Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 68, No. 3 (2006) 561–572

Exploring the analytical potential of comparing citing and cited source items

OLLE PERSSON

Inforsk, Department of Sociology, Umeå University, Umeå (Sweden)

Comparing properties of citing and cited source items opens a wide variety of analytical possibilities. In a study of citations among papers in the journal *Scientometrics* a number of analytical themes are identified. The analysis shows: the way in which a citation graph can be decomposed into different subparts; country specific citation patterns; the effects of self-citations and domestic citations; the mapping of cited author relationships using direct citation and co-citation links; and time slicing effects on impact ranking of countries and papers.

Introduction

Citation studies are mostly focussed on either the citing or the cited side of citation links. Less attention has been given to the analytical potential of comparing the source items of both sides of the citation link. For example, when the cited paper is also a source item we can attribute citations to all cited authors and not just to those listed first, as in a conventional author citation and author co-citation analysis. Such an approach has been shown to give a much more valid picture of influential authors in a field compared to first-author citation analysis (PERSSON, 2001). We can also exclude self-citations by comparing the list of citing authors with that of the cited document.

Received December 20, 2005

Address for correspondence:
OLLE PERSSON
Inforsk, Department of Sociology, Umeå University, SE-901 87 Umeå, Sweden
E-mail: olle.persson@soc.umu.se

0138–9130/US $ 20.00

In this paper some of the potentialities of looking closer at both sides of the citation link will be illustrated by studying the citations among papers in *Scientometrics*, from its start in 1978 up to year 2004. The following analytical aspects will be covered:

- chronological analysis and decomposition of the network of papers
- self-citations and domestic citations
- country specific citation patterns
- citations between country regions
- author mapping based on direct citations and co-citation links
- length of paper and reference list for citing and cited papers
- time slicing of citing and cited windows.

Data

Web of Science was used to download papers from *Scientometrics* 1978–2004. 1655 papers that have addresses and cited references form the basis of this study. Author names were standardized to harmonize with how first authors are represented in the reference lists, "Van Raan" becoming "VanRaan" etc. To find citation links among these papers a search key was made containing the last name of first author, publication year, volume and start page, for example "Moed, 1985, V8, P149". All in all 3904 citation links among the papers where found. For each of these links we have full bibliographic information of the citing and cited paper. One might rightly argue that the design of this study makes it a study of journal self-citations, leaving aside all citation links going beyond its own archive. I would suggest however that the study can still serve as a basis for illustrative examples, and to provide a method for examining more closely the phenomenon of journal self-citation itself.

Chronological analysis

The dynamics of a research field can be studied by establishing the chronological order of citing and cited documents. This yields a more detailed picture, than the one derived from using publication years. Provided we can establish a reasonable chronology of papers, the citations between them can be plotted in a "citations from-citations to" diagram. This is easily done for a single journal. Papers were sorted by year, volume, issue and page. It could be argued that papers from the same issue should have the same number since they all appear simultaneously. The effects of that should be studied in more detail in future research.

We can denote the number of papers separating the citing and the cited paper as chronological distance. For example, if the 100^{th} paper cites the 50^{th} then the chronological distance is 50 papers. This type of chronological analysis was first

conceived by PRICE (1965) to illustrate the growth of a citation network amongst papers in a research speciality. The typical triangle form of such a graph can be decomposed into the research front represented by dots around the diagonal citing backwards x number of papers. Foundational papers are highly cited papers that appear as horizontal lines of dots, while vertical lines of dots are the typical effect of review papers that cite a large number of earlier papers. The Price-graph was applied by HARGENS (2000), who compared seven different research fields, all having different so called reference-network graphs.

Figure 1 displays the reference-network graph of *Scientometrics*. The distribution of dots is denser close to the diagonal and sparser towards the lower right hand side of the graph. This indicates a certain progression of the field, since authors tend to give recent papers more attention than older ones. Generally, as shown by HARGENS (2000), life and natural science specialities tend to demonstrate a greater density of dots just below the diagonal compared to social and humanities research areas.

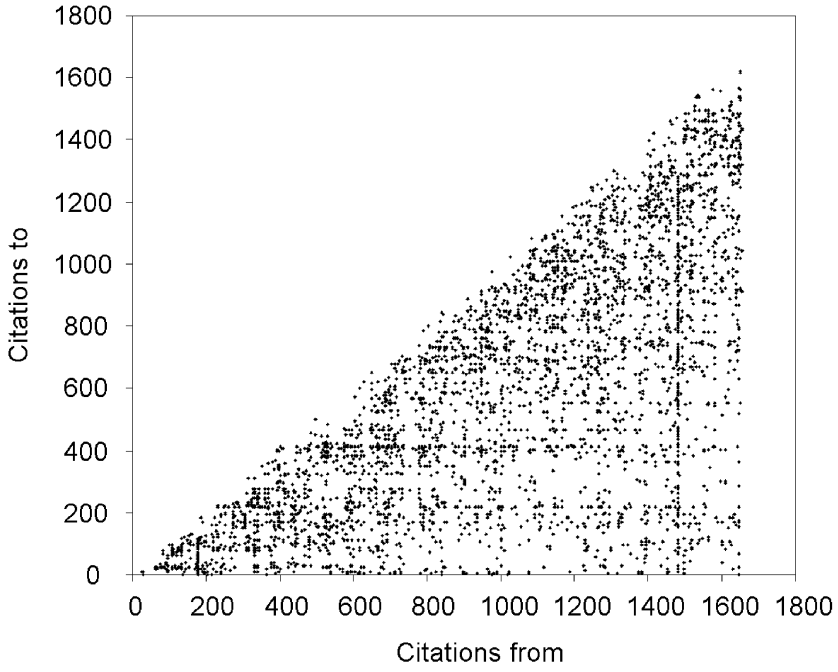

Figure 1.The reference-network graph for *Scientometrics* 1978–2004

Figure 2 highlights three typical parts of the graph. The research front in the journal *Scientometrics* is represented by citations going back no more than 100 papers, and appears as a cluster of dots close to the diagonal. The density of the front is similar as we move forward. However, there are some intervals that are empty. One of these comes from the Nalimov Memorial Issue in 2001 (papers no 1356 to 1373) that did not cite recent *Scientometrics* papers.

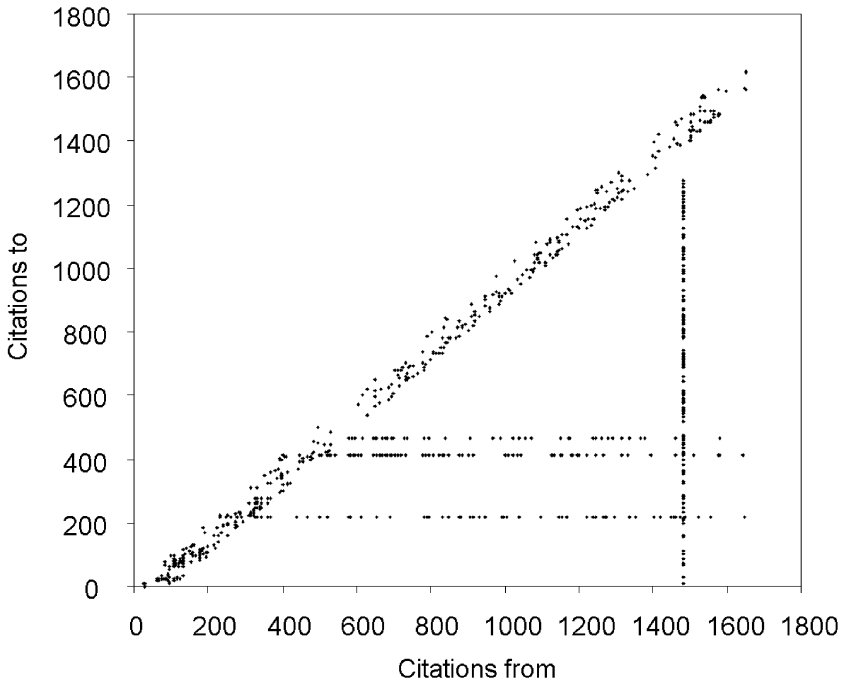

Figure 2. Some highlighted features of the reference-network graph for *Scientometrics* 1978–2004
Note: The research front is seen here as a cluster of dots close to the diagonal and representing citations to the 100 most recent papers counting back from the citing paper The vertical line of dots is a review paper, and the horizontal lines of dots are the three most cited papers.

The vertical line of dots is a review paper by GARG (2003): "An overview of cross-national, and institutional assessment as reflected in the international journal *Scientometrics*". This paper cited 154 of earlier *Scientometrics* papers.

The horizontal lines represent the three most cited papers. The most cited of these is: SCHUBERT, A., GLÄNZEL, W., BRAUN, T. (1989): "Scientometric datafiles – a comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields 1981–1985"; the second most cited is: SCHUBERT, A., BRAUN, T.

(1986): "Relative indicators and relational charts for comparative-assessment of publication output and citation impact"; and the third most cited is also by these two authors from 1990: "International collaboration in the sciences 1981–1985". The strong and lasting impact of the Hungarian indicators group is clear here.

Self-citations

Self-citations can be identified through overlapping author sets of citing and cited papers; the same author name occurring amongst the authors of the citing and the cited document may be taken to represent self-citation. Similarly, we can identify domestic and foreign links by matching citing paper countries with cited paper countries.

Table 1 shows the chronological distance, measured as the mean number of papers separating citing and cited papers, for various types of citation link. Self-citations appear to be relatively short-term phenomena, since the "distance" is on average 158 papers shorter than for non-self-citation links. This is in line with GLÄNZEL et al.'s observation (2004) that self-citations decrease over time. Self-citing domestic links span the shortest distance, on average 278 papers, compared to non self-citing foreign citations, which have a mean distance of 456 papers. This implies an effect of space or closeness on citing behaviour. It should be born in mind however that self-citation links are only 14 percent of all citation links, and domestic links constitute 30 percent of all links.

Table 1. Mean number of papers separating citing and cited papers for various types of citation links

	Not self-citation	Self-citation	All
Domestic citation	380	278	337
Foreign citation	456	342	455
All	441	283	419

Citations to countries and among countries

The reference-network graph can also be decomposed by country by highlighting citations to a given country. For US papers, which are cited by 786 documents, there seems to be a sparser pattern within the research front at the end of the time frame compared to the first years of the journal (Figure 3).

A "who cites whom" analysis can be made on all aggregation levels, among authors, departments, universities, countries and regions. Table 2 shows how regions cite each other. Europe and North America dominate the journal and are also the most citing and cited regions.

Table 2. Citation links between transnational regions

	Cited country region						
Citing country region	Africa	Australia & New Zealand	Europe	North America	South & Latin America	South & East Asia	Total
Africa	4	0	11	11	1	1	28
Australia & New Zealand	2	14	67	28	2	6	119
Europe	18	23	1825	515	52	140	2573
North America	4	10	191	189	6	18	418
South & Latin America	1	2	94	29	85	17	228
South & East Asia	13	18	411	131	36	215	824
Total	42	67	2599	903	182	397	4190

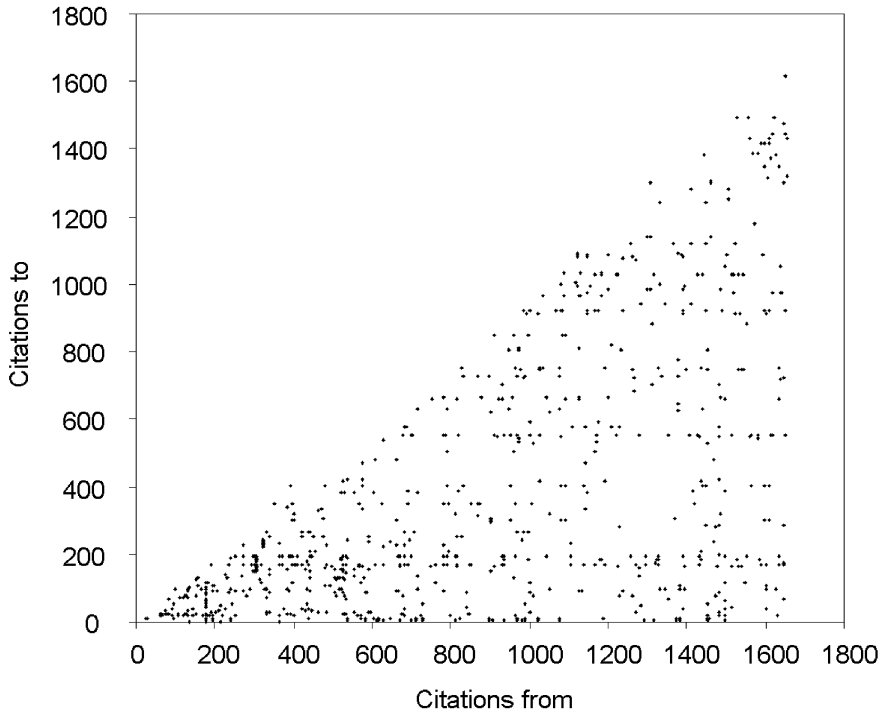

Figure 3. Scatter of citations to US papers in *Scientometrics*

North America gives Europe 191 citations but receives 515 citations from Europe. The tendency to cite within the region is much stronger for Europe compared to North America, which indicates a certain Euro-centrism in the field. Looking at the marginal distributions North America is the winner of this citation game.

If we look at the chronological distance it takes on average 262 papers before US papers get their first citation, 222 for Hungarian papers, 156 for Dutch papers, 266 for UK papers and 213 for German (author self-citing papers excluded). This gives the impression of a scientific community which is more interactive in Europe than in North-America. In fact this can also be studied. Taking the more frequent country citation links, the links between European countries and US are on average separated by 429 papers, while country links within Europe are separated by 350 papers.

Maps of authors

Citation based mapping of authors can be based on direct citation links between them or some indirect measure, like co-citations or shared cited authors. The latter two are generated from an underlying citation graph of direct citation links. Mapping based on direct citation links is rarely conducted, and it would be interesting to see whether a map based on direct links differed from one based on co-citations. In this study we are of course able to include all authors of citing and cited papers.

Figure 4. Map of most cited authors based on direct citation links
Note: Multi Dimensional Scaling was used to make the map from a matrix of direct citations among the 47 authors. The size of the circles is proportional to the citation frequency (self-citations excluded). The closer authors are the more citation links among them

The first step was to select authors to be mapped. First papers from 1995–2004 were selected to get a somewhat more contemporary picture. Then the 47 most cited authors cited by at least 10 papers were selected for mapping, whilst excluding self-citations in terms of overlapping author sets as defined above. The 47 most cited authors represent just a small fraction of totally 949 cited authors. The maps in Figure 4, based on direct citation links, and Figure 5, based on co-citations, do show some differences in the positions of cited authors. The names are the same but their positions vary. When direct citation links are used, collaborating authors are somewhat more apart compared to the co-citation map. A significant example is Melin and Persson who co-authored a highly cited paper in 1996 on co-authorship analysis. In Figure 4 they are quite apart, but much closer in Figure 5. When that paper is cited the authors are both co-cited once.

Figure 5. Map of most cited authors based on co-citations
Note: Multi Dimensional Scaling was used to make the map from a matrix of co-citations among the 47 authors. The size of the circles is proportional to the citation frequency (self-citations excluded). The closer authors are the more often they are co-cited

However, the direct citation link introduces the citing author who is coupled with each of them, while no link is established among the two co-authors. In addition, because self-citations are excluded no direct citation link from one of the authors to the co-authored paper will qualify. Another example is Braun who drifts away from Glänzel and Schubert in the direct citation graph. In other words, direct citation mapping eliminates the effect of collaboration. The reduction of links is obvious since the direct citation graph has 213 links between the authors while the author co-citation graph has 501 links. Ignoring collaboration is to conceal the fact that papers in most cases are co-authored and that credit should accrue to all of the collaborating authors.

Length of papers and reference list

Although not always made explicit, one of the purposes of writing a scientific paper is to be cited by other papers. The style of writing may have some effect on citation patterns, but it is unknown exactly what stylistic characteristics. Two aspects of style can be studied quite easily by counting the number of pages in a paper and the number of references it makes.

In Table 3 these two indicators are shown for citing and cited papers. It is quite clear that writing long papers does not pay off, since the cited papers generally have fewer pages. It is rather the "punch per word" that counts, not the number of them. But, when it comes to the length of reference lists, cited papers have somewhat longer lists than all papers. The fact that citing papers have many more references is quite logical since the more references a paper has the more likely it is that it also cites papers in the same journal.

Another observation is that there is a zero-correlation between the number of citations a paper gives to the journal and the number it gets from it. Loyalty of this kind apparently doesn't pay off.

Table 3. Mean paper length and mean number of references for citing, cited and all papers
Note: Only genuine articles are included

	Mean
Number of pages:	
Citing papers	18.15
Cited papers	16.28
All papers	19.10
Number of references:	
Citing papers	34.71
Cited papers	19.64
All papers	15.70

Time slicing

Publication years, or chronological intervals, can be used for studying the dynamics of a research field. Such time slicing has been used by CHEN (2004) to identify intellectual turning points. If we use the same time slices for citing and cited document we arrive at rankings of the most cited countries over the decades. Papers from US had the strongest impact during the 1980s, but were outscored by the Netherlands in the two later periods.

Table 4. Rank of most cited countries for different citing-cited windows
Note: International co-authored papers ignored

1978–1989	N of citations	1990–1999	N of citations	2000–20004	N of citations
US	185	Netherlands	126	Netherlands	33
Russia	75	Germany	125	India	25
Hungary	61	US	109	France	21
Netherlands	36	Hungary	94	US	20
Canada	18	France	74	UK	18
Germany	15	India	69	Belgium	15
India	14	UK	52	Denmark	15
UK	14	Spain	45	Spain	13
England	10	Brazil	18	Germany	11
Japan	9	Israel	17	Hungary	11

Table 5. The authors of most cited papers by chronological interval
Note: Self-citations removed

Paper interval	Authors of most cited titles	Citations	All citations to the interval
1–100	Beaver DD; Rosen R	49	365
101–200	Small H; Sweeney E; Greenlee E	30	409
201–300	Schubert A; Braun T	49	275
301–400	Braun T; Glänzel W; Schubert A	35	220
401–500	Schubert A; Glänzel W; Braun T	64	275
501–600	Narin F; Stevens K; Whitlow ES	24	219
601–700	Luukkonen T; Tijssen RJW; Persson O; Sivertsen G	23	279
701–800	Braun T; Glänzel W; Maczelka H; Schubert A	21	234
801–900	Moed HF; Debruin RE; Vanleeuwen TN	28	231
901–1000	Luukkonen T	12	245
1001–1100	Narin F; Olivastro D	12	187
1101–1200	Glänzel W; Schubert A; Czerwon HJ	14	136
1201–1300	Bjorneborn L; Ingwersen P	8	121

We can also have a closer look at the most cited papers in different chronological intervals. By focussing on cited papers rather than countries we get a view of the most influential researchers. This is done in Table 5, which for each interval of 100 papers shows the authors of the most cited papers. If we have a look at Figure 1, the slicing is made on the vertical axes for every 100th paper.

In the first interval we find two papers from 1979 on scientific collaboration by Beaver & Rosen. In the second period the co-citation clustering approach was presented by Small et al., in 1985. The next 3 periods were dominated by the Hungarian indicators group with Braun et al., as already indicated in Figure 2. The lead paper in the following period was by Narin et al. on citations to multinational papers. After that came Luukkonen et al. with a paper on measuring collaboration. In 1994 Braun et al. presented output and citation impact figures for countries. The Moed et al. paper from 1995 was also about assessment of national performance. In 1997 Luukkonen discussed Latour's theory of citations and in 1998 Narin & Olivastro presented their study of patents citing papers. In 1999 Glänzel et al. presented one paper on subject classification of papers in multidisciplinary journals, and another paper on EU-collaboration. During the last period Björneborn & Ingwersen introduced the field of webometrics in 2001. Several of these papers do mark essential steps in the development of bibliometric indicators.

Discussion

Comparing citing and cited source items can yield a wide variety of analysis and interesting information about a field of research. Some of the examples given are based on a chronology of papers, which is fairly easy to construct for a given journal such as *Scientometrics*. If papers from a research speciality are studied, papers will be scattered over several journals, making it harder to establish a chronology within a year for citing and cited papers. At least, it will demand more effort to make a reasonable order of papers.

One attractive outcome of having full bibliographic information on both citing and cited papers is that we have closer control of self-citations, or for that matter, citations within and between groups, institutions and countries. We can also study the chronological distance between citing and cited papers for different types of citation links.

Citation based author mapping can be made much more accurate by incorporating all citing and cited authors in the analysis. One apparent limitation of this approach is that the share of papers citing within the set will vary from field to field, as an effect of the citation behaviour of authors as well as the coverage of cited items in the database. On the other hand using citations among source items, and in combination with time slicing, may weed out less relevant cited documents.

Several of the examples given in this paper need further exploration. At this stage it is enough to conclude that there are a number of interesting openings for future research in comparing the properties and time order of citing and cited papers.

*

I am greatly thankful to Dr Ingrid Schild for suggestions and critical comments.

References

CHEN, C. M. (2004), Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America,* 101 : 5303–5310.

GARG, K. C. (2003), An overview of cross-national, national, and institutional assessment as reflected in the international journal Scientometrics. *Scientometrics,* 56 (2) : 169–199.

GLÄNZEL, W., THIJS, B., SCHLEMMER, B. (2004), A bibliometric approach to the role of author self-citations in scientific communication. *Scientometrics,* 59 (1) : 63–77.

HARGENS, L. L. (2000), Using the literature: Reference networks, reference contexts, and the social structure of scholarship. *American Sociological Review,* 65 (6) : 846–865.

LUUKKONEN, T., TIJSSEN, R. J. W., PERSSON, O., SIVERTSEN, G. (1993), The measurement of international scientific collaboration. *Scientometrics,* 28 (1) : 15–36.

PERSSON, O. (2001), All author citations versus first author citations. *Scientometrics,* 50 (2) : 339–344.

PRICE, D. J. (1965), Networks of scientific papers. *Science,* 149 : 510–515.

SCHUBERT, A., BRAUN, T. (1986), Relative indicators and relational charts for comparative-assessment of publication output and citation impact. *Scientometrics,* 9 (5-6) : 281–291.

SCHUBERT, A., BRAUN, T. (1990), International collaboration in the sciences, 1981–1985. *Scientometrics,* 19 (1-2) : 3–10.

SCHUBERT, A., GLÄNZEL, W., BRAUN, T. (1989), Scientometric datafiles – a comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields 1981–1985. *Scientometrics*, 16 (1-6) : 3–478.