Journal of Informetrics 4 (2010) 415–422

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Identifying research themes with weighted direct citation links

Olle Persson

Inforsk, Department of Sociology, Umeå University, SE-901 87 Umeå, Sweden

ARTICLE INFO

Article history:
Received 27 November 2009
Received in revised form 15 February 2010
Accepted 22 March 2010

Keywords:
Weighted direct citation links
Shared inlinks and outlinks
Mapping
Library and information science

ABSTRACT

In this study direct citations are weighted with shared references and co-citations in an attempt to decompose a citation network of articles on the subject of library and information science. The resulting maps have much in common with author co-citation maps that have been previously presented. However, using direct citations yields somewhat more detail in terms of detecting sub-domains. Reducing the network down to the strongest links of each article yielded the best results in terms of a high number of clusters, each with a substantial number of articles similar in content.

1. Introduction

Direct citations, shared references and co-citations have been used as similarity measures to detect past and emerging research themes (Chen, 2006; Small, 2006; Small & Upham, 2009; Zhao, 2006). Shibata, Kajikawa, Takeda, and Matsushima (2009) found that direct citations performed better in detecting research themes earlier compared to co-citations, the main explanation being that it takes time to build up co-citations. They also found that papers connected by direct citations had the strongest clustering tendency suggesting that they are more similar in content compared to those connected by shared references or co-citations.

However, assuming that a direct citation is a good enough measure of similarity is perhaps too simplistic. References in a paper may vary considerably in terms of similarity to the topic of the citing paper. In a citing paper, the cited papers can be ranked by the number of references they have in common with the citing paper. It is reasonable to assume, then, that a direct citation link between two papers is stronger, in terms of similarity, with the more shared references they have. Furthermore, we can assume that a direct citation becomes stronger if the two papers are co-cited by other papers. Then, the strongest direct citation links would be those that share many references, and are frequently co-cited. This is partly similar to the idea that citation links form diamond lattices as the basic building blocks of a citation network (Egghe & Rousseau, 2002; Fang & Rousseau, 2001).

The suggestion here is to integrate direct citations with shared references and co-citations into one measure of citation strength, which we call *weighted direct citations* (*WDC*). If we weight a citation link with indirect citations we might also come to terms with the topic drift that is inherent in reference lists. If the citing and cited paper of a direct citation link neither share a reference nor are being cited together by other papers, this might imply that the cited paper is out of topic of the citing paper.

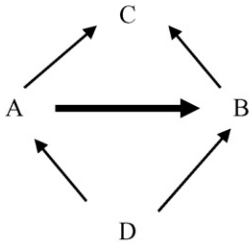
In this study we are analyzing citations among papers within a given set of papers, which mean that we do not study all shared references or all co-citations. It would not constitute a major problem to find all shared references since that data would be available once we have full reference information of a given set of papers. On the other hand,

E-mail address: Olle.Persson@soc.umu.se.

1751-1577/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2010.03.006

finding all co-citations would require access to all papers in which they are cited, which is a much more demanding task.

The diagram below shows how the measure WDC is calculated. The direct citation link from paper A to paper B is strengthen by the fact that A and B cite C, and D cites A and B. In this case WDC gets the value 3, which is the sum of one direct link and two shared links, and this is indicated by a thicker arrow from A to B. For each additional shared reference or co-citation the WDC will grow by 1.



Some papers cite substantially more papers than other papers, and some papers are considerably more cited than others. Therefore, one could normalize a given shared reference by the number of citations to that particular paper. For example, if C has received 10 citations we could take the inverse (1/10) as a normalized shared reference to C. Similarly, for a given co-citation we could take the inverse of the number of papers that the citing paper cite, e.g. if D cites 5 papers the normalized co-citation would be 1/5. Then, in our example the *normalized weighted direct citation* (*NWDC*) for A citing B is 1 + 1/10 + 1/5 = 1.3.

There are other approaches to assigning weights to citation links. The "search path count" method means that a citation link becomes stronger the more frequently it is passed when following citations paths through the citation graph (Hummon & Doreian, 1989). This has been integrated into HistCite software (Garfield, 2004) and used in a study by Lucio-Arias and Leydesdorff (2008). It is reasonable to assume that citation weight is positively correlated with the measure suggested here, since citation links that are strongly connected backward and forward will frequently be part of citation paths. On the other hand, citation path weights are indications of information flows rather than similarity. Typically, a link that is a bridge leading from one major cluster to another will get high citation weights without sharing references or being co-cited.

Decomposing networks is a prevailing challenge in bibliometrics. Even if we focus on a narrow subject the number of nodes and links must be reduced in order to be able to identify, read and understand traditional and emerging research themes. Networks of papers citing each other are often much more complex compared to aggregated structures (i.e. by authors, journals, and institutions). On the other hand, networks of papers offer more detail and are much closer to actual research activities.

In an author co-citation network, authors are forced into only one location, although they might be active in different sub-domains. Klavans, Persson, and Boyack (2009) launched the idea of multi-point location of cited authors by linking co-cited pairs. The method seems to position some authors in several relevant sub-domains, but it is still difficult to decipher the networks without looking more closely at the papers from which the author locations have been generated.

We can apply different strategies to decompose a directed citation network among papers. One approach is to remove links by setting a threshold for the strength of the citation links. In order to radically remove more links, we can allow each paper to be represented by its strongest link. Furthermore, papers that are not cited, or have few citations could be excluded. Moreover, although it is not recommended here, one could also remove self-citing links, defined by overlapping author sets, to avoid clusters of papers dominated by the same authors or groups. However, removing self-citations would probably reduce the citation network of a research theme too sharply.

The approach taken in this paper is to apply some of the criteria discussed above to reduce a directed citation graph in order to explore what kind of research themes emerge. Of course, this is a fine-tuning exercise; we will arrive at different results depending on how the criteria are set and different interpretations depending on who reads the results.

2. Data and methods

Knowledge about the research field under study is more or less a necessity for validating the results. Therefore, a set of journals familiar to the author of this paper was selected. The data for this study is a set of downloaded records from *Web of Science*™, covering papers from 13 journals in the field of library and information science (LIS), see Table 1. This journal set was chosen to enable comparisons with the mapping of the field made by Zhao and Strotmann (2008) and White and McCain (1998). Only genuine articles were selected. Review papers were excluded since they have an indirect role in relation to actual research, and a strong tendency to structure the network because they cite a lot of other papers as well as being cited by many. This also implies that all papers from *Annual Review of Information Science and Technology* had to be excluded, although some papers have been classified as articles. The period covered is 1945 up to September 2009. Compared to earlier studies, this opens up a longer time perspective on change patterns. However, some journals are not covered over the whole period, either because of late start or late inclusion in Web of Science

Table 1
Articles from 12 journals in library and information science.

Journal	No. of articles	Coverage from year
Electronic library	576	1984
Information processing and management	1431	1975
Information technology and libraries	515	1982
Journal of documentation	875	1945
Journal of information science	959	1979
Journal of the American society for information science	1626	1970
Journal of the American society for information science and technology	838	2001
Library and information science research	424	1984
Library resources and technical services	1074	1957
Proceedings of the American society for information science	803	1972
Program-automated library and information systems	261	1979
Scientometrics	1580	1978
Total	10,962	

Note: *Annual Review of Information Science and Technology* is excluded since it only contains review papers. Still, the document type "article" appears 141 times in the Web of Science records.

Bibexcel software was used to prepare the network data (Persson, Danell, & Wiborg Schneider, 2009), and then Pajek was used for decomposing the networks, and drawing the networks using the Kamada–Kawai layout (Pajek, 2009). Duplicate records and papers without cited references were removed.

Direct citation links between the papers were identified using a search key consisting of the cited first author's last name, year, volume, and start page. Shared references and co-citations were calculated from these direct citation links. All in all 27.841 direct citation links were identified, connecting 8.166 papers.

Table 2 shows the data for one paper by Ahlgren, Jarneving, and Rousseau (2003). Reading the full text, this article is a critique of the author co-citation metrics used in White and Griffith (1981) and White and McCain (1998). These two articles have the highest weighted citations (WDC) with respect to the articles that Ahlgren's paper cites. In terms of normalized values (NWDC), the article by Persson (2001) scores higher than White and Griffith (1981), which is less reasonable since Persson discusses all-author counting rather than co-citation metrics. If we look at the articles that cite Ahlgren's article, White (2003) is by far the strongest link and it is a direct response to the critique in Ahlgren's article. On the whole, it appears that WDC and NWDC give sound weights to the direct citation links, which otherwise would have the same value. Pearson's correlation between WDC and NWDC is 0.92, which means that using normalized values will have small effects on the network structure.

Table 2
Data example for one article: Ahlgren et al. (2003).

Citations from Ahlgren et al. (2003)			Citations to Ahlgren et al. (2003)		
Cited article	WDC	NWDC	Citing article	WDC	NWDC
White, 1981, V32, P163	10	1.67	White, 2003, V54, P1250	16	2.99
White, 1998, V49, P327	9	1.99	Rousseau, 2004, V55, P513	6	1.86
Persson, 2001, V50, P339	8	1.83	Zhao, 2007, V58, P1285	4	1.75
Marion, 2001, V52, P297	4	1.75	Zuccala, 2006, V57, P152	4	1.75
Bayer, 1990, V41, P444	4	1.43	Leydesdorff, 2004, V60, P159	4	1.18
Lin, 2000, V47, P143	3	1.50	Leydesdorff, 2005, V56, P769	4	1.18
Eom, 1996, V47, P941	3	1.17	Leydesdorff, 2004, V55, P991	3	1.17
Dore, 2001, V52, P763	2	1.25	Leydesdorff, 2004, V60, P371	3	1.15
Mccain, 1991, V42, P290	2	1.25	Leydesdorff, 2005, V56, P1469	3	1.15
Mccain, 1995, V32, P153	2	1.25	Jarneving, 2005, V65, P245	2	1.25
Ojasoo, 1999, V45, P81	2	1.25	Musgrove, 2003, V58, P657	2	1.25
Braam, 1991, V42, P233	2	1.13	Schneider, 2007, V58, P1586	2	1.25
Dore, 1996, V47, P588	2	1.08	Wagner, 2005, V62, P3	2	1.25
Braam, 1991, V42, P252	1	1.00	Zhou, 2007, V58, P223	2	1.14
Coulter, 1998, V49, P1206	1	1.00	Leydesdorff, 2007, V71, P391	2	1.07
Todorov, 1992, V23, P319	1	1.00	Kostoff, 2007, V33, P21	1	1.00
			Leydesdorff, 2006, V57, P1470	1	1.00
			Leydesdorff, 2006, V57, P1616	1	1.00
			Leydesdorff, 2006, V67, P231	1	1.00
			Leydesdorff, 2007, V58, P1303	1	1.00
			Leydesdorff, 2007, V58, P25	1	1.00
			Leydesdorff, 2007, V70, P693	1	1.00
			Waltman, 2007, V58, P1701	1	1.00
			Vaughan, 2006, V57, P1178	1	1.00
			Zhao, 2006, V42, P1578	1	1.00
			Zhou, 2007, V72, P185	1	1.00
			Zuccala, 2006, V57, P1487	1	1.00

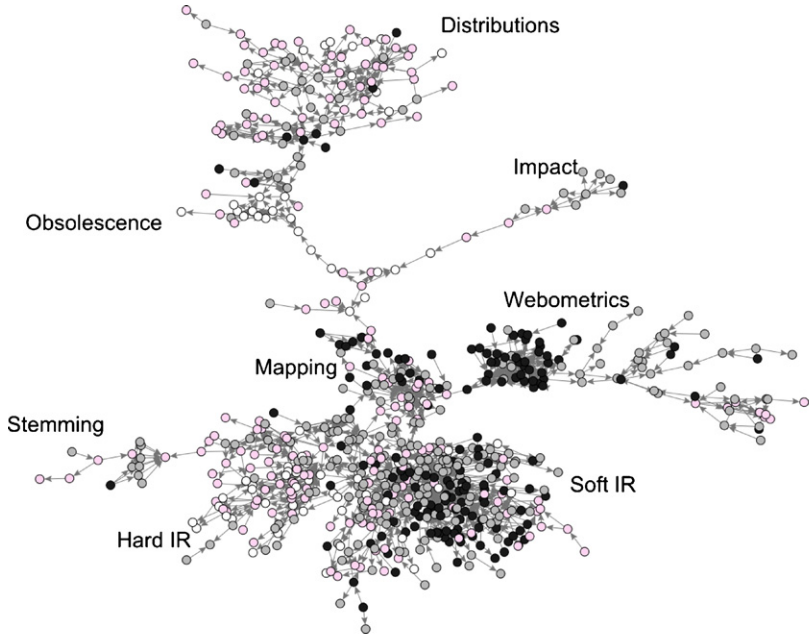


Fig. 1. Network reduced to the main component. *Note*: Generated from a network with all papers and links included. Reduction made using at least five inlinks and WDC > 4. Grey scale indicates decade: 1945–1979 = white; 1980–1989 = light grey; 1990–1999 = dark grey; 2000–2008 = black.

3. Results

In order to make a map that captures the overall structure of the field, we needed to reduce the considerably large citation graph with the help of Pajek. When limiting articles to those with at least five citations (indegree), and then removing links with lower line values than five (WDC < 5), 1.626 articles remained. The largest component of that network was extracted containing 681 articles. The remaining components have less than 10 articles.

The map of the main component shown in Fig. 1 was constructed using the Kamada–Kawai layout. The basic divide between informetrics and information retrieval is, in many respects, similar to author based co-citation maps of LIS presented by White and Griffith (1981), Persson (1994), White and McCain (1998), Astrom (2007) and Zhao and Strotmann (2008). Webometrics is the youngest sub-domain with articles mostly from year 2000 or later. This is indicated by black circles. Mapping and soft IR are still active subfields.

The hard IR sub-domain seems to have few articles from the last 10 years. For this study we have chosen a limited set of journals within library and information science. Some fields of study will obviously be excluded already from the start, and others may disappear simply because papers within a domain have moved to other journals. This might very well be the reason behind the decline of hard IR-research.

It is quite interesting to find obsolescence studies (e.g., the aging of literature) that date back to the 1970s, such as articles by Line (1970), Line and Sandison (1974) and Sandison (1971), and the stemming cluster which refers to the algorithm for suffix stripping, which was developed by Porter (1980). The long time window is probably the reason why these two domains show up.

Since the map is cluttered with links, it was reduced to be able to read it in more detail. The first step was to find a threshold for WDC that split the network in several fairly large components. If we start with the network in Fig. 1 and raise the strength from 5 to 8 we get five components with least 20 articles (Fig. 2). These components are what we really could expect from the first map, where they make up the largest sub-domains. They all seem to fit well with the general view of the field, and the time dimension comes out very clear.

The loss of articles appears to be substantial, and raising the threshold of WDC further would result in even smaller clusters and fewer articles. One could expect that this might be due to the fact that we initially reduced the papers to those having received at least five citations from within the network. However, if we start over with the whole network of 8.166 articles and only reduce by WDC, the network starts to fall apart at the same level of strength, and just about the same articles and clusters are left. If we use NWDC instead we get essentially the same effect, although it allows for reduction in smaller steps.

In order to decompose the network without losing too many articles, we can try another approach. First, only the strongest link of each article is accepted. All links are sorted by WDC as the first sort key and NWDC as the second. If an article has several links to other articles with the same highest number of WDC, we will select the link with the highest NWDC. However, if the strongest links also have the same normalized value for a given article, the selection of the strongest link will depend

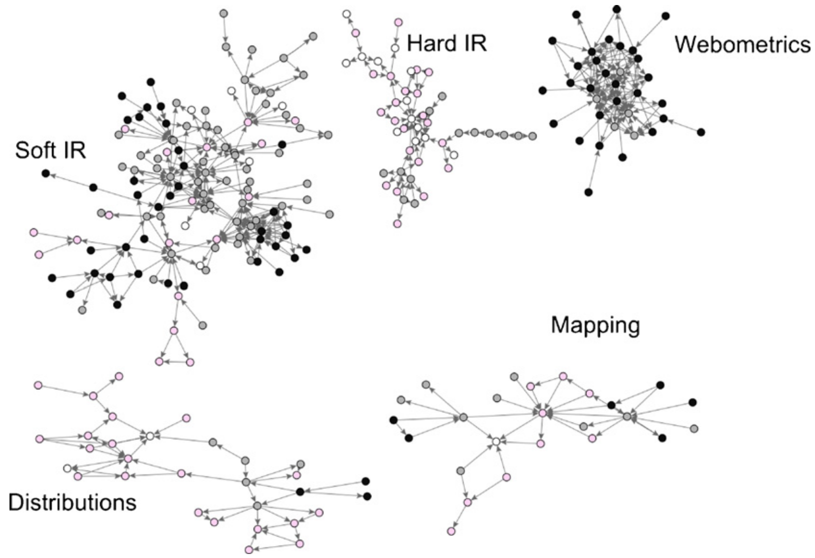


Fig. 2. Network in Fig. 1 further decomposed. *Note*: Reduced to links with WDC > 7, and minimum component size set to 20.

on an arbitrary sort order. In order to handle this problem, one could study the robustness of the clustering result when the sort order for the ties of a given paper is reversed.

Second, we require that each of the selected strongest links must have a weighted directed citation of at least 2. This will maintain a much larger set of articles, while drastically reducing the number of links. The result of this method is shown in Fig. 3. Here the network falls into a number of components, out of which the largest 22 with at least 25 papers are displayed.

In order to study the tie problem, we used a reversed sort order of ties, and the clustering result was quite similar. The total number of clustered articles, in clusters with minimum 25 articles, increased by 34 articles, 1114 compared to 1080. 19 of the 22 clusters were identical. Cluster number 4 got one additional article, number 6 three and number 20 four. Cluster 11 got one less. The remaining 27 articles formed an extra cluster having at least 25 articles.

The clusters in Fig. 3 differ in size and form. Several of the largest clusters have more than one central node, each with its own backyard. Some clusters have more recent articles than others. To tell more about the research themes one really have to take a closer look on each of them. Here we will study a few examples more closely.

Similarity is the basis of this strongest link approach since the strongest links are those in which the articles have the highest number of shared references and co-citations. Still, one might use WDC to reduce the number of papers when mapping a component. This has been done in Fig. 4, which covers papers in mapping. The initial 101 papers were reduced to 88 when the strongest links had to have a WDC value of at least 3. At this level we can actually start to read the map in detail. It appears to have two parts, and three central nodes. White and Griffith (1981) introduces the author co-citation method

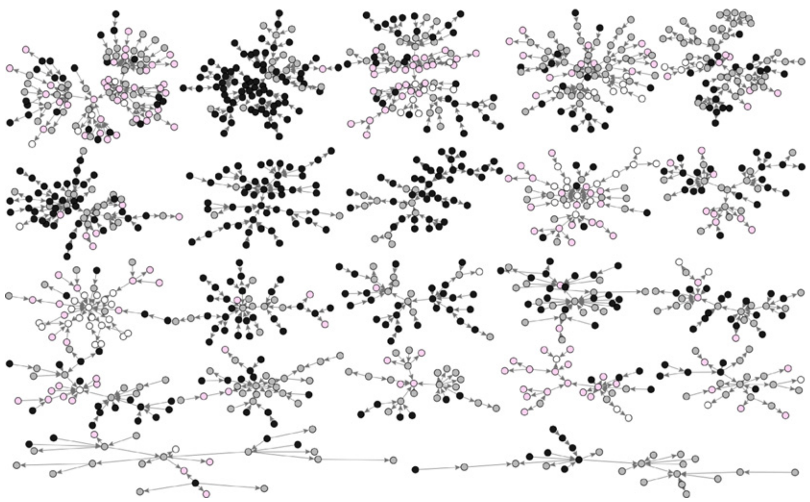


Fig. 3. Network components based on strongest links. *Note*: Generated from a network with only the strongest link of a paper included. WDC > 1, and minimum component size set to 25.

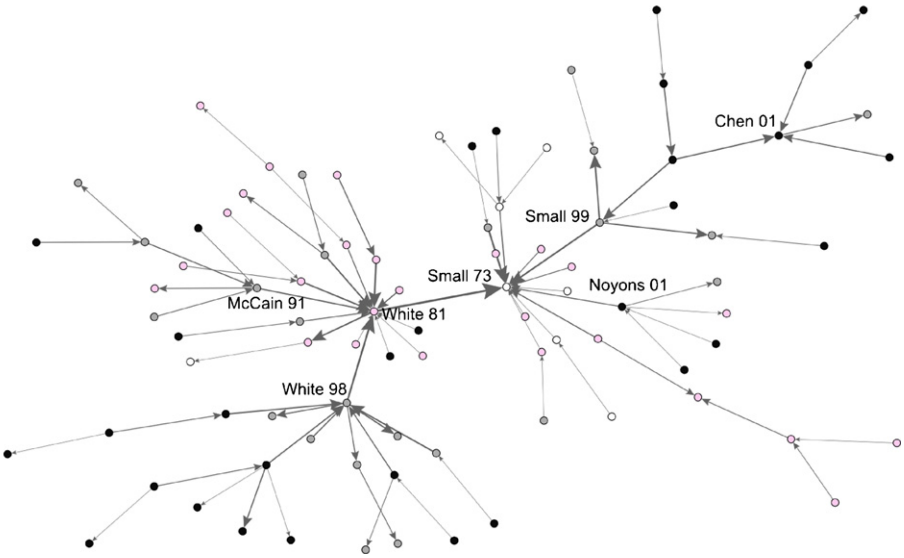


Fig. 4. The mapping component from Fig. 3. *Note*: Links removed with WDC < 3. Size of lines and arrows = ln(WDC).

in a study of library and information science, and White and McCain (1998) is a follow up study of the same field. McCain (1991) is a journal co-citation study of economics. The left part deals primarily with author co-citation analysis, while papers connected to Small (1973) in the upper half more often deal with co-citations among papers. Visualization is the main focus of articles further up to the right (Chen, Paul, & O'Keefe, 2001; Small, 1999). Noyons (2001) discusses bibliometric mapping in general and its applications in a science policy context.

Interestingly, the papers by Ahlgren et al. (2003) and White (2003) are not part of this map, although we can see from Table 2 that Ahlgren's paper has strong links to articles by White on the map. Instead, they are part of a smaller component of eight articles discussing or using the correlation coefficient as a similarity measure (Fig. 5). Thus, in this particular case the strongest link approach created a sub-domain with quite strong topic similarity.

Information seeking makes up another component in Fig. 3 which is more closely mapped in Fig. 6. There are two parts. One is composed of articles citing Kuhlthau (1991): *inside the search process* and the other is centred on Wilson (1999): *models in information behaviour research*.

The largest cluster in Fig. 3 represents the cognitive perspective on information retrieval. From Fig. 7 we can see that it has at least five central nodes. Ingwersen (1996) made a broad overview of cognitive perspectives in information retrieval research, and his articles has a strong link to the early article on Anomalous States of Knowledge by Belkin, Oddy, and Brooks (1982). Bates (1986) has also a strong link to Belkin, but deals with design issues for subject searching of online catalogues. Chen and Dhar (1991) is linked to Bates and discuss how cognitive processes could form the basis for design of information retrieval systems. Larson (1991) reports on a decline of subject searching in online catalogues.

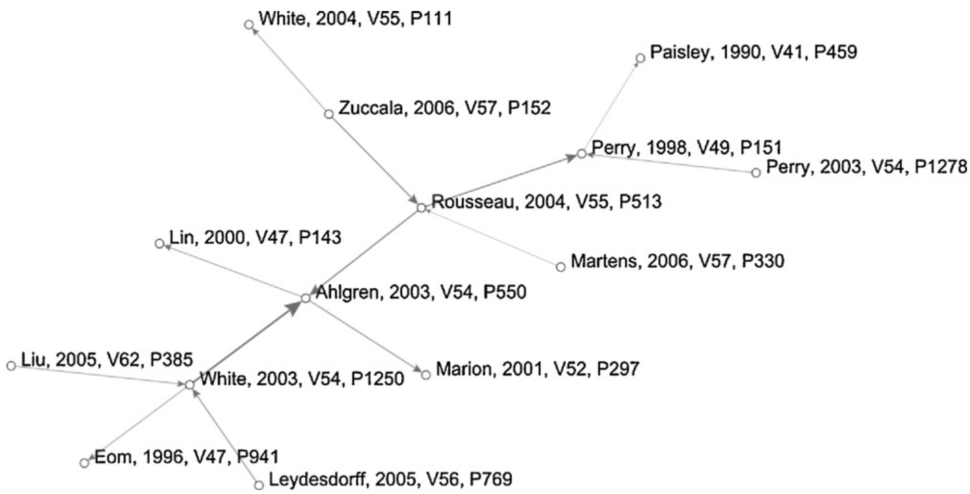


Fig. 5. The dispute on using correlations as a similarity measure. *Note*: Strongest links and with WDC > 1. Size of lines and arrows = ln(WDC).

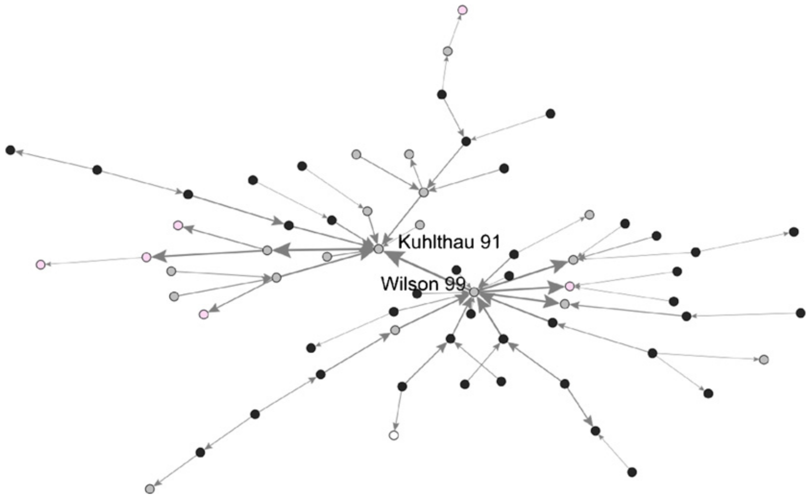


Fig. 6. The information seeking component from Fig. 3. *Note*: Links removed with WDC < 3. Size of lines and arrows = ln(WDC).

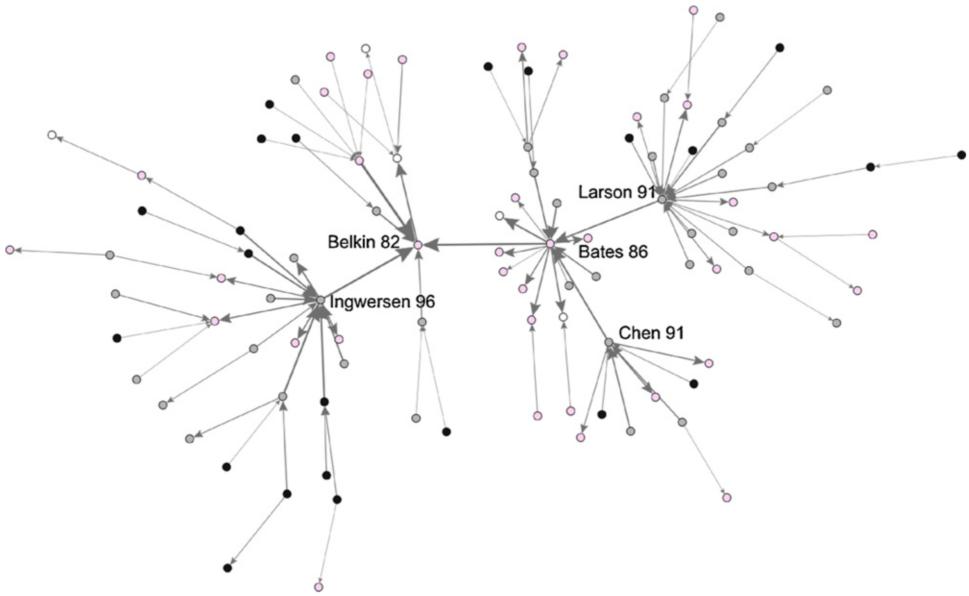


Fig. 7. The cognitive veiwpoint component from Fig. 3. *Note*: Links removed with WDC < 3. Size of lines and arrows = ln(WDC).

Of course, all clusters and all nodes in them do not have the same strength in terms of similarity and coherence. To fully validate all of them would require much more time, expertise and probably another publication form. On the other hand, comprehensiveness has not been the goal of this article. The examples given are significantly clear and interesting to motivate further research.

4. Conclusions

Using shared references and co-citations as a way of weighting the strength of direct citations appears to be a useful tool for decomposing a network of papers. In scientific disciplines such as library and information science, the citation network is quite large in terms of papers and citations among them. It is possible to obtain meaningful sub-domains by removing links below a certain weight, and by removing less frequently cited papers. However, network complexity is still very high; raising the threshold further will result in a great loss of papers.

The strongest link approach, which only allows for the strongest links of each paper, produced interesting results with quite many and meaningful sub-domains of research. However, there is a need for further testing of clustering routines, which might make better use of the proposed measure. Above all, the examples presented here have demonstrated that weighted direct citations (WDC) contribute to the structuring and interpretation of the research themes.

The weighting approach could also be applied to co-citation analysis since a co-citation should receive a higher value if the co-cited papers cite each other and if they share references. The same kind of reasoning could be applied if the focus is on shared references, or bibliographic coupling. Another suggestion for future research is to incorporate all cited references in counting shared references as well as finding all co-citations to a pair of papers. This paper has demonstrated that weighted direct citation analysis offers an interesting alternative to the co-citation approach that has dominated the field so far.

Acknowledgements

The author is thankful for comments from two anonymous reviewers and Rickard Danell in my own department.

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.
Astrom, F. (2007). Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990–2004. *Journal of the American Society for Information Science and Technology*, *58*(7), 947–957.
Bates, M. J. (1986). Subject access in online catalogs—A design-model. *Journal of the American Society for Information Science*, *37*(6), 357–376.
Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information-retrieval. 1. Background and theory. *Journal of Documentation*, *38*(2), 61–71.
Chen, C. M. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, *57*(3), 359–377.
Chen, C. M., Paul, R. J., & O'Keefe, B. (2001). Fitting the jigsaw of citation: Information visualization in domain analysis. *Journal of the American Society for Information Science and Technology*, *52*(4), 315–330.
Chen, H. C., & Dhar, V. (1991). Cognitive process as a basis for intelligent retrieval-systems design. *Information Processing & Management*, *27*(5), 405–432.
Egghe, L., & Rousseau, R. (2002). Co-citation, bibliographic coupling and a characterization of lattice citation networks. *Scientometrics*, *55*(3), 349–361.
Fang, Y., & Rousseau, R. (2001). Lattices in citation networks: An investigation into the structure of citation graphs. *Scientometrics*, *50*(2), 273–287.
Garfield, E. (2004). Historiographic mapping of knowledge domains literature. *Journal of Information Science*, *30*(2), 119–145.
Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network—The development of DNA theory. *Social Networks*, *11*(1), 39–63.
Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation*, *52*(1), 3–50.
Klavans, R., Persson, O., & Boyack, K. W. (2009, July). Coco at the copacabana: Introducing co-cited author pair co-citation (Coco) analysis. In *Paper Presented at the 12th International Conference of the International-Society-for-Scientometrics-and-Informetrics* Rio de Janeiro, Brazil.
Kuhlthau, C. C. (1991). Inside the search process—Information seeking from the users perspective. *Journal of the American Society for Information Science*, *42*(5), 361–371.
Larson, R. R. (1991). The decline of subject searching—Long-term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science*, *42*(3), 197–215.
Line, M. B. (1970). Half-life of periodical literature—Apparent and real obsolescence. *Journal of Documentation*, *26*(1), 46.
Line, M. B., & Sandison, A. (1974). Progress in documentation—Obsolescence and changes in use of literature with time. *Journal of Documentation*, *30*(3), 283–350.
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.
McCain, K. W. (1991). Mapping economics through the journal literature—An experiment in journal cocitation analysis. *Journal of the American Society for Information Science*, *42*(4), 290–296.
Noyons, E. (2001). Bibliometric mapping of science in a science policy context. *Scientometrics*, *50*(1), 83–98.
PAJEK. (2009). *Pajek: Program for large network analysis*. http://vlado.fmf.uni-lj.si/pub/networks/pajek/
Persson, O. (1994). The intellectual base and research fronts of JASIS 1986–1990. *Journal of the American Society for Information Science*, *45*(1), 31–38.
Persson, O. (2001). All author citations versus first author citations. *Scientometrics*, *50*(2), 339–344.
Persson, O., Danell, R., & Wiborg Schneider, J. (2009). How to use Bibexcel for various types of bibliometric analysis. In *Celebrating scholarly communication studies: A festschrift for Olle Persson at his 60th birthday*. International Society for Scientometrics and Informetrics., pp. 9–24.
Porter, M. F. (1980). An algorithm for suffix stripping. *Program-Automated Library and Information Systems*, *14*(3), 130–137.
Sandison, A. (1971). Use of older literature and its obsolescence. *Journal of Documentation*, *27*(3), 184–199.
Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2009). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science and Technology*, *60*(3), 571–580.
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.
Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, *50*(9), 799–813.
Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics*, *68*(3), 595–610.
Small, H., & Upham, P. (2009). Citation structure of an emerging research area on the verge of application. *Scientometrics*, *79*(2), 365–375.
White, H. D. (2003). Author cocitation analysis and Pearson's. *Journal of the American Society for Information Science and Technology*, *54*(13), 1250–1259.
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.
Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, *55*(3), 249–270.
Zhao, D. Z. (2006). Towards all-author co-citation analysis. *Information Processing & Management*, *42*(6), 1578–1591.
Zhao, D. Z., & Strotmann, A. (2008). Comparing all-author and first-author co-citation analyses of information science. *Journal of Informetrics*, *2*(3), 229–239.