It offers a scale-independent indicator for impact assessment
3. It offers a generation-independent indicator for impact assessment
4. It offers an “area-of-research-independent” indicator for impact assessment
5. It offers a time trend analysis
6. It offers a visualisation of coherence in the research group.

On Bibliometric Indicators

International scientific influence is the first and foremost parameter in the assessment of research performance. International influence, or impact, can be considered as an important, measurable aspect of scientific quality. Therefore, standardized, bibliometric procedures have been developed to assess research performance. The bibliometric approach works well in the large majority of the natural, the medical, the applied sciences, and in several fields within the social and behavioral sciences. At least for the Swedish case there are a number of other databases that can be used for the study of researchers within social science and the humanities. A method for this is discussed in two of my papers at forskningspolitik.se.[1]

Bibliometric data should be treated as basic information complementing other, e.g. expert opinion (peer review).[2] When used for funding decisions it forces experts to better precision and might provide new insights. Particularly at the level of research programs this method can be used as a tool for decision-making in science policy, above all in priority setting.[3]

In the following, research output is defined as the number of articles, as far as covered by the Web of Science, mainly the Science Citation Index (SCI), the Social Science Citation Index (SSCI), or the Arts & Humanities Citation Index (AHCI). As "article" the following publication-types are considered: normal articles (including proceedings papers published in journals), notes, letters and reviews, but *not* meeting abstracts, obituaries, corrections, editorials, etc.

Part I. A Methodology for Bibliometric Identification

A crucial part of the methodology concerns the retrieval of documents, or what I call the bibliometric identification. The normal procedure for this is a matching of names of team member or items in a publication list with the publication database.[4] Due to problems of accuracy the result of this process has to be verified by the researchers. Consequently, this procedure might be time-consuming and laborious. A magnitude of problems might appear: researchers might be abroad, they might be tired of evaluations, and also they might have a low level of accuracy in their own publication list.

Therefore, I propose another method for the bibliometric identification. This method relies on the full CV (including publication list), but does not implicate a verification procedure.

Depending on the objective all names of a research group or the four-five most senior researchers should be subject to bibliometric identification. Basically, the process is manual and relies on the online version of Web of Science (WoS). Each name is searched on the WoS (articles, notes, letters and reviews) for a specific period e.g. a period of seven years (1998-2004). In several cases, if last name and initial is commonly used, a detailed analysis has to be performed. For that purpose the full CV from all group members is used. The CV gives information on all institutional addresses of interest. In most cases this information is sufficient if there is also an indication on the subject categories where the researchers are active. If this is not the case, titles can be double-checked against articles listed in the CV.[5] It should be underlined

[1] See the reports: *Utvärdering av forskningens produktivitet och kvalitet med hjälp av friktionsmodellen* (Sept 2000) and *Relevansbedömning av forskning i ett processperspektiv* (Oct 2000), both available at www.forskningspolitik.se. Enclosure 4 shows that "History" is the fifth largest subject category on Web of Science considering no. of journals.

[2] See Martin, B.R. & Irvine, J. (1983) "Assessing basic research", *Research Policy*, 12, 61-90.

[3] Here I follow the line of argument and methodological points of departure put through by the Leiden group. See van Raan, A. F. J. (2004) "Statistical Properties of Bibliometric Indicators: Research Group Indicator Distributions and Correlations", *Journal of the American Society for Information Science and Technology*, November 12, 2004; van Raan, A. F. J. (2003)."The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments", *Technikfolgenabschätzung-Theorie und Praxis/Technology Assessment-Theory and Practice*, 1, 12, March, p. 20-29. For a critique of "amateur bibliometrics", see van Raan, A. F. J. (2005). "Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods", *Scientometrics*, 62, 133–143.

[4] For a more detailed description of the procedure, see Moed, H.F. (2005) *Citation Analysis in Research Evaluation*. Springer Verlag, pp. 72-74.

[5] My experience it that only a small portion of papers listed in CVs is not found. But, it should be mentioned that the bibliometric identification process might consist of some minor mistakes, e. g. due to misspelling of names, due to

that information of that sort often is available from the Internet or from the department management. If the evaluation is performed in the context of competition for grants then there is almost always a section for CV in the applications.

In my opinion this method is more secure than other methods. As mentioned above, bibliometricians match lists of publications from a group against the database. This should be efficient from a time perspective, if the publication list is fixed according to the standards of Thomson/ISI. Otherwise, the matching procedure will not work or might be laborious. Also, this procedure involves the researchers for validation and this is not an efficient use of their time. A manual procedure using the WoS has many advantages: if made with prudence it is much more controlled. If made by experienced senior investigators, it is also relatively time efficient. According to my experience it is possible to perform 75-100 individuals per day, i.e. seven to ten research groups per day. Having performed 2,500 identifications myself this is based on experience.

Another advantage with this procedure is that it lay the basis for visualizations (see part III below). Also, it gives data on the individual level, which might be of importance either for a funding body or for the management of a department/institute.

Part II. Methodology for Bibliometric Analysis

For a comprehensive bibliometric analysis it is necessary to have data from the Thomson/ISI, preferably data comparable to the Web of Science. With this database in operation, using advanced programming, it is possible to perform an analysis of quality indicators based on citations to journal articles.

In order to find out whether a certain volume of citations, or a certain citation-per-publication value, is high or low, it is necessary to have an international reference value. Proposed here is a normalization method using a reliable measure of *relative, internationally field-normalized impact*. Citation rates are constantly increasing.[6] Therefore it is important to normalize the measured impact of a research group or institute (number of citations per publication, *CPP*) to international reference values.[7]

Bibliometrics with normalization between areas

The first step is to calculate the average citation rate of all papers on a world level in the journals in which the research group has published. This indicator can be called *JCS* or, to be correct, JCS*(mean)*[8] and reveals the mean Journal Citation Score of the research group's "journal set". As an indicator *JCS* defines a world-wide reference level for the citation rate of the group.

Type of article (article, letter note or review) is taken into account **so that different reference values** are designed for each type. This is important as citation rates for reviews compared to

---

different usage of acronyms over time, and lack of addresses in some journals. Although a journal is indexed a small number of journal issues occasionally might not have been registered by ISI.

[6] See Olle Persson et al (2004) on indicator inflation.

[7] The methodology was originally developed by Braun & Glänzel, see T. Braun, W. Glänzel, A. Schubert, "Publication Productivity. From Frequency Distributions to Scientometric Indicators". *J. Inf. Sci.*, 16 (1990) 37-44 and has been taken further by the Leiden group; e.g. van Raan, AFJ. (1996). "Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises". *Scientometrics*, 36(3):397–420. **This part of the paper strictly follows the format for Leiden reports**.

[8] For reasons of simplification we take this procedure of counting averages for granted in the following.

articles might be relatively higher. Also, reference values should be designed for each year. The ratio *CPP/JCS* help us to observe whether the measured impact is *above* or *below* international average.

Comparison of the group's citation rate *(CPP)* with the average citation rate of its journal set (*JCS*) takes us to the question "degree of difficulty". If one group publishes in high impact journals, and another group in low level journals, the citation rate of articles published by both groups may be equal *relative to* the average citation rate of their respective journal sets. Evidently, the first group performs better than the second.

Therefore, we need a second international reference level, a *field*-based average called Field Citation Score, *FCS*.[9] Field in this context refers to the subject categories used by Thomson/ISI in their classification of journals. The FCS-indicator is based on the citation rate of *all* papers published in *all* journals of the field(s) in which the group is active, and not only in the journals in which the group's researchers publish their papers. A publication in a less prestigious journal might give a (relatively) high *CPP/JCS* but a lower *CPP/FCS*, and a publication in a more prestigious journal should result in a higher *CPP/FCS*. Normally publications in prestigious journals generate an impact above the field-specific average (which is the logic behind Journal Impact Figures delivered by Thomson/ISI in their Journal Citation Reports). In my opinion, university departments should use the difficulty level as an indicator of where to publish results.

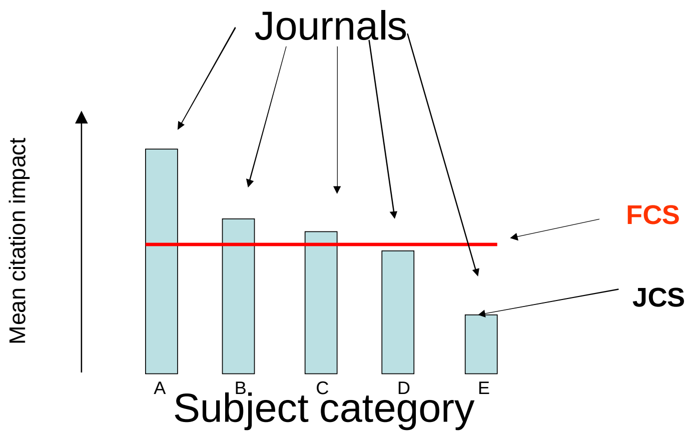


Figure 1. The normalization procedure for journal and field citation scores.

Here, the subject category consists of journals A-E. For each of these journals a JCS (journal citation score) can be calculated. This is the journal mean citation level for the year under investigation. A specific article might have actual citations (CPP) above, below or on par with this mean level. All journals in the subject category together form the basis for the FCS (field citation score). Accordingly, a specific article might have CPP over or under the FCS mean level. A researcher publishing in journal A will probably find it easier to reach the mean FCS-level than a researcher publishing in journal E. Note that the mean number of journals in a subject category (first classifications) in the Web of Science (full version) is 41 journals (period 2002-2003, SCI-data). For a full picture we should add that most journals are multi-assigned so with the help of almost 250 subject categories Thomson can accurately cover the content of journals.

The procedure applied in the calculation of *JCS* is also used for *FCS*. Almost always a research group is active in more than one field. In such cases a weighted average value is calculated, the

[9] To be precise we should talk about FCS (mean) for the same reasons as indicated above.

weights being determined by the total number of papers published by the group in each field. If the group publishes in journals belonging to genetics and heredity, as well as to cell biology, then the *FCS (mean)* of this group will be based on both field averages. Thus, indicator *FCS* represents a *world average* in a specific (combination of) field(s).[10]

If the ratio *CPP/FCS* is above 1.0, the impact of the groups papers exceeds the field-based (i.e., *all* journals in the field) world average. Further, also the ratio *JCS/FCS* is an interesting indicator. When this indicator is above 1.0, the mean citation score of the group's journal set exceeds the mean citation score of all papers published in the field(s) to which the journals belong.

For an analysis of groups in social sciences a five-year citation window is preferred. As journals normally have fewer articles per number and fewer issues per year in the social sciences the peak for citations comes later than in the sciences. Although there is a high correlation between CPP/FCS_2 and CPP/FCS_5 figures are corrected for the high variations that often occur with a shorter citation window in areas of science characterized by a low production of journal articles.[11]

The *internationally standardized impact indicator CPP/FCS* is regarded as the **"crown"** indicator by the Leiden bibliometricians. With this indicator you can immediately observe whether the performance of a research group or institute or country is significantly far below (indicator value < 0.5), below (indicator value 0.5-0.8), around (0.8-1.2), above (1.2-1.5), or far above (> 1.5) the international impact standard of the field.

In this specific analysis of rather small research groups it should be underlined that in the measurement of scientific impact one has to take into account the *aggregation level of the entity* under study. The higher the aggregation level, the larger the volume in publications and the more difficult it is to have an impact significantly above the international level. Based on international experiences, the following is a rule of thumb:

Generally, at the research group level a *CPP/FCS* value above 2.0 indicates a very strong group; values in the magnitude of 3.0 indicate that the group should be considered as "excellent and comparable to top-groups at the best European universities".[12]

Bibliometric Results: An Example

In table 1 publication data for 12 research groups is reported (see below). The groups are active in a diverse set of research fields. The first column contains the number of papers published, P, which gives a good indication of the size of a group, but the differences between groups are substantial. In the second column we find the average number of citations per paper (CPP), using a two-year citation window. The third column uses a five-year citations window. Data are not corrected for self-citations. In the fourth column we show figures for mean number of authors per paper. Average number of international collaborations is displayed in column 5 (operationalized as number of countries per article).

---

[10] To achieve a high precision the methodology accounts for multi-assignation of subject categories. An article in a journal with three subject categories receives one third of FCS-values from each category.
[11] See D. Aksnes dissertation *Citations and their use as indicators in science policy*. University of Twente, (March 2005), c.f. Aksnes, D. (2003), "A macro study of self-citation", *Scientometrics* vol 56, no. 235-246.
[12] van Raan op cit. (2003).

The core analytic scheme is as follows.[13] Time period is 1998-2004. For papers published in 1998, citations are counted during the period 1998-2000, for 1999 papers citations in 1997-2001, and so on, i.e. we use a two-year window for citations. There is ample empirical evidence that in the natural and life sciences – basic as well as applied – the average "peak" in the number of citations is in the third year after publication and a large part of the citations comes during the first two years. The correlation between 2-year window and 5-year window is as large as 0.98 (Pearson r) in this particular material.[14]

The following columns – 6, 7 and 8 – show the result of the citation analysis using the method described above. In the table groups are ranked from top to bottom on the basis of column 8, i.e. the crown indicator. Interestingly, the result indicate that the method is scale-independent, and that groups active in applied areas (e.g. applied energy research) can be as successful as groups active in basic science.

Table 1. Bibliometric Indicators for Research Groups

Column	1	2	3	4	5	6	7	8
	Papers	CPP	CPP	AU	Internat	CPP/	JCS/	CPP/
GROUP	P	(2 yr)	(5 yr)	mean	collab	JCS2	FCS	FCS2
Chemistry	131	7,68	16,72	5,46	1,65	1,92	1,35	3,04
Energy	46	3,87	6,37	5,52	1,96	2,26	1,06	2,43
Biology	80	8,15	13,73	6,29	1,84	1,31	1,77	2,07
Ecosystems	96	4,95	10,23	3,8	1,44	1,32	1,54	1,92
Plant Physiology	100	7,68	16,27	4,21	1,67	1,27	1,66	1,91
Geoscience	82	3,54	6,28	4,04	1,6	1,51	1,31	1,81
Food Science	172	4,72	8,89	7,39	2,03	1,39	1,21	1,75
Microbiology	119	3,09	5,89	3,76	1,49	1,13	1,2	1,32
Forestry	41	1,63	4,29	2,88	1,12	1,34	0,99	1,25
Genomics	117	2,67	5,03	3,88	1,5	1,16	1,09	1,14
Biodiversity	62	2,27	4,74	3,4	1,34	1,15	1,08	0,96
Building Tech	21	1,43	2,19	2,43	1,24	0,7	0,96	0,71

Note: Team size in each group is about 4.

Time trends

Table 1 show figures for average performance over a specific period of time. But, normally we also want to look into the development over time for research groups. Figures 2-13 below display two aspects of performance: firstly, the number articles per year (bar), secondly, the field citations score per year (line). Again, remember that FCSm-data for year 2004 are unstable and not reliable. Notice that scales are heterogeneous between groups as the variation in productivity and quality is substantial. Accordingly, this analysis should be focused on the trend and not the relative position.

[13] The methodology was developed by the Leiden group, see van Raan, A. (2004) "Measuring Science: Capita Selecta of Current Main Issues", *Handbook of Quantitative Science and Technology Research: The use of publication and patent statistics in studies of S&T systems.* Dordrecht/Boston/London: Kluwer Academic Publishers 2004, pp.19-50. See also van Raan, A.F.J. (2005) "Measurement of central aspects of scientific research", *Measurement*, 3(1), 1–19.As mentioned above this section strictly follows the format for Leiden reports.

[14] Here is used a two-year citation window while the Leiden group normally use a window with de-escalation, i.e. the window contain of articles with 4-year, 3-year, 2-year and 1-year windows simultaneously.

It is interesting to see that groups with a production of 15 articles per year or more have a high stability over time. Several groups, e.g. chemistry, biology and genomics, have only marginal shifts over time[15] and can with high prediction be said to produce high quality research. From the perspective of a funding body it should not be venturesome to invest in these groups. Other groups, plant phys. and ecosystems, seem to have a more negative development. Some of the groups in this sample show an unpredictable pattern with ups and downs. Almost always is this the case for groups with few articles, five to ten per year, which is an argument for having a wider base for the bibliometric identification – groups should be able to produce at least 30 articles per year (whole counts). When this criterion is fulfilled we might use these bibliometric indicators and the results will have a high stability.

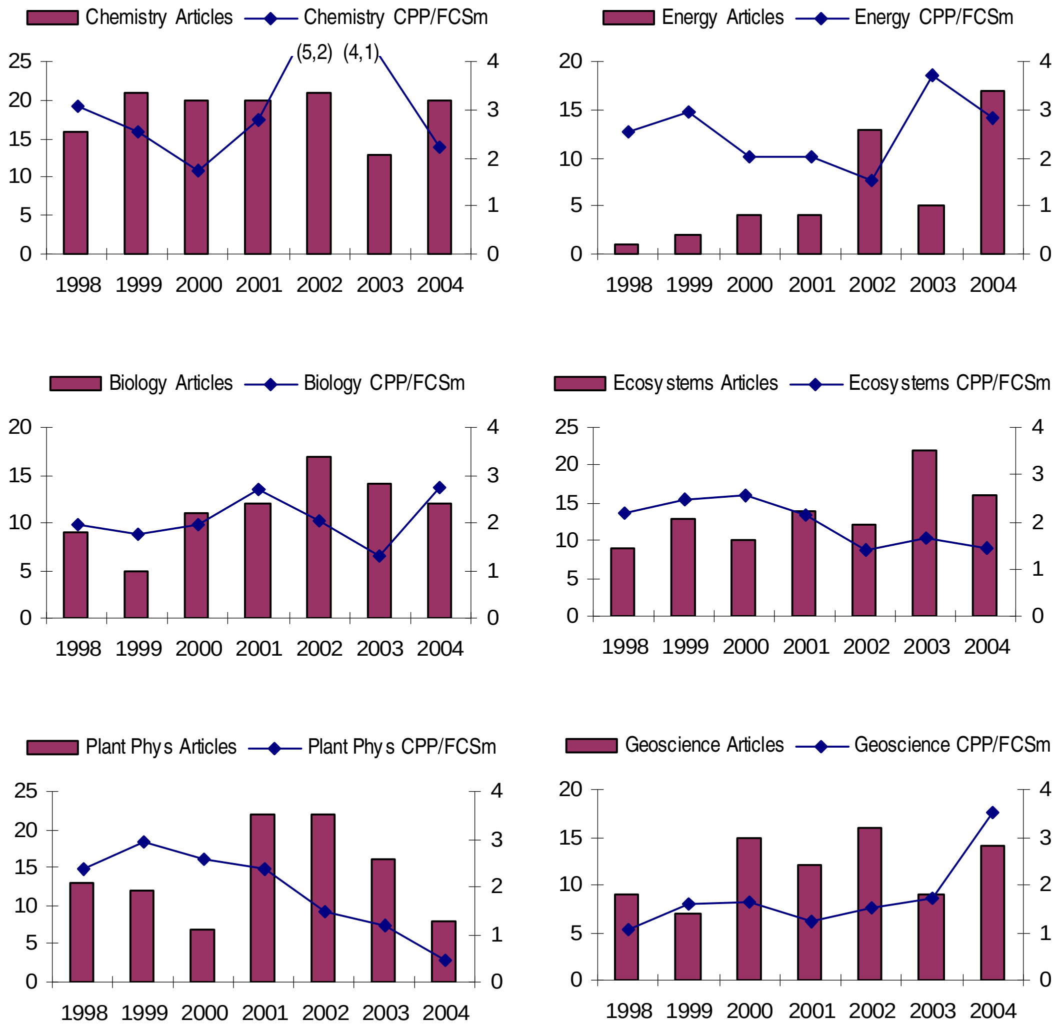


Figure 2-7: Articles (bars) and CPP/FCS2 (line) per year and group.

[15] When there are shifts we should be able to explain these shifts for the group: have they published a specifically interesting papers; have they lost a highly productive researcher; or other explanations of that type.

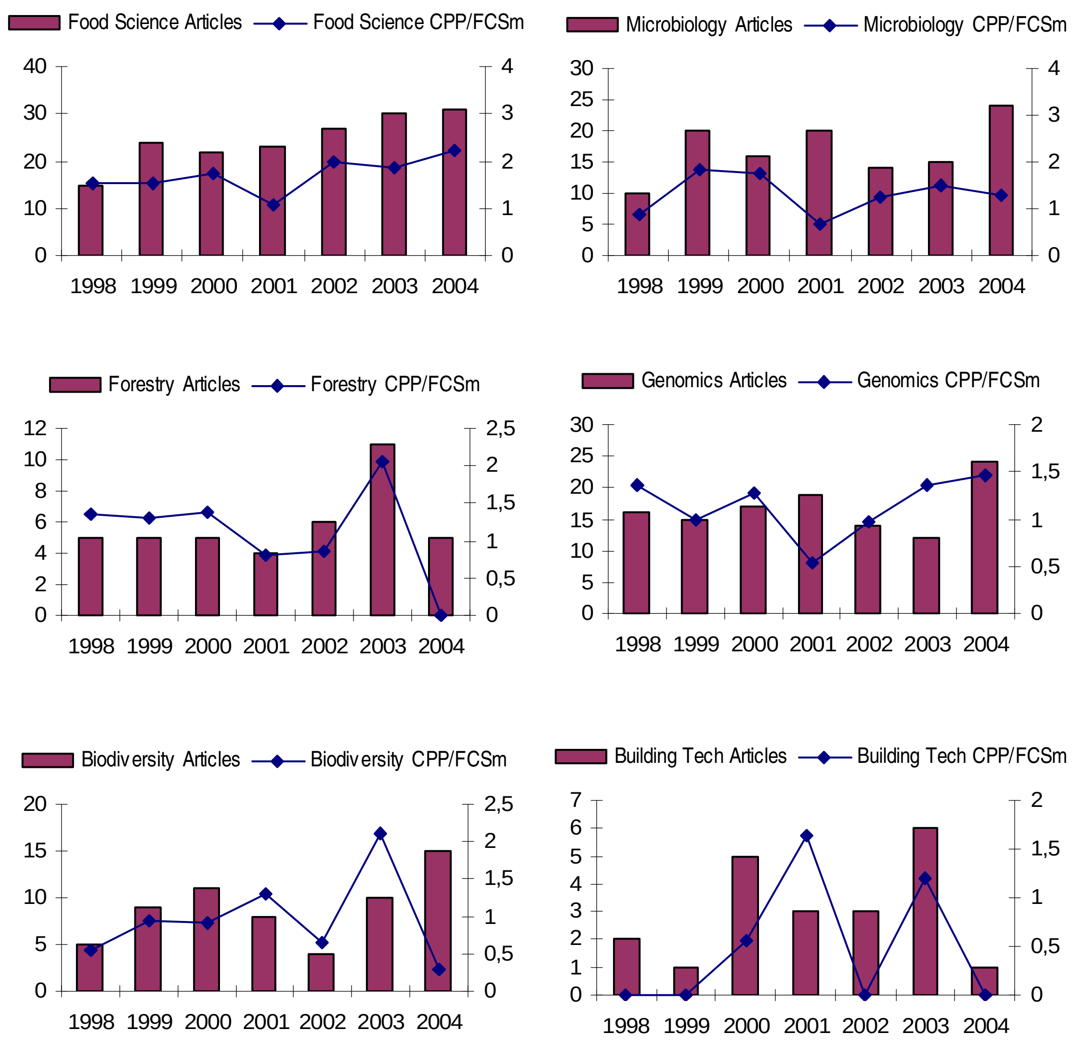


Figure 8-12: Articles (bars) and CPP/FCS2 (line) per year and group.

Fractional counts of publications and citations

In most fields of research scientific work is done in a collaborative mode. Collaborations make it necessary to differentiate between whole counts and fractional counts of papers and citations.[16] Fractional counts give a figure of weight for the contribution of the group to the quantitative indicators of all their papers. By dividing the number of authors from the group with the number of all authors on a paper we introduce a fractional counting procedure. In order to indicate the use of this procedure we use the acronym FRAC for these figures in the following, see table 2. Column 9 gives the fractionalised number of papers and column 10 the percentage of own contributions from the group (column 9 divided by column 1). In this example data are from excellence groups at Swedish universities competing for large grants.

Fractional counting is a way of controlling for the effect of collaboration when measuring output and impact, and used in combination with whole counts the effects of collaboration can be studied. In consequence, from FRAC-figures we can see to what extent the group receives many citations on collaborative papers only, or if all papers from the group are cited in the same manner. In short, when column 12 figures are lower than column 8 it follows that the group normally receive fewer citations on their own papers than in collaborative papers, and vice versa.

Table 2. Bibliometric Indicators for Research Groups with fractional counts

Column	1	2	4	5	6	7	8	9	10	11	12
								FRAC		*FRAC*	*FRAC*
GROUPname	**P**	**CPP2**	**AUm**	**IntColl**	**CPP/JCS2**	**JCS/FCS2**	**CPP/FCS2**	*_P*	*%*	*_CPP2*	*_CPP/FCS2*
Group 1	150	19,75	3,96	1,61	2,68	1,53	4,22	71	0,47	10,49	2,30
Group 2	83	3,40	2,07	1,67	1,53	1,69	3,33	48	0,58	3,72	3,40
Group 3	442	13,26	4,42	1,61	1,81	1,39	2,68	203	0,46	9,82	2,38
Group 4	287	5,32	4,61	1,52	1,40	1,29	2,06	126	0,44	4,97	1,89
Group 5	384	5,44	4,18	1,40	1,11	1,44	1,59	115	0,30	5,72	1,56
Group 6	570	3,39	4,65	1,74	1,21	1,41	1,52	155	0,27	3,76	1,51
Group 7	418	2,85	4,51	1,59	1,14	1,32	1,49	120	0,29	3,06	1,43
Group 8	167	2,89	4,02	1,52	1,08	1,36	1,45	50	0,30	2,80	1,38
Group 9	503	3,29	4,44	1,70	1,11	1,43	1,41	144	0,29	3,60	1,57
Group 10	402	2,94	5,13	1,68	1,02	1,39	1,29	94	0,23	2,68	1,15
Group 11	356	4,86	5,58	1,51	1,21	1,00	1,29	81	0,23	4,09	1,02
Group 12	240	7,44	4,01	1,51	0,94	1,45	1,24	73	0,31	7,33	1,15
Group 13	141	6,17	5,28	2,74	1,21	1,02	1,10	46	0,33	6,62	1,17

Note: All groups have 10 members.

[16] "When whole counts are used, the data should be read as the number of papers in which a particular unit occurs. This procedure leads to double counting. We can avoid the double counting by dividing the number of papers into fractions according to the number of units that have produced them, that is, by using *fractional counts*. The sum of all fractions will be equal to the actual number of papers in the data set. This allows us to measure the percentage of all papers that has been produced by a particular unit. If a paper is co-produced by two or more units from different countries and has been cited by ten other articles, each of the two units will get half of an article and five citations. The difference between whole and fractional counts for a given unit is a measure of the degree of collaboration. For example, if the whole count gives an institution 100 papers and the fractional count 50 papers, we can say that the difference is the effect of collaboration." *from* O. Persson et. al. *A bibliometric study of Finnish science* (2000), VTT working papers. (www.vtt.fi/ttr/pdf/wp48.pdf).

Part III. Methodology for Visualization of Research Groups

Strong research environments can be analyzed by citation analysis. Our interest is mainly focused on the mapping of author collaboration networks and co-citation networks. For each of the environments we produce two maps: firstly a map on co-authorships and secondly a document co-citation analysis.[17] For methodological reasons I will also display a third type of map showing journal co-citations.

An important task in mapping co-citation networks is to identify clusters of articles in correspondence to the underlying thematic groupings as perceived by scientists themselves. Co-citation links represent how often two articles are referenced together by a subsequent article. The strength of a co-citation link provides a very informative measure of the association between two articles.

For the purpose of visualizing the research groups we prefer to use the software CiteSpace developed by Chaomei Chen at Drexel University. There are a number of other possibilities.[18] CiteSpace is developed for analyzing and visualizing co-citation networks.[19] CiteSpace facilitates the creation of co-citation networks and processes them using network scaling algorithms such as Pathfinder network scaling. The purpose of using Pathfinder network scaling is to reduce the complexity of networks by retaining only the most salient links.

Pathfinder network scaling is a structural-modelling tool developed by cognitive psychologists. It provides an effective way to extract the most essential relationships from a given set of proximity data and simplifies a network by retaining only the strongest paths. Pathfinder uses a filtering criterion known as the triangle inequality condition to determine whether to remove or retain each link in the original network. Triangle inequality requires that the length of a path connecting two points in the network should not be longer than the length of other alternative paths connecting the two points, but go through extra intermediate points.

Pathfinder offers a better alternative to traditional layout and link-reduction methods such as MDS and minimum spanning trees. MDS provides no explicit representations of links, making it difficult to interpret the nature of each dimension in an MDS solution. Pathfinder explicitly represents the strongest links, and the interpretation relies on linkage instead of relative positions along each dimension. A very useful feature of Pathfinder networks is that the most significant and relevant work tends to be located in the centre of its Pathfinder network representation, whereas other relevant but less predominant work appears at the outskirt of such networks.[20]

CiteSpace allows the user to control the sampling procedure by selecting various thresholds for citation, co-citation, and co-citation coefficients. The lower a threshold, the more articles will be qualified for subsequent modelling and analysis. The network scaled co-citation networks are

---

[17] Other possible analyses are: author co-citation analysis and journal co-citation analysis.

[18] For example, SITKIS tool from Helsinki University of Technology (http://users.tkk.fi/%7Ehschildt/sitkis/); Olle Persson's Bibexcel software (www.umu.se/inforsk/Bibexcel/); Pajek (http://www.leydesdorff.net/jcr04) used by Leydesdorff and for some purposes a program at (http://www.csiss.org/clearinghouse/FlowMapper/) called Tobler's flow mapping.

[19] Chen, C & Paul, R.J.: "Visualizing a Knowledge Domain's Intellectual Structure", *IEEE Computer* March 2001, pp.65-71, see also Chen C. "Searching for intellectual turning points: progressive knowledge domain visualization", *Proceedings of the National Academy of Sciences of the USA*, April 6, 2004, vol. 101, suppl. 1. pp. 5303-5310.

[20] Chen, C., Paul R.J & O'Keefe, B. :"Fitting the Jigsaw of Citation: Information visualization in Domain Analysis", *Journal of the American Society for Information Science and Technology*, 52(4):315-330.

subsequently merged with the Pathfinder's topological properties preserved; in essence, all links in the final network must not violate the triangle inequality.

The merged network can be visualized. A number of colour encoding schemes are used to convey a variety of information. Co-citation links between articles are coloured by the time when the first instance of the co-citation was made. Cold colours (from blue to green) are from 1998-2000. Warmer colours (yellow and orange) are from 2001-2005. *Nodes with high centrality are marked with an extra purple ring.*

As an illustration two research groups are visualized using co-authorship and co-citation analysis, see enclosure 1-2. In enclosure 3 there is an illustration of journal co-citations analysis.

For the interpretation it can be said that a group with many nodes at a large distance from each other (Enclosure 2) probably lack an intensive collaboration during the period. Instead, a nested picture implicates a collaborative network (Enclosure 1).

The other picture, so called document co-citations, reveals what can be called "the intellectual space" for the group. To what extent are the groups members active in the same areas and to what extent are they citing the same journals? When interpreting network pictures it should be considered that collaborations not always becomes manifest in co-authorships.[21]

Normally, a scattered map can be interpreted as a scattered group with not so many intellectual commonalities, but there is also room for other interpretations. A group can be highly collaborative but at the same time they might be working on so many different topics that a scattered picture would be the expected outcome. The colours indicate either the stability in relations and reference behaviour, or an innovative behaviour – if there are many new collaborators and many new references used.

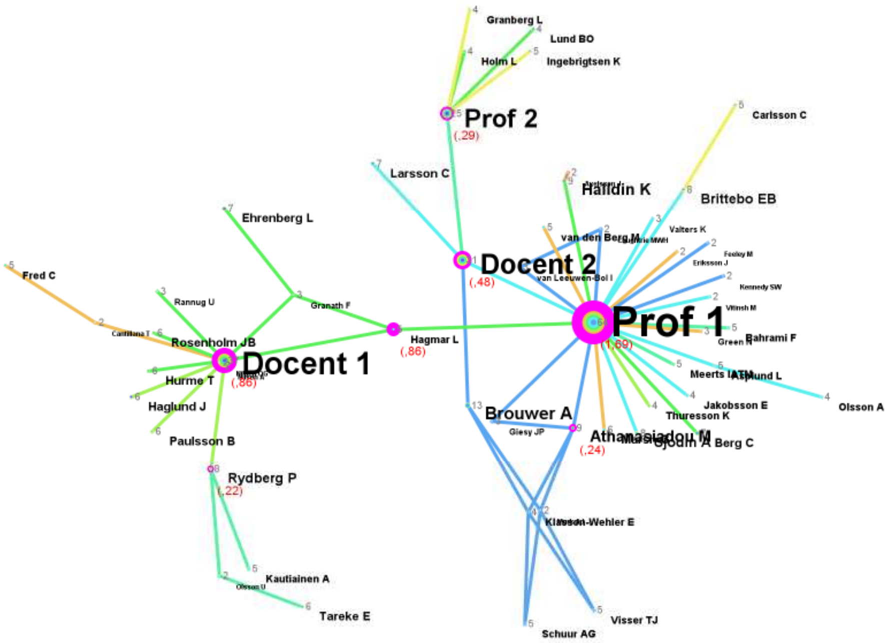
Finally, there is a section (enclosure 3) with maps that display the journal co-citations. With this analysis it is possible to see on the one hand the breadth and extensiveness or on the other hand the concentration of sources that are used in the articles published by the group. Which are the most central journal sources used by several of the members in the group?

An important question is whether the same thresholds should be used for all groups. If that line of operation is chosen, groups with a large production will have dense maps and groups with fewer papers will not appear as dense. Even though this necessarily gives a partly biased picture, my experience is that it is possible to see the relative frequency of collaborations within the group as well as to find an indication of the relative density of the intellectual space.

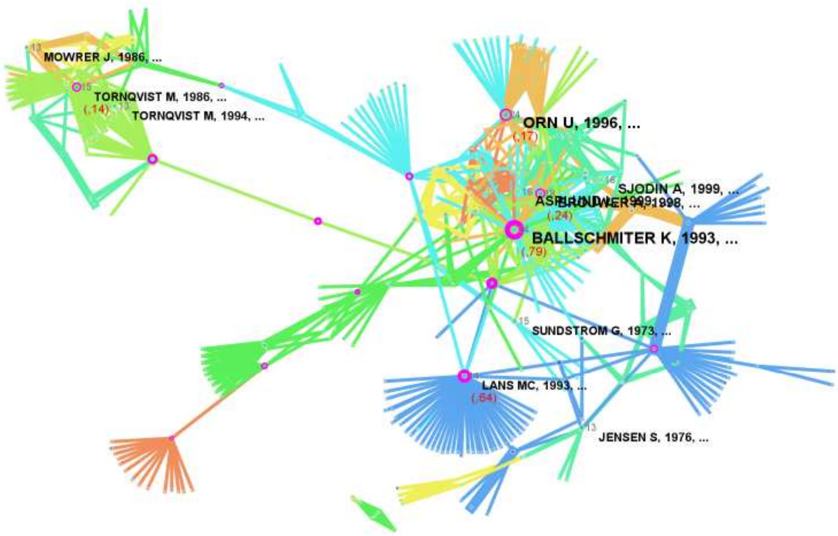
For the visualization work it is possible to use articles up until the date of identification. Added to the analysis are the most recent articles that are available from the on-line version of Web of Science. But, data for the bibliometric analysis is normally restricted to the year before.

[21] See e.g. the excellent paper by Katz JS, Martin BR, "What is research collaboration?" *Research Policy* 26 (1): 1-18, March 1997.

Enclosure 1

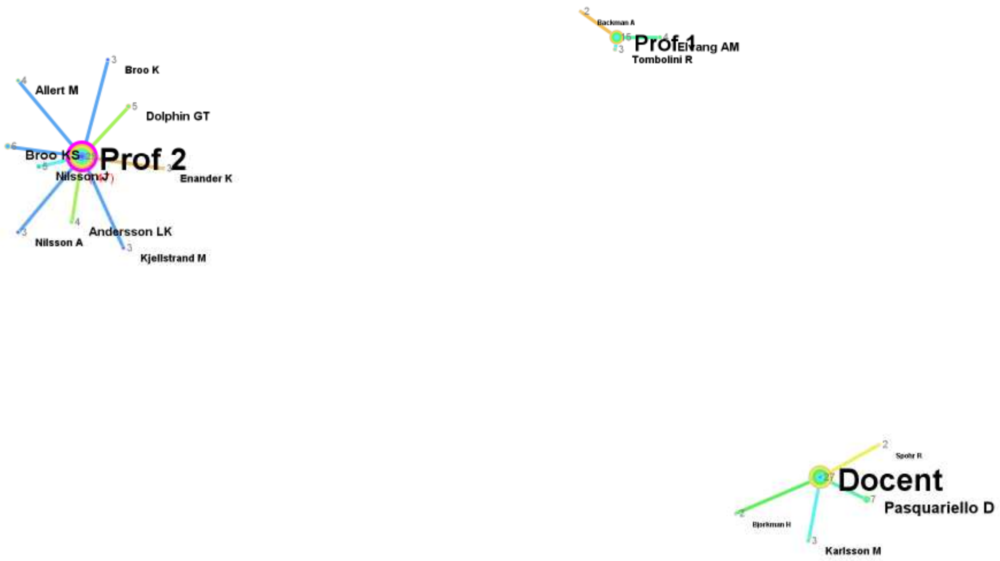


ABOVE: Co-author analysis___________BELOW: Document co-citation analysis

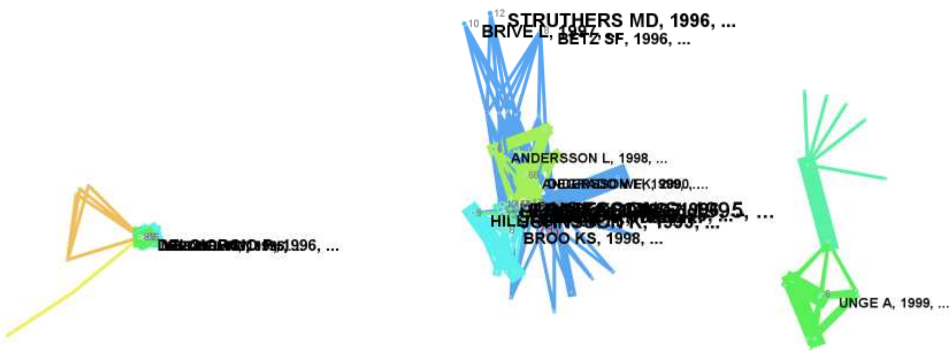


Group A (131 documents)

Enclosure 2

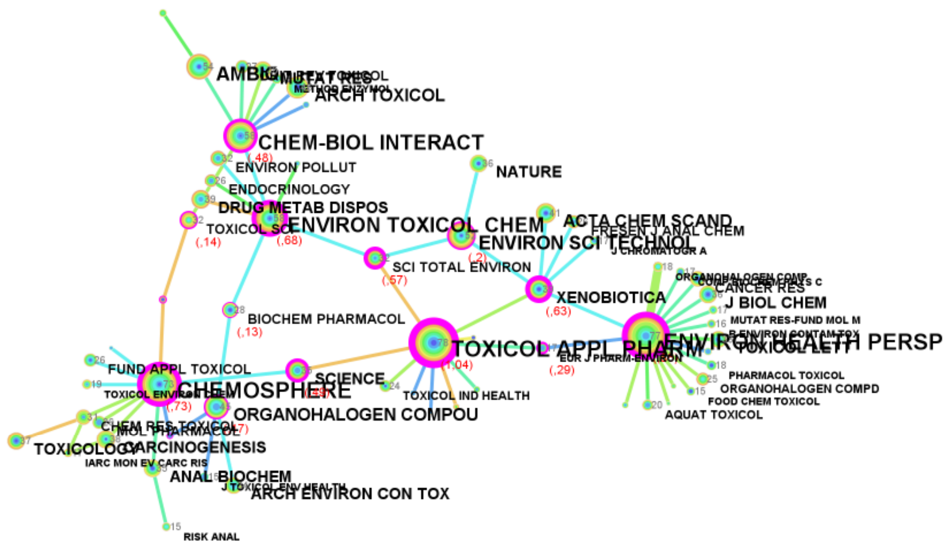


ABOVE: Co-author analysis___________BELOW: Document co-citation analysis

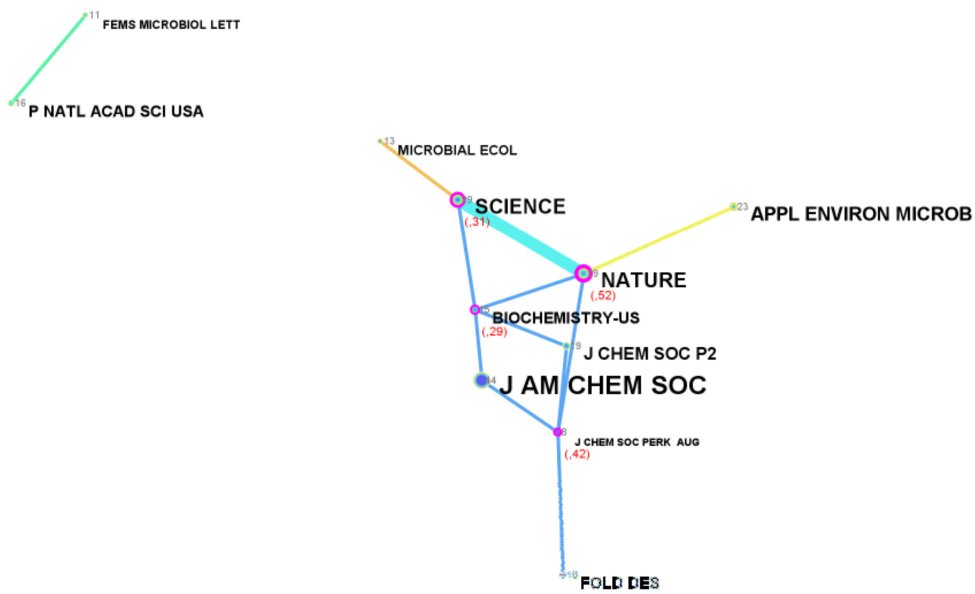


Group B (103 documents)

Enclosure 3. Journal co-citations for group A and B



ABOVE: Group A____________BELOW: Group B



Enclosure 4. The ten largest subject categories counted from number of journals.

Subject category	No of journals
COMPUTER SCIENCE, THEORY & METHODS	299
BIOCHEMISTRY & MOLECULAR BIOLOGY	198
MATHEMATICS	183
COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE	175
HISTORY	164
CHEMISTRY, MULTIDISCIPLINARY	152
PHARMACOLOGY & PHARMACY	137
NEUROSCIENCES	133
MATERIALS SCIENCE, MULTIDISCIPLINARY	132
ECONOMICS	126
MATHEMATICS, APPLIED	124
ENGINEERING, ELECTRICAL & ELECTRONIC	123
VETERINARY SCIENCES	109
PLANT SCIENCES	108
CELL BIOLOGY	107

References

Aksnes, D. (2003), "A macro study of self-citation", *Scientometrics* vol 56, no. 235-246

Aksnes, D. (2005) *Citations and their use as indicators in science policy*. University of Twente, (Diss. March 2005)

Braun, T., W. Glänzel, A. Schubert (1990), "Publication Productivity. From Frequency Distributions to Scientometric Indicators". *Journal of Information Science*, 16, 37-44

Chen C. (2004), "Searching for intellectual turning points: progressive knowledge domain visualization", *Proceedings of the National Academy of Sciences of the USA*, April 6, vol. 101, suppl. 1. pp. 5303-5310

Chen, C & Paul, R.J. (2001), "Visualizing a Knowledge Domain's Intellectual Structure", *IEEE Computer* March, pp.65-71

Chen, C., Paul R.J & O'Keefe, B. :"Fitting the Jigsaw of Citation: Information visualization in Domain Analysis", *Journal of the American Society for Information Science and Technology*, 52(4):315-330

Katz JS, Martin BR, "What is research collaboration?" *Research Policy* 26 (1): 1-18 MARCH 1997

Martin, B.R. & Irvine, J. (1983) "Assessing basic research", *Research Policy*, 12, 61-90

Persson et. al. (2000). *A bibliometric study of Finnish science*, VTT working papers.

van Raan, AFJ. (1996). "Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises". *Scientometrics*, 36(3):397–420

van Raan, A. F. J. (2003)."The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments", *Technikfolgenabschätzung-Theorie und Praxis/Technology Assessment-Theory and Practice*, 1, 12, March, p. 20-29

van Raan, A. (2004) "Measuring Science: Capita Selecta of Current Main Issues", *Handbook of Quantitative Science and Technology Research: The use of publication and patent statistics in studies of S&T systems*. Dordrecht/Boston/London: Kluwer Academic Publishers 2004, pp.19-50

van Raan, A.F.J. (2005) "Measurement of central aspects of scientific research", *Measurement*, 3(1), 1–19

van Raan, A. F. J. (2005). "Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods", *Scientometrics*, 62, 133–143

van Raan, A. F. J. (forthcoming) "Statistical Properties of Bibliometric Indicators: Research Group Indicator Distributions and Correlations", *Journal of the American Society for Information Science and Technology*, November 12, 2004