

# Short, Strong and Simple Mapping of Research Fields

Olle Persson

[olle.persson@soc.umu.se](mailto:olle.persson@soc.umu.se)

Inforsk, Department of Sociology, Umeå University, SE 901 897 Umeå, Sweden

## Abstract

Even within a narrowly defined research speciality, the citations among its papers may form a dense and complex network. In this paper a new mapping approach is suggested and tested on a network of papers on co-citation analysis. One way of mapping the dynamics of the field is to reduce the number of direct citation links by successively minimizing the citation year lag. Another approach is to remove the weakest links from the network. When the two approaches are combined, keeping only short and strong citation links, interesting decompositions of the original citation network was found.

## Introduction

Over the years quite a number of methods to map research specialities have been suggested, and reviewed by Morris & Martens (2008) and Bar Ilan (2008). There is now a set of different similarity measures and mapping techniques in use (Schneider & Borlund 2007a, 2007b), and some of them heavily debated (Ahlgren et al 2003, White 2003a, Bensman 2004). White (2003b) used a pathfinder algorithm to layout the structure of the information science domain. To study development over time Garfield (2004) has integrated a main-path analysis, inspired by Hummon & Doreian (1989), in his HistCite software, which has been used by Lucio-Arias & Leydesdorff (2008) and others. New advanced algorithms for clustering complex networks have been applied to citation data (Rosvall & Bergstrom 2008).

More attention should be given to mapping of direct citation relationships since they form the basis for the other, more fashionable, indirect relationships based on them, such as co-citation and shared references. Building on the tradition set out by Price (1965) this paper will focus on the mapping of research fronts using quite simple and straightforward ways of reducing a citation network.

One way of mapping the dynamics of a field is to reduce the number of direct citation links by the citation time lag. This was what Price (1965) suggested to identify the research front of a speciality. If we successively shorten the citation time lag we could hope to find one or several components linked by papers published in the same year range. As suggested by Persson (2010), direct citations could also be weighted by the number of indirect citations coupled to it. The idea is that if paper A cites paper B, that citation would be stronger if both papers are co-cited by other papers or share citations to other papers. This measure is called Weighted Direct Citations (WDC).

## Data

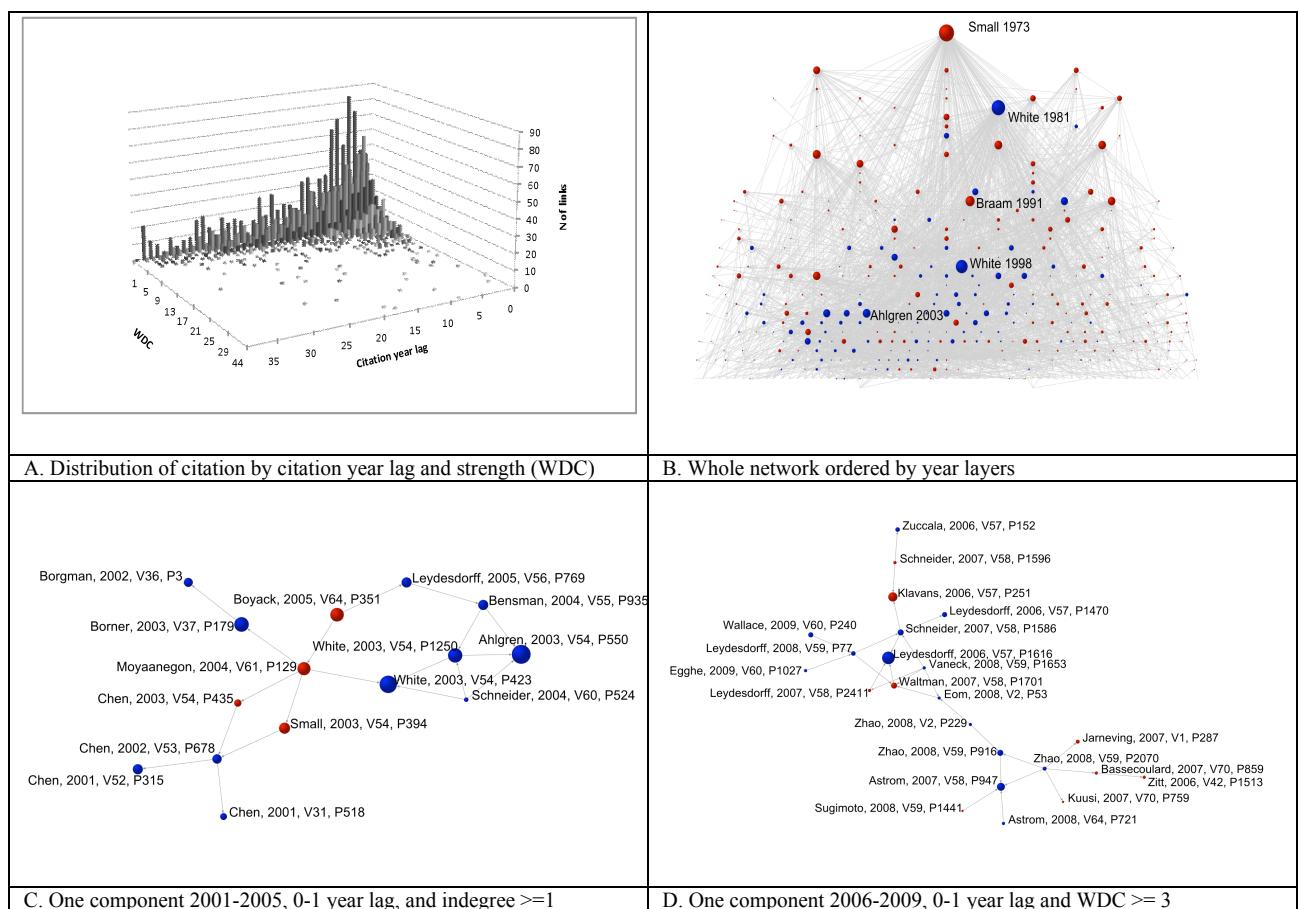
This study is based on a download from Web of Science (updated 2011-03-04) of 569 articles & proceedings papers from SCI, SSCI, A&HCI, and searched by “co-citation\* OR cocitation\*\*” in topic. The Bibexcel software (Persson 2011) was used to indentify 2895 direct citation links among the papers, as well as for calculating the citation year lag and WDC for each citation link. To identify papers on author co-citation (ACA), all papers were classified as either ACA or not if the string “author co-citation” or “author cocitation” was found somewhere in the downloaded full record. The Pajek software was used to draw maps.

## Results

Figure 1A shows the distribution of citation links by citation year lag and WDC. About 40 percent of all links are within five years, and 21 percent have a WDC of five or more. When

the whole network is ordered in year layers (Figure 1B), a quite clear picture emerges with major papers appearing in big red and blue circles. On the top we find the paper by Small (1973) that stated the whole field of co-citation studies, and below it the ACA papers by White & Griffith (1982) and further down White & McCain (1998). Two other significant papers are indicated by their first author, Braam et al (1991) and Ahlgren et al (2003).

In Figure 1C, the citation year lag is set to max one year and shows one of several components. The right part is centred on the Ahlgren (2003) and White (2003) dispute regarding the use of Pearson's correlation, and the left part contains other mapping studies from the same period. Figure 1D has one of the components that emerge when WDC is set to  $>=3$ . It has 23 fairly recent papers on mapping, covering parts of the area at the bottom of Figure 1B.



**Figure 1. Maps of papers on co-citation studies.**  
Blue circles= ACA studies, and red circles=other studies.

## Conclusions

It appears that a combined approach, reducing the citation network to short and strong citations, gives an adequate view of research fronts and subfield dynamics. When the two methods are used separately, they also produce relevant mappings. Above all, this approach shows that it is possible to decompose a complex citation network using quite straightforward elimination of citation links.

## References

- Ahlgren, P., Jarneving, B., & Rousseau, R. (2003). Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology*, 54(6), 550-560.
- Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century - A review. *Journal of Informetrics*, 2(1), 1-52.
- Bensman, S. J. (2004). Pearson's r and author cocitation analysis: A commentary on the controversy. *Journal of the American Society for Information Science and Technology*, 55(10), 935-935.
- Braam, R. R., Moed, H. F., & Vanraan, A. F. J. (1991). Mapping of science by combined cocitation and word analysis .1. Structural aspects. *Journal of the American Society for Information Science*, 42(4), 233-251.
- Garfield, E. (2004). Historiographic mapping of knowledge domains literature. *Journal of Information Science*, 30(2), 119-145.
- Hummon, N. P., & Doreian, P. (1989). Connnetivity in a citation network - the development of DNA theory. *Social Networks*, 11(1), 39-63.
- Lucio-Arias, D., & Leydesdorff, L. (2008). Main-path analysis and path-dependent transitions in HistCite (TM)-based historiograms. *Journal of the American Society for Information Science and Technology*, 59(12), 1948-1962.
- Morris, S. A., & Van der Veer Martens, B. (2008). Mapping research specialties. *Annual Review of Information Science and Technology*, 42, 213-295.
- Persson, O. (2011). Bibexcel - A toolbox for bibliometriicians: Inforsk, Department of Sociology, Umeå university, Sweden, www.umu.se/inforsk/Bibexcel.
- Price, D. J. D. (1965). Networks of scientific papers. *Science*, 149(3683), 510-&.
- Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, 105(4), 1118-1123.
- Schneider, J. W., & Borlund, P. (2007). Matrix comparison, part 1: Motivation and important issues for measuring the resemblance between proximity measures or ordination results. *Journal of the American Society for Information Science and Technology*, 58(11), 1586-1595.
- Schneider, J. W., & Borlund, P. (2007). Matrix comparison, part 2: Measuring the resemblance between proximity measures or ordination results by use of the mantel and procrustes statistics. *Journal of the American Society for Information Science and Technology*, 58(11), 1596-1609.
- Small, H. (1973). Cocitation in scientific literature - New measure of relationship between 2 documents. *Journal of the American Society for Information Science*, 24(4), 265-269.
- White, H. D. (2003). Author cocitation analysis and Pearson's r. *Journal of the American Society for Information Science and Technology*, 54(13), 1250-1259.
- White, H. D. (2003). Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology*, 54(5), 423-434.
- White, H. D., & Griffith, B. C. (1981). Author cocitation - a literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32(3), 163-171.
- White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science*, 49(4), 327-355.