Mapping of Cross-Reference Activity Between Journals by Use of Multidimensional Unfolding: Implications for Mapping Studies

Jesper W. Schneider

jws@db.dk
Royal School of Library and Information Science, Fredrik Bajers Vej 7K, DK-9220 Aalborg S (Denmark)

Abstract
The present paper explores simultaneous modelling of cross-reference activity between journals by use of asymmetric proximities and multidimensional unfolding. We thereby model and map both citing and cited relations between journals in a common space. This enables a more comprehensive comparison of the journal's dual roles of citing and being cited in a reference network. We examine the set of journals most often used to delimit information science in mapping studies, in order to demonstrate the need for exploratory data analyses. The general findings of the study include the ability of unfolding analysis to identify specialized journals and the influence of publication activity in mapping studies.

Introduction
The present paper explores the modelling of cross-reference activity between journals by use of asymmetric proximities and multidimensional unfolding (Schneider, 2008). Our aim is to explore simultaneous modelling and mapping of both citing and cited relations between journals in a particular reference network. We thereby examine the different roles of journals, their "citing behaviour" and their "cited reception". This is essential as journals are perhaps the most important entity in bibliometric activities. Everything seems to depend on journals. Bibliometric data resides in publications and publications are assembled in journals (we do not consider proceedings in this paper). It is primarily journals that are included in citation indices, and it is journals which are indexed with subject categories in such indices. The scope of bibliometric analyses depend upon accession policies of the hosts of citation indices. Constructs such as Thompson Reuter's journal subject categories affect our conception of scientific discourse and to a strong degree bibliometric analyses. Most importantly for the present study, subject categories are very often used to delimit fields of study in bibliometric analyses. Some or all of the journals indexed in a subject category become the target for data collection. Within subject categories, journals can be vey different. Not all journals are scientific. Some are subject specific, others more general. Some have large annual publication rates others have considerably lower. Such factors influence bibliometric studies. In mapping studies, for example, we set out to explore the "intellectual structure" of a "discipline" (White & McCain, 1998). Characteristically the "discipline" is delimited by selection of journals from one or several subject categories. The aim is to uncover "structure" from the selected set of data. Notice, data structure is influenced by publication productivity of the selected journals. And more essentially, structure does not necessarily come from the data alone. Data may instead be in the constructs relied upon such as subject categories. If so, one could argue that we are representing "the world" of Thompson Reuters or Scopus, and not the "real" structures of science. The influence of constructs most likely varies with level of analysis. At the macro level of mapping, the accession policies of Thompson Reuters and Elsevier are the primary boundary. Journal classification structures can be identified independently of the imposed categories, creating new journal based constructs (e.g., Klavans & Boyack, 2005; Leydesdorff, 2006). The problem becomes more apparent at the meso and micro levels of mapping.
In this article we focus upon the well known and often mapped subject category of "Information Science & Library Science". White and McCain's (1998) modus operandi for defining information science from a limited set of journals listed in the Thompson Reuters

Proceedings of ISSI 2009, pages 443–454.
Edited by B. Larsen and J. Leta.

(former ISI) subject category for "Information Science & Library Science" has been adopted in several later studies (e.g. Åström, 2007; Zhao & Strotmann, 2008). With minor alterations, this core set of journals (and one proceeding) has provided the data when mapping the structures of information science.

We find it questionable that most mapping studies do not reflect upon their data selection process. Mapping studies of information science conveniently proceeds with the original journal set delimited by White and McCain (1998). Mapping of other domains apply a similar perfunctory approach where journal characteristics such as cross-reference patterns and productivity are not considered. Now it may very well be that the various mappings of information science are essentially "good" mappings, but of what? In order to interpret and evaluate maps we must reflect upon the chosen data before and after an analysis. This is a good and necessary empirical standard. In the midst of the booming methodical interest in science mapping, we seem to pay much less attention to more subtle methodological and theoretical issues. The present study therefore examines some of the tacitly accepted methodological norms of mapping studies. Modelling cross-reference activity between journals by use of asymmetric proximities and multidimensional unfolding can be considered an exploratory data analysis (EDA) where journal cross-reference data and publication activity are screened. Consequently the purpose of the present study is twofold: 1) more generally to explore the utility of unfolding analysis to characterize a set of journals considered for mapping studies; and 2) more specifically to examine the well-known set of journals implicitly delimiting information science. This leads to the following research questions:

- To what extent can asymmetric proximities and multidimensional unfolding model and map journal cross-reference activity in a restricted set of journals?
- What are the general and specific findings of the unfolding analyses?

The structure of the paper is as follows: The next section outlines the background and related research for the present analysis; the following section introduces the data selection and methods applied; the section after that outlines and discusses the results; and finally we conclude with a brief summary of the major findings.

Background

Journal reference networks are multidimensional, with citing and cited dimensions. In such networks, journals have different profiles according to their role as either citing or cited. The references actively given by authors in a journal determine the "behavioral profile" of the journal, whereas the more passive role of being cited creates a "profile of reception" for the journal. This entails that journals may have different roles and structural patterns in the citing and cited network respectively. Interestingly, these two patterns may differ considerably. The motivation behind the present study is to explore the utility of unfolding analysis to simultaneously model and map the multidimensional roles and structural patterns of journals in a reference network. This is a novel approach.

Cross-reference activity of journals is modeled in square asymmetric transaction matrices (Price, 1981; Noma, 1982). Cross-reference activity is asymmetric as the information flow expressed by references given by one journal to another, is not necessarily reflecting a mutual relationship. For a set of journals, it is possible to construct a matrix of cross-referencing activity, where the rows represent journals giving references *to* the other journals, and the columns represent the same journals receiving citations *from* the other journals. Notice, the diagonal in a square transaction matrix contains the number of self-references for each journal.

Modeling and mapping journal cross-reference transactions is not novel (e.g., Tijssen, De Leeuw and Van Raan, 1987; Leydesdorff, 2001; 2006). Common to these studies are that they model cross-structural relations in the transaction matrix, and subsequently find one position for a journal in the map, which reflects both its citing and cited dimension. The present paper is a continuation of these studies. However, we apply multidimensional unfolding in order to position journals simultaneously in both their citing and cited dimensions (Borg & Groenen, 2005). Multidimensional unfolding is a data analysis technique for two-way two-mode proximity data (Borg & Groenen, 2005). Distance-based multidimensional unfolding can be considered an extension of multidimensional scaling (MDS). In MDS models we have $d(x_i, x_j) = d(x_j, x_i)$, whereas in unfolding models we generally find $d(x_i, y_j) \neq d(x_j, y_i)$. Thus distances in unfolding are inherently asymmetric. Further, in MDS we must have $d(x_i, x_i) = 0$, whereas in general we have $d(x_i, y_i) \neq 0$ in unfolding. Thus, the unfolding model also allows modelling of the diagonal of a square matrix, unlike the MDS model. In theory, this implies that the unfolding model is able to reflect all three relational aspects present in a reference transaction matrix, citing, cited and self-citing relations. For an asymmetric matrix, MDU finds coordinates in a low-dimensional space for *two* sets of objects with a common quantitative scale. Until recently, unfolding models suffered from degenerate solutions in the majority of cases. A degenerate solution satisfies a given optimality criterion perfectly, while at the same time resulting in a substantively trivial, non-informative spatial representation (Borg & Groenen, 2005). Recently, solutions to the degenerate problem have been found in two novel unfolding algorithms, PREFSCAL (Busing, Groenen, & Heiser, 2005) and GENEFOLD (Van Deun et al., 2007) respectively. The next section presents the methods applied for the present study.

Methods & data selection

We make two similar multidimensional unfolding analyses, one with 10 journals and one with 16 journals. Otherwise procedures are identical. Our method consists of four steps: 1) collection of reference data for the chosen journals in Journal Citation Reports® (JCR); 2) construction of transaction matrices of cross-reference activity, 3) calculation of asymmetric proximities; and finally 4) multidimensional unfolding of the proximity matrices.

The data selection for the present analysis comes from the citing and cited journal transactions available from the JCR. We use the 2007 data for the subject category "Information Science & Library Science". Notice only references given or citations received by the journals two or more times are available from the JCR.

In order to examine the utility of multidimensional unfolding as an EDA tool for exploration of journal cross-reference patterns and productivity prior to mapping, we investigate the journal set of information science originally delimited by White and McCain (1998). Notice, *Program – Automated Library and Information Systems* is no longer indexed by Thompson Reuters. We also exclude *Proceedings of the American Society for Information Science* as data were not readily available in JCR. As such, the exclusions are not problematic for the present study. We are still able to explore the potentials of unfolding analysis to model and map cross-reference activity. To elaborate on the White and McCain journal set, we modify it for a further analysis with six additional journals. These journals are chosen on the basis of the similarity with the identified core journals of the first unfolding analysis. Remember, we do not claim that a specific journal set is invalid. Our focus is to explore journal cross-reference activity prior to their selection. The journals are listed in Table 1.

Table 1: Journals included in the two unfolding analyses; the journals delimited by White and McCain, and the six extra journals included in the second unfolding analysis.

Journals	Abbreviation
Information science	
Annual Review of Information Science and Technology	ARIST
Information Processing & Management	IPM
Journal of the American Society for Information Science and Technology	JASIST
Journal of Documentation	JDOC
Journal of Information Science	JIS
Library & Information Science Research	LIB INF SCI RES
Scientometrics	SCIENTOMETRICS
Library automation	
Electronic Library	ELECT LIBR
Information Technology and Libraries	INF TECHN LIBR
Library Resources & Technical Services	LIB RES TECH SER
Extra journals included in the modified study	
Aslib Proceedings	ASLIB
Information Research	INF RES
Journal of Librarianship & Information Science	JLIS
Knowledge Organization	KNOW ORG
Library Quarterly	LIB QUA
Library Trends	LIB TREND

Remember that the unfolding solution is based on the cross-reference activity among the chosen set of journals. Obviously, the journals give references to journals outside the set and receive citations from journals outside the set. The unfolding solution is therefore distinctive for the given set of journals.

Journal cross-reference activity data is represented in two asymmetric square transaction matrices, where rows and columns correspond to the journals giving references and journals receiving citations. The transaction matrices are transformed into simple proximity matrices of odds ratios of observed to expected cross reference activities, corrected for main effects in the data. Main effects reflect the tendency of some journals to have consistently higher frequencies than others. Our purpose is to model journal structure where the magnitude of individual journal reference activity does not dominate.

We denote frequencies f_{ij}, where i is the citing and j is the cited dimension. If journals give references to each other in a random fashion, we would expect the joint frequencies to satisfy the formula for the expected frequencies (e_{ij}) under independence:

$$e_{ij} = N\left(\frac{f_{i+}}{N}\right)\left(\frac{f_{+j}}{N}\right) = \frac{f_{i+}f_{+j}}{N}, \tag{1}$$

that is, the product of the estimated probability of citing and the estimated probability of being cited times the total number of references N (here, the + in the marginal totals replaces the index over which we have summed). As a measure of proximity to be used in the unfolding model, we define the odds of journal a_i giving a reference to journal b_j against what we expect under independence:

$$\rho(a_i, b_j) = \frac{f_{ij}}{e_{ij}} = \frac{Nf_{ij}}{f_{i+}f_{+j}}. \tag{2}$$

We are not able to present the matrices of odds ratios for the cross-reference activity between the journals in the present paper due to space limitations. In return, the matrices are available as Table 2 and 3 at www2.db.dk/jws/crossrefodds.htm. Note that ρ (a_i, b_j) = 1 if journal a_i gives references to journal b_j as expected according to the size of cross-reference activities of the two journals; consequently ρ (a_i, b_j) < 1 is less than expected according to size, and ρ (a_i, b_j) > 1 is greater than expected. To give an example, the odds that ELECT LIBR gives a reference to INF TECHN LIBR rather than exchanging references with one of the other journals

is 5 to 1. In the unfolding solution, ρ (a_i, b_j) < 1 will lead to a relatively large distance d (x_i, y_j) and ρ (a_i, b_j) > 1 to a relatively small distance d (x_i, y_j).

The asymmetric proximity matrices of cross-reference odds ratios are the basis for the multidimensional unfolding analysis. In the present study we apply the GENEFOLD algorithm (Van Deun et al., 2007). GENEFOLD performs an alternating least squares and iterative majorization procedure. The procedure starts with a random configuration of dimensionality and then alternates between transformation updates, configuration updates and regression weight updates. This is to ensure that the sequence of loss function values converges to a local minimum to provide an optimal configuration and avoid degenerate solutions. We apply an ordinal unfolding approach where distances in the map have a monotone relation with the odds ratios in the proximity matrices. The following section presents and discusses some of the characteristics findings of the multidimensional unfolding analyses.

Results

In multidimensional unfolding, locating the citing and cited positions of a journal, assessing the mutual distance between positions, and interpreting the positions relative to other nearby objects, indicates a journal's degree of resemblance in its different roles of citing and being cited, as well as its overall structural positions in the studied reference network. Journals with similar patterns, citing, cited or both, will tend to group. A good modelling solution will locate a journal's citing and cited positions according to the most dominant cross-reference proximities in their profiles. According to Heiser and Busing (2004), it is important to focus on the origin of the map and then identify in which direction citing odds ratios ≥ 1 go. Such an analysis reflects coherence and specialization among the studied journals – which journals are in the "core" and which ones are on the "fringe" of the set. Vague profiles (constant odds ratios below 1) will tend to isolate journals. Remember that the ratios are the odds of a row journal citing a column journal against the expected values under independence. For example, the odds that a publication in INF TECHN LIBR gives a reference to LIB RES TECH SER (in this particualr data set) rather than exchanging references with one of the other journals is 11 to 1. Notice that the citing profile characterizes the journal's reference activity in 2007 and that the cited profile characterizes how *all* previous publications in the journal was used in 2007. This can create two very different "images" of a journal. The present citing image may differ considerably from the cited image, if the present citing behaviour is not mutually reflected in the received citations. The latter can be the effect when a considerable number of citations go to older volumes of the journal. With this in mind, we now examine some of the characteristic findings of the unfolding solutions.

The results of the multidimensional unfolding analyses are presented in Figures 1, 2, 4 and 5. Figures 1 and 2 are the unfolding solution for White and McCain's core journal set, and Figures 4 and 5 are the unfolding solution for the extended journal set. First we examine the characteristic findings in the unfolding solution of White and McCain's core journal set. Subsequently we examine noticeable differences between this solution and the one obtained with the extended journal set.

Multidimensional unfolding of the White and McCain journal set.

The results are presented in Figure 1 and 2 below; Figure 2 illustrates the main citing directions for the collective groupings of the journals. Closed circles are the citing positions of the journals in 2007 and open circles are the cited positions of journals.

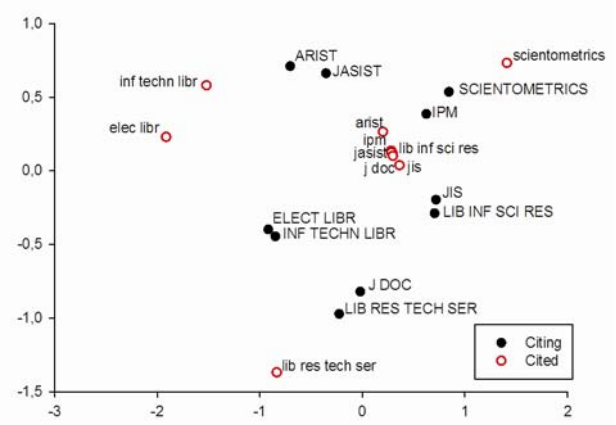

Figure 1: Multidimensional unfolding of 10 journals from the core set delimited by White and McCain.

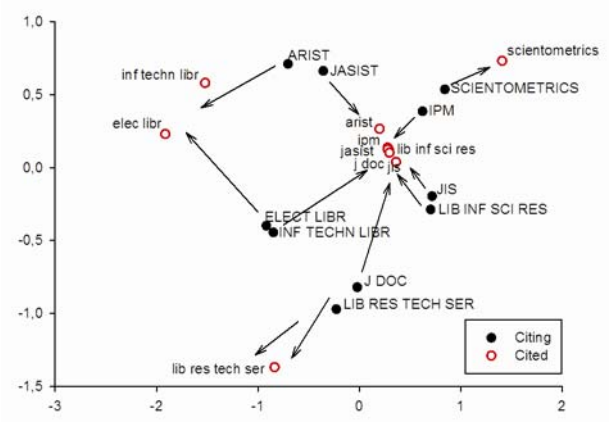

Figure 2: Main citing directions for collective groupings of journals in the White and McCain set.

Several collective groupings appear from the unfolding solution. Moving clockwise around the map, starting at twelve o'clock, we find the citing positions of ARIST and JASIST. The direction of their citing activity is similar: 1) towards the origin of the map, where a cited core is apparent, and 2) towards INF TECH LIBR and ELEC LIBR in the upper left fringe of the map. Both journals have several citing odds ratios close to or above 1 with other journals. ARIST seems to have a high odds ratio of citing INF TECHN LIBR. Among the rest, only ELECT LIBR and JASIST have citing odds ratios towards INF TECHN LIBR, i.e., ELECT LIBR with a very high odds ratio of 5.02, whereas JASIST is below 1.

At one o'clock we find the citing positions of IPM and SCIENTOMETRICS. IPM'S citing profile solely points inward to the core cited set in the origin of the map. The citing direction of SCIENTOMETRICS, conversely, points outward towards its own eccentric cited position. Notice all citing and cited odds ratios for SCIENTOMETRICS are well below 1 except its self-reference activity of 3.53; this determines its position. We discuss this important characteristic below.

At three o'clock we find the citing positions of JIS and LIB INF SCI RES. Again we see a dominant citing direction that points inward to the core set of cited journals.

At six o'clock we find the citing positions of JDOC and LIB RES TECH SER. The central citing odds ratios for JDOC in 2007 is split between an inward direction towards the cited core in the origin of the map, and outwards to the position of LIB RES TECH SER. Similar to SCIENTOMETRICS, the cited position of LIB RES TECH SER is eccentric on the fringe of the map. However, contrary to SCIENTOMETRICS several of the other journals do have high odds ratios of citing LIB RES TECH SER, i.e., ELECT LIBR 1.55, INF TECHN LIBR 10.94 and JDOC 3.25. In spite of that, the reason why LIB RES TECH SER'S cited position is on the fringe of the map, similar to SCIENTOMETRICS, is the extremely high self-reference odds ratio of 44.78. This is clearly so dominant that it positions LIB RES TECH SER far away from the core set of cited journals. The citing position of LIB RES TECH SER is just below JDOC. Besides its very high self-reference activity, LIB RES TECH SER does not have odds ratios ≥ 1 for any of the other journals in the set.

At eight o'clock, and closer to the origin of the map than JDOC and LIB RES TECH SER, we find INF TECHN LIBR and ELECT LIBR. Commonly these journals cite in three directions, inwards towards the core cited set in the origin, outwards towards their own cited positions in the upper left fringe of the map, and finally outwards toward LIB RES TECH SER in the lower left fringe of the map. We should note that the cited position of these journals at eleven o'clock is eccentric in the upper left fringe of the map. Noticeably, both journals have a number of odds ratios above 1 for citing the other journals in the set. They also both have high odds ratios of self-reference activity, and besides themselves, only ARIST and JASIST are likely to cite them

as expected. Even though we have corrected for the size of reference activity through main effects, this finding suggests that the activity for these journals is considerably lower than that of the other journals. We elaborate on this below.

Finally, at the origin of the map, its centre, we find very close to each other, the cited positions of ARIST, JASIST, IPM, JDOC, LIB INF SCI RES and JIS. These cited journals are no doubt a core in the set. Journals citing inwards towards this core set can be considered, to a varying degree, to have citing profiles that match the core. In other words, the visible circle of citing journals positioned around the core cited set in the centre should be evaluated for their citing directions. The more they point to the core the more they match the core. Of interest also is how far citing journals, which are also present in the cited core, are from the origin and whether their dominant reference activity points to the core set. ARIST and JDOC seem to waver a bit, as ARIST also cites outwards towards INF TECHN LIBR and ELECT LIBR, and JDOC towards LIB RES TECH SER. This indicates that the citing profiles for the journals in 2007 – the way the journals present themselves – are somewhat different from how authors in the other journals use them. The case of JDOC is informative. The cited position of JDOC is determined because other journals citing the core set cite JDOC with odd ratios ≥ 1; otherwise, they do not seem to reflect the citing profile of JDOC. When we investigate the cited distribution for JDOC in 2007, we can clearly see that a significantly larger share of citations go to older volumes of the journal, compared to the distributions of most other journals in the set. JDOC underwent considerable editorial changes around 2003-2004, which may have altered the citing profile and thus to some extent explains the currently marked difference in citing and cited profile.

Consequently, the cited positions of SCIENTOMETRICS, LIB RES TECH SER, ELECT LIBR and INF TECHN LIBR are eccentric, as they do not belong to the core. As pointed out by Heiser and Busing (2004), citing outwards, away from the origin, towards eccentrically positioned objects most likely reflects specialization. Notice that dimensions and scales are interpretable in unfolding analyses. If we examine the *x* and *y* axes in Figure 1, we can see that the cited positions of LIB RES TECH SER, ELECT LIBR and INF TECHN LIBR are positioned to the left of the centre, where LIB RES TECH SER is positioned at the lower negative end of the *y* axis and ELECT LIBR and INF TECHN LIBR at the upper positive end of *y* axis. Conversely, the cited position of SCIENTOMETRICS is positioned to right of the centre, in the upper positive end of the *y* axis. These opposite positions inform us that these journals when cited can be considered as specialized. Either only some of the journals tend to cite them, or they tend to cite only themselves in this set. This is very important when we consider which journals should be included in mapping studies. Before we discuss this we need to address the related issue of reference activity.

Reference activity

As stated in the introduction section, major differences in annual publication activities between journals selected for a mapping study will influence the mapping result. Reference activity is a mirror reflection of publication productivity. According to the core and scatter phenomenon in bibliometrics, we can assume that the more publications annually in a selected journal, the more likely it will strengthen the core cited entities for that particular journal (and eventually its authors). Consequently, the citing behaviour of journals with relatively high annual publication productivity will dominate mapping results, at least when selected together with journals that have lower annual publication productivity. Reference activity is to a certain degree also dependent on the length of reference lists in journal publications (in the present case ARIST produced only 15 publications in 2007 but still have a considerable references activity).

As indicated in the methods section, the asymmetric proximity measures applied in the present study remove main effects from the data. Recall that main effects reflect the tendency

of some journals to have consistently higher frequencies than others. Levelling out reference activity between journals means that those journals with a distinctively lower reference activity, compared to the others, will be more visible than otherwise. This enables modelling of cross-reference activity; however, it does not remove the basic challenge of reference activity. Figure 3 below illustrates the actual reference activity among the journals in the White and McCain set for 2007. What we see is a considerable variation in the journal set for 2007. JASIST is clearly the most active journal in the set giving 1790 references to journals in the set (including itself). Number two, with less than half of the activity of JASIST, but still considerably more than number 3 is SCIENTOMETRICS. The total reference activity for the whole set is 3397.

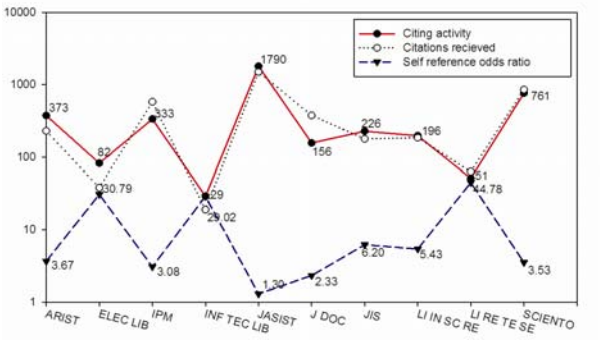

Figure 3: Reference activity among journals in the White and McCain journal set for 2007.

It is evident that a map of information science for 2007, based on the set of journals delimited by White and McCain, will be strongly influenced by the reference patterns in JASIST and to lesser degree SCIENTOMETRICS, and as a consequence that other journal's reference patterns will most likely be suppressed.

Journals with lower publication activity tend to have unstable profiles due to fewer references, and they seem to have larger self-reference odd ratios. Notice from Figure 3 that ELECT LIBR, INF TECHN LIBR and LIB RES TECH SER have the lowest reference activity and the highest self reference odds. If we compare these findings, to the structural patterns of their cited positions in Figure 1 and 2, we can infer that they all seem to have eccentric positions, with citing activity going outwards towards them. This is an indication that the majority of citations in the present set is expected to come from themselves. This suggests that these journals are specialized journals among the current selected journals. The reference activity is lower, and the resulting patterns markedly different from the rest.

The case of SCIENTOMETRICS is somewhat different, as SCIENTOMETRICS has a considerably larger reference activity, though still in an eccentric cited position similar to the three other journals. Notice (from Table 2) that SCIENTOMETRICS has *no* odds ratio ≥1 with the other journals, only a considerable self-reference ratio of 3.5 to 1. Of the 849 citations SCIENTOMETRICS received in 2007 in the current set, 570 are self-references. JASIST is the only other large contributor of references to SCIENTOMETRICS with 114. However, adjusted for main effects and cross-reference activities across the set for both journals, the odds ratios of SCIENTOMETRICS being cited by JASIST and vice versa turns out to be low. This leaves SCIENTOMETRICS'S cited position on the fringe of the set. The only citing position with favourable odds is SCIENTOMETRICS itself and the citing direction is outwards, away from the core set of cited journals. SCIENTOMETRICS is also a specialized journal.

SCIENTOMETRICS has a large reference activity. In fact it also has a relatively large self-reference odds ratio considering its large reference activity. The odds that a publication in SCIENTOMETRICS gives a reference to SCIENTOMETRICS (in the given data set) rather than exchanging references with one of the other journals is 3.5 to 1. This is in fact a considerable

number when one considers that in 2007 SCIENTOMETRICS published 129 publications according to JCR; and it has consequences for mapping studies.

Interpretation and consequences for mapping studies

The present empirical findings about specialized journals are consistent with White and McCain (1998, p. 330). Here, ELECT LIBR, INF TECHN LIBR, and LIB RES TECH SER are designated as belonging to "library automation". Interestingly, these journals have a relatively low reference activity and high self-reference odds ratios. The citing role of ELECT LIBR and INF TECHN LIBR do not deviate considerably from the ones that constitute the circle around the core cited set. They cite inward but also outward. The citing role of LIB RES TECH SER seems more eccentric, the journal cites itself, leaving its citing position some distance away from the circle. However, if we include these journals in a mapping study of information science for 2007, based on the White and McCain journal set, in order to reflect the "library automation", we will most likely not succeed, that is, if we do not reflect upon the varying degrees of publication productivity among the chosen journals. Clearly, these journals annually publish a considerable lower number of publications, compared to the rest of the set. By mere size, their specialized reference profiles will disappear among giants such as JASIST or SCIENTOMETRICS. Since most mapping studies eventually choose objects to be mapped based on frequencies, the likelihood that the special profile of low activity journals will have an impact upon the study is poor.

SCIENTOMETRICS is designated as information science in White and McCain (1998, p. 330). In the present study, however, we show that at least in 2007, SCIENTOMETRICS should be considered a specialized journal within this set based on its cross-reference activity. Perhaps SCIENTOMETRICS is on the periphery of information science and more a multidisciplinary journal? At least if we identify the core of information science as the cited journals in the centre of the map, and those citing journals that circulates around them and clearly cites inwards. We examined the overlap of authors between core information science journals and SCIENTOMETRICS for 2007 and for a five-year period from 2003-2007. In 2007, 7% of the authors in SCIENTOMETRICS also published in core information science journals. In the five-year period from 2003-2007, 18% of the authors in SCIENTOMETRICS also published in core information science journals. These figures support the claim that SCIENTOMETRICS is on the periphery of information science.

In contrast to ELECT LIBR, INF TECHN LIBR, and LIB RES TECH SER, selecting SCIENTOMETRICS for a mapping study of information science, based upon the White and McCain journal set, will most likely have a significant effect due the size and "ego-centric" pattern of its reference activity. It is apparent, with an annual publication rate around 130, and a self-reference activity odds ratio of 3.5 to 1, that a considerable number of these references will go to publications where the subject most likely will be scientometrics, informetrics, bibliometrics or webometrics. Compared to the core cited set of journals, where we can expect a more evenly distributed treatment of major subjects, due to the more evenly balanced mutual cross-reference activity between citing journals, bringing in a journal that is highly specialized, and at the same time highly productive, will no doubt affect the mapping solution. The specialized subject area may be "overemphasized". Notice, we do not claim that it is wrong to include for example SCIENTOMETRICS or other specialized journals in mapping studies. It clearly depends on aim and intention with a particular study. We simply emphasize the necessity of exploring data before analyses, as publication productivity and cross-reference patterns between journals are determinants in mapping studies.

In the next subsection we will demonstrate what happens if we include six extra journals into the White and McCain set.

Multidimensional unfolding of the White and McCain journal set.

In order to explore what happens if we enlarge White and McCain's core set of journals, we identified six extra journals. These journals are those among the total set of 56 in the subject category of "Information Science & Library Science" in 2007 that have the most similar profiles to those citing the core in the previous unfolding analysis. We will present major findings, and emphasize where the structural patterns in the map changes from the previous analysis presented in Figures 1 and 2 above. The results of the extended unfolding analysis are presented in Figures 4 and 5 below; Figure 5 illustrates the main citing directions for journals.

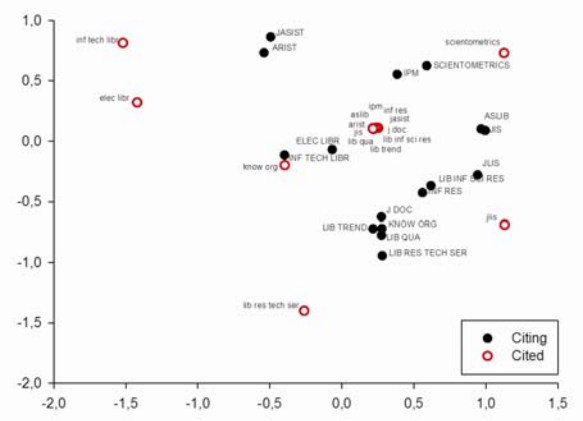

Figure 4: Multidimensional unfolding of the extended set of 16 journals.

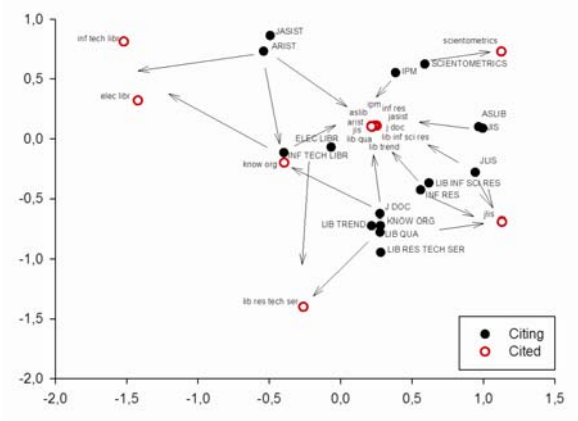

Figure 5: Main citing directions for collective groupings of journals in the extended set.

The two unfolding solutions are quite similar. We still have a core of cited journals at the centre of the map. Of the extra journals, ASLIB, INF RES, LIB QUA and LIB TREND appear in the cited core. Notice that the journal positions are almost on top of each other in the cited core. Notice, it is not a degenerate solution. It is, however, an indication of even more density in the core than in the first unfolding solution. Two of the journals have different cited positions. Like SCIENTOMETRICS, JLIS has a cited position to the right of the centre albeit on the negative scale of the *y* axis, compared to SCIENTOMETRICS. The journals citing JLIS cite outwards. This indicates that JLIS is a specialized journal in the current set, and that the specialization is different from that of for example SCIENTOMETRICS. The same is true for KNOW ORG. The cited position of the journal is to the left of the core. To a certain degree, the group of journals most likely to give references to KNOW ORG, cite in a direction away from the core (which some of them actually also cite into). So of the six new journals, four goes into the core and two have individual positions.

Examining the structural patterns of the citing journals after inclusion of the six new journals still shows the circle around the core. However, we see some stronger more coherent groupings in the present unfolding analysis. Remember position of citing journals is based on resemblances in citing behaviour, not merely who they cite, but more importantly, how likely is it that the journal cites another journal. We still have JASIST and ARIST at the top, albeit this time ARIST also have a citing direction towards KNOW ORG, followed clockwise by IPM and SCIENTOMETRICS. At three o'clock we find JIS, however, this time matched with ASLIB and not LIB INF SCI RES. This is meaning, since the UK based JIS and ASLIB have a considerable overlap of authors. At four and five o'clock, we find the citing positions of the new journals JLIS and INF RES together with LIB INF SCI RES. The citing profiles for these journals are alike, they both cite inwards toward the core and outwards toward JLIS. Perhaps the most interesting new structural pattern in the present unfolding analysis is the stronger grouping around JDOC and to a lesser degree LIB RES TECH SER. The new journals LIB TREND, LIB QUA and KNOW ORG

are positioned here close to JDOC. This indicates a strong resemblance of citing behaviour with JDOC. Interestingly, while both LIB QUA and LIB TREND have their cited position in the core set, KNOW ORG'S position is not. From Table 3 (www2.db.dk/jws/crossrefodds.htm), we can see the favourable odds ratios for KNOW ORG to receive citations come from this group. Notice also, as discussed above, that more specialized journals tend to have high self-reference odds ratios, especially if they do not have a high annual publication output. This seems to be the case for KNOW ORG. Consequently, extending the analysis with six journals elaborates the structural patterns found in the original set of White and McCain. The specialized journals from the first analysis remain specialized in the extended analysis. Three of the six extra journals in the extended analysis also show signs of specialty. Consequently, ten journals in the extended unfolding analysis seem to have balanced mutual citing and cited relations which bring them into the core of the selected journal set.

Summary and conclusions

The present paper explores the modelling of cross-reference activity between journals by use of asymmetric proximities and multidimensional unfolding. In contrast to previous studies, we want to model and map both citing and cited relations between journals in a particular reference network. We therefore examine the different roles of journals, their "citing behaviour" and their "cited reception", simultaneously. Modelling cross-reference activity between journals by use of asymmetric proximities and multidimensional unfolding can be considered an exploratory data analysis, where journal cross-reference data and publication activity are viewed. One intention with the present study is to demonstrate that we need to pay more attention to the data we include in mapping studies. We should not tacitly adhere to *ad hoc* methodical conventions. Like all empirical studies, we need to carefully explore and characterize our data in order to interpret them more accurately and to be able to reflect upon eventual biases. The present study demonstrates that analyzing asymmetric cross-reference activity between journals by use of multidimensional unfolding is an exploratory data analysis for such purposes. At least for the present journals sets modelled, the unfolding solutions give sensible solutions. Most importantly, the technique is able to model the different roles of journals, both when citing and being cited. Both roles say something about a journal and its positions in the reference network.

The present study demonstrates that it is appropriate to model reference activity between journals as asymmetric. The application of a simple asymmetric proximity measures where we calculate the odds of row journals citing column journals against the expected values under independence, enables modelling and mapping of journals with widely different reference activities. It is important to remember that this is an asymmetric probabilistic measure and *not* a symmetric similarity measure. We find that the multidimensional unfolding analyses perform satisfactorily, in that their results are not only stable over two analyses, but also straightforward to interpret. Unfolding the odds against independent citing gives an understanding of the self-reference behaviour of journals hitherto unseen.

Some of the specific structural characteristics found in the unfolding solutions are: 1) a core set of cited journals at the origin of the map; 2) a number of cited journals with eccentric positions on the fringes of the map; 3) a circular pattern of citing journals around the cited core set in the centre of the map.

Analyzing dominant citing directions of journals reveals that: 1) cited journals that are positioned in the core set at the origin of the map, and whose dominant citing direction is inward towards this core, have more mutually balanced cross-reference activities between them; they are alike; 2) cited journals that are positioned in the core set at the origin of the map, and whose dominant citing directions are both inward and outward, are good "senders" and "receivers" (Heiser & Busin, 2004); they are more general in their current scope, but the

reception of them is more in line with the others in the core; 3) cited journals positioned outside the core set, and whose dominant citing direction is inward towards this core, have more unbalanced asymmetric relations with other journals in the set; their citing and cited profiles differ; and 4) cited journals positioned outside the core set at the fringes of the map, and whose dominant citing direction is outward towards the cited position on the fringe, are different from the rest of the set; they are specialized journals. This leads to the following general findings for the multidimensional unfolding analysis: 1) citing inward towards the origin of the map indicates whether the journal set has a core; 2) citing outwards towards the fringes identifies specialized journals.

A final observation: the influence of publication activity is extremely important when selecting journals for mapping studies. High publication activity leads to high reference activity. Eventually, the latter will suppress the reference patterns from low activity journals selected for mapping studies. A consideration seldom reflected upon!

References

Borg, I. & Groenen, P.J.F. (2005). *Modern multidimensional scaling: Theory and applications*, 2nd edition, Springer Science: New York.

Busing, F., Groenen, P. & Heiser, W. (2005). Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika*, Vol. 70(1), p. 71-98.

Heiser, W. J. & Busing, F. M .T. A. (2004). Multidimensional scaling and unfolding of symmetric and asymmetric proximity relations. In: D. Kaplan, *The Sage Handbook of Quantitative Methodology for the Social Sciences*; London: Sage Publications, pp.25-48.

Klavans, R. & Boyack, K. (2005). Mapping the backbone of science. *Scientometrics*, Vol. 64(3), p. 351-374.

Leydesdorff, L. (2001). *The Challenge of Scientometrics*. *The development, measurement and selforganization of scientific communication*. 2nd Edition. Universal Publishers: uPublish.com

Leydesdorff, L. (2006). Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports? *Journal of the American Society for Information Science & Technology*, Vol. 57(5), 601-613.

Noma, E. (1982). An improved method for analyzing square scientometric transaction matrices. *Scientometrics*, Vol. 4(4), p. 297-316.

Price, D.J.D. (1981). The analysis of square matrices of scientometric transactions. *Scientometrics*, Vol. 3(1), p. 55-63.

Schneider, J.W. (2008). Analyzing asymmetric proximity relations in transaction matrices: The case of journal cross-reference matrices. IN: *Proceedings of the 10th International Conference on Science and Technology Indicators*, Ed. E. Schiebel, 17–20/9 - 2008, Vienna.

Tijssen, R.J.W., De Leeuw, J. & Van Raan, A.F.J. (1987). Quasi-correspondence analysis on scientometric transaction matrices. *Scientometrics*, Vol. 11(5-6), p. 351-366.

Van Deun, K., Marchal, K., Heiser, W., Engelen, K., & Van Mechelen, I. (2007). Joint mapping of genes and conditions via multidimensional unfolding analysis. *BMC Bioinformatics*, Vol 8, 181

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*, Vol. 49(4), p. 327-355.

Zhao, D., & Strotmann, A. (2008). Information Science during the first decade of the Web: An enriched author co-citation analysis. *Journal of the American Society for Information Science & Technology*, 59(6), 916-937.

Åström, 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 & Technology*, 58(7), p. 947-957.