Mapping the intellectual structure of information retrieval studies: an author co-citation analysis, 1987-1997

Ying Ding, Gobinda Chowdhury and Schubert Foo

Journal of Information Science 1999; 25; 67
DOI: 10.1177/016555159902500107

Mapping the intellectual structure of information retrieval studies: an author co-citation analysis, 1987–1997

Ying Ding, Gobinda Chowdhury and Schubert Foo

Nanyang Technological University, Singapore

Received 18 May 1998
Revised 28 August 1998

Abstract.

Author co-citation analysis (ACA) has been used to explore changes in the intellectual base of the information retrieval (IR) field over two consecutive time periods: 1987–1991 and 1992–1997. Thirty-nine highly cited IR researchers were selected as the research sample. Multidimensional scaling (MDS) and clustering techniques (CT) were used to create the two-dimensional maps to display the dynamic intellectual structure of IR, based on scholars citing their work over these two time periods. Factor analysis (FA) has been used to reveal the 'breadth' of the authors' research areas. ACA offers a good technique that contributes to the understanding of intellectual structure in the sciences and possibly in other areas to the extent that those areas rely on formal scholarly communication such as serial publications. Nonetheless, obvious drawbacks exist in ACA. These include the subjective nature of the interpretation of results, the difficulty of readily identifying clusters and the inability to distinguish collaborative research relationships between authors. Thus, ACA by itself is insufficient. However, ACA can be enhanced significantly when combined with FA to give a more accurate and useful picture of the MDS results.

Correspondence to: Y. Ding, Division of Information Studies, School of Applied Science, Nanyang Technological University, Nanyang Avenue, Singapore 639798. E-mail: P143387632@ntu.edu.sg

Introduction

As the structure of our society becomes increasingly complex, individuals and groups need to communicate an increasing volume of information. Schramm [1], a dominant figure in the communication sciences, once said that the field of communication was like 'an oasis in the desert, where many trails cross, and many travellers pass, but only a few tarry'. Bibliometrics or 'the application of mathematics and statistical methods to books and other media of communication' provide a method for examining communication among scholars in a field through their scholarly publication [2].

The introduction of the bibliometric method into communication science has led to a rapid growth in both the number and types of studies [3, 4]. Bibliometrics encompasses a number of empirical methods, citation and co-citation analysis [5]. In particular, they can be applied to the formal record of scholarly communication from different points, such as authors, journals and textual content. Author co-citation analysis (ACA), a well-established technique in bibliometrics, is a potentially productive method of examination of the advantages of the cognitive/intellectual structure of science specialities. Numerous studies use ACA to map the intellectual base of specialities that have included, for example, organisational behaviour [6], communication science [7], information science [8], marriage and family research [9], scholarly communication in the sociology of science and information science [10] and others [11].

This paper describes the use of author co-citation mapping to study changes in the intellectual structure of information retrieval (IR) over the past consecutive eleven years. This is further divided for analysis into two time frames; namely, 1987–1991 and 1992–1997.

Methodology

ACA's approach (as described in detail by White and Griffith [12]) is based on the frequency with which any work by an author is linked to any work by another author to a third and later work. It assumes that the more frequently two authors are cited together, and the more similar their patterns of co-citations with others, the closer is the relationship between them.

The validity of the method was demonstrated by many researchers (e.g. see [6, 8, 9, 12, 13, 14, 15]). In any case, the theoretical underpinning from the philosophy of science and the sociology of science is the same: their citation together is effected by the citing author, whose work contributes to the cumulative advancement of science by repeating old, and making new, linkages of prior contributions [16, 17, 18].

Selection of authors

The main bulk of IR research has been carried out by researchers from the disciplines of library and information science, computer science and other smaller related disciplines. These works are reported in *Social Science Citation Index* (*SSCI*) and *Science Citation Index* (*SCI*) [19]. However, a preliminary test conducted on *SCI* and *SSCI* confirmed that there are about 4,500 publications on information retrieval field during the period of study. Thus, including both the sets of data would prove too much to manage within the stipulated time because the total number of citations in those 4,500 orso publications would be more than 100,000. Therefore, this study focuses on data in *SSCI*. All the relevant papers were selected in *SSCI* via Dialog. Other useful records were added from *Library and Information Science Abstracts* (*LISA*) CD-ROM. From this, a total of 1,466 IR-related papers was selected from 367 journals with 44,836 citations. The citations include journal articles, books, conference papers and technical reports. If the citations are books, they are treated the same as the other kinds of citations and the whole book is considered as one citation. All the citations in this study only rely on the first authors because of the limitation of *SSCI* (the citations in this database only include first authors) and the co-authorship of the citations was not taken into account. Two databases were set up, based on these IR papers. These are ***Source Database*** and ***Citation Database*** respectively.

Thirty-nine most highly cited authors were selected from these two databases as the author sample of this research (Table 1). The most highly cited papers of each of these authors are listed in Appendix 1. The author co-citation frequency was calculated based on the ***Citation Database***. Among the highly cited authors, some are psychological researchers whose researches were related to psychology and consequently were reported in journals on psychology but not in *LISA*. Hence, the list of these authors (as shown in Appendix 2) was removed from the author sample. A separate study to derive the ACA maps using the author sample that included these psychological researchers has been reported in [20].

Table 1
Authors selected for ACA

Salton, G. (772)	Belkin, N.J. (318)	Croft, W.B. (247)
Saracevic, T. (219)	Robertson, S.E. (206)	Borgman, C.L. (198)
Van Rijsbergen, C.J. (194)	Bates, M.J. (187)	Sparck Jones, K. (166)
Blair, D.C. (141)	Swanson, D.R. (136)	Cooper, W.S. (136)
Fox, E.A. (135)	Lancaster, F.W. (131)	Bookstein, A. (127)
Ellis, D. (105)	Fidel, R. (104)	Marchionini, G. (102)
Ingwersen, P. (102)	Harman, D. (87)	Harter, S.P. (85)
Losee, R.M. (84)	Dervin, B. (80)	Markey, K. (80)
Meadow, C.T. (78)	Wong, S.K.M. (73)	Cleverdon, C.W. (71)
Rada, R. (67)	Radecki, T. (61)	Smeaton, A.F. (61)
Yu, C.T. (60)	Spink, A. (59)	Kuhlthau, C.C. (58)
Maron, M.E. (56)	Oddy, R.N. (56)	Fuhr, N. (55)
Chen, H.C. (53)	Turtle, H. (50)	Zadeh, L.A. (50)

Note: The numbers in brackets indicate the cited frequency of the authors over the study period.

Author co-citation matrix

A Foxpro database program was written to calculate the co-citation frequency of every two of these 39 authors from ***Citation Database*** in the two different time periods (i.e. 1987 to 1991 and 1992 to 1997). An author co-citation matrix was set up, based on these frequencies of author co-citation. The diagonal data were calculated according to White and Griffith [12]. The data in the matrix are each author's profile of co-citation with every other author on the list. This resulted in a 39×39 matrix.

Data analysis

The matrix of raw co-citation counts was analysed using multidimensional scaling and clustering displays. Factor analysis techniques were used to complement these two methods. Author co-citation frequency reflects the similarity of author pairs. A high co-citation frequency shows that these two authors are more similar or more related to each other.

Results of multidimensional scaling and cluster analysis

Multidimensional scaling analysis

Multidimensional scaling (MDS) is a set of techniques used to create visual displays (maps) from matrices, so that the underlying structure within a set of objects can be studied [21, 22]. The major output of MDS is a display of points in two or three dimensions. Points are placed on the map according to their proximity in the author co-citation matrix (where high values reflect high similarities). Points representing authors with high similarities will be placed close together, while points representing authors with low similarities will be placed farther apart in the map.

In ACA, the major uses of MDS are twofold: to provide an information-rich display of the co-citation linkages and to identify the salient dimensions underlying their placement. Figs. 1 and 2 show the two-dimensional MDS mapping generated by the ALSCAL program (part of the Statistical Package for the Social Sciences (SPSS) [21]) with good fit for the two time periods.

Clustering techniques (CT) are used to group authors so as to provide insights into the intellectual organisation of a given field. The cluster-generating programs are also available in SPSS. Authors are grouped within boundaries as shown in solid lines. The name of the cluster was chosen based on the common research topics of each author in this cluster. (The details of MDS and CT are discussed in detail in references [21, 22] and [23] and will not be repeated here.)

General structural stability

In general, there are some schools to thought which both appear in the two periods' ACA map, such as

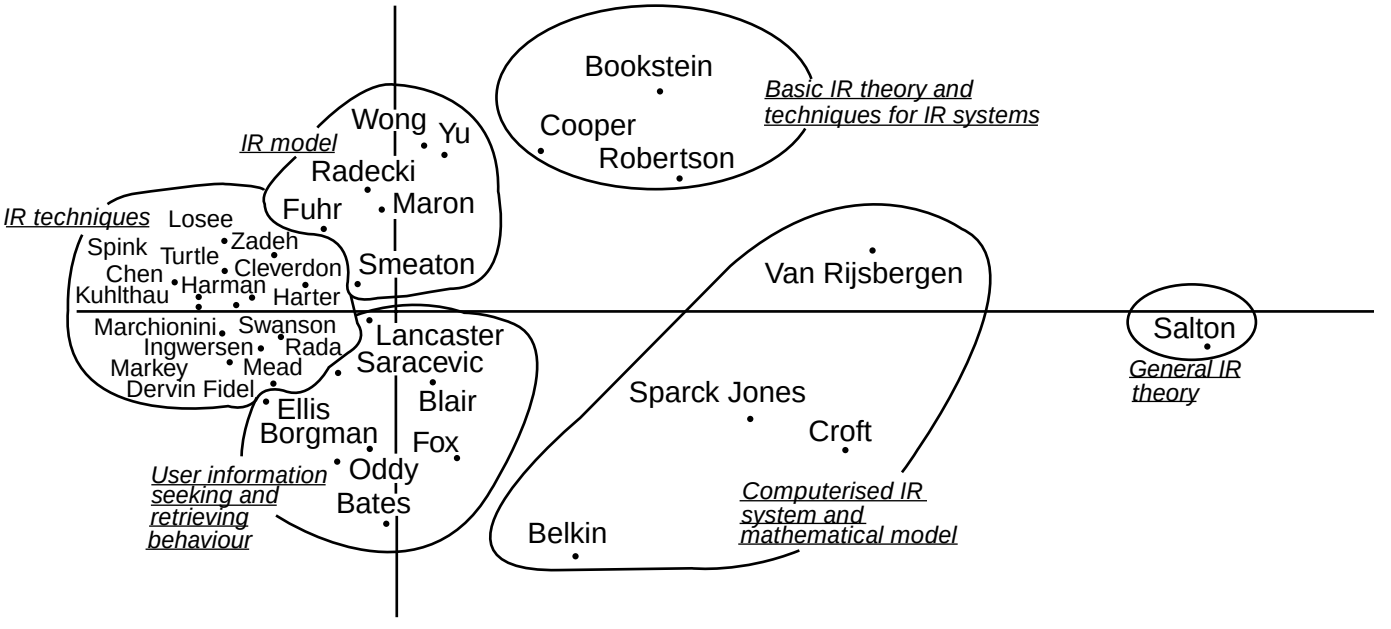

Fig. 1. Author co-citation mapping of IR (1987–1991).

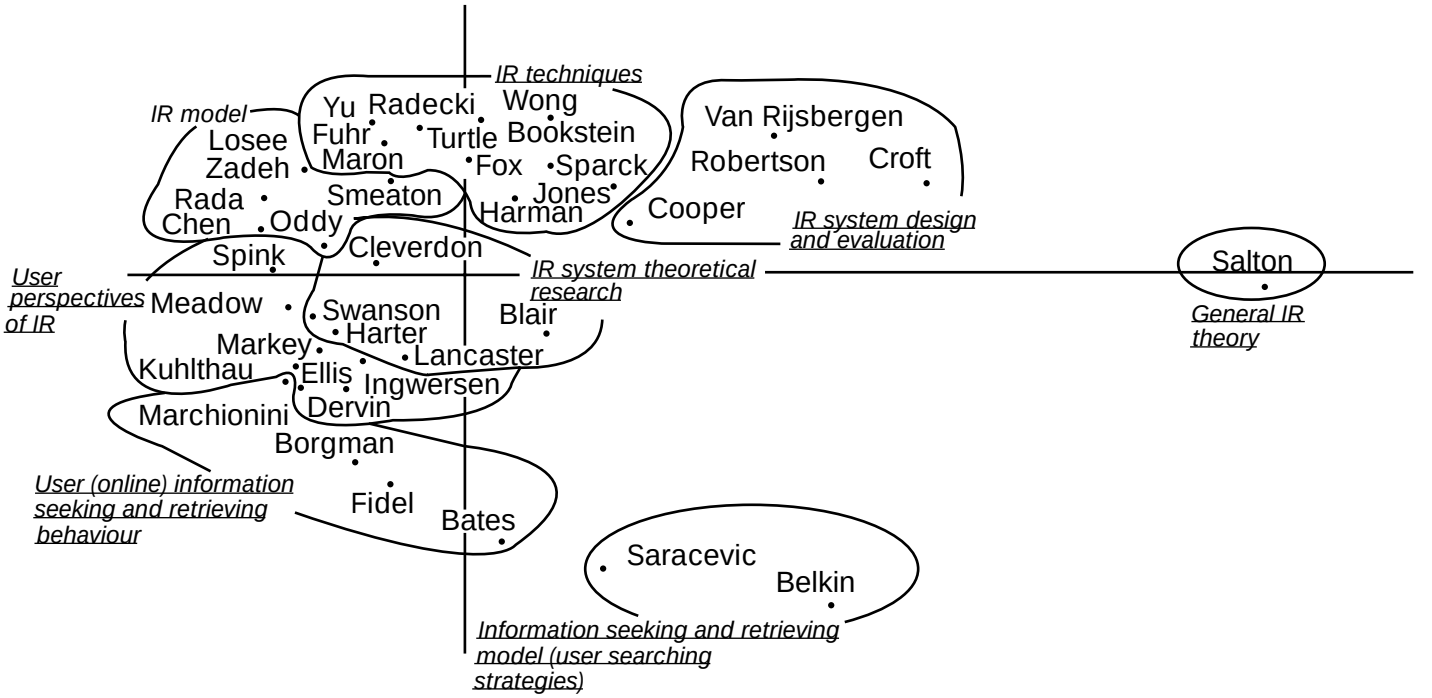

Fig. 2. Author co-citation mapping of IR (1992–1997).

general IR theory, IR model, IR techniques, information seeking and retrieving behaviour and so on.

General IR theory, the main sole contributor of which is Salton, is distinct and located on the right-hand of both the maps. This result coincides with that of another research study [24].

In contrast, user information seeking and retrieving behavioural research of Fig. 1 has been divided more distinctly into user perspectives of IR and online information seeking and retrieving behaviour, plus information seeking and retrieving model (user searching strategies) in Fig. 2. All these groups are located together or are in the middle of the map. This reflects that they are vital areas of IR research.

The horizontal axis (from left to right) of the eleven-year period seems to represent more specific and technological researches to more general and basic theoretical researches. The vertical axis (from top to bottom) for the same period appears to see a shift from theoretical system design to application and evaluation and, finally, to user searching behaviour.

Scholarly migration

The general spatial orientation of authors and their cluster assignment have not changed much in these two periods. However, no single author has maintained exactly the same position on the map during these two time periods. This does not imply that all authors are moving to new research areas. Even if an author's research area remains, the position of that author on different maps might change, since these maps were derived from the citation relationships of all the authors. For example, Sparck Jones's position on these two maps is very different: one is above the horizontal axis and one is below the horizontal axis, although her research areas did not change much.

On the other hand, when a comparison of the position of an author is made on these two maps, we can see that, although some authors' positions did not change, the research groups which they belonged to have changed. For example, Belkin's position on both the maps did not change, but the other authors who shared the same research interests with him changed from Van Rijsbergen, Croft and Sparck Jones to Saracevic. In this instance, the research topics have changed on these two maps from computerised IR systems and mathematical models to information seeking and retrieving models (user searching strategies). This is an interesting phenomenon. This is similarly observed with Wong, Yu and Robertson. This may be attributed to a number of factors. First, the 'minus' migration of some authors'

research topics can change the cluster that the author belonged to and the structure of MDS; second, the reasons for one author to be cited can vary from paper to paper, person to person and time to time. Finally, since one paper can have many parts, with each part having different sub-topics, different authors will inevitably cite different parts of the paper. Such distinctions cannot be picked up using ACA.

All authors within group boundaries share similar co-citation profiles. Closely placed pairs have very similar profiles, which is equivalent to a consensus that the work of the two authors is highly related. Authors having positive correlation with other authors across group boundaries deserve further study. They created a bridge for two research groups and often borrow new ideas from other groups to improve or extend the research of their own groups. Usually, new research areas were created across the boundary of different research groups. For example, Maron and Smeaton, who were in the same group of IR models in Fig. 1 were separated into two groups of IR models and IR techniques in Fig. 2, even though they are located very close to the boundary of these new groups.

Factor analysis

Factor analytic technique (or factor analysis (FA)) has been applied in this study to complement MDS and CT. This technique can explain the interrelationships observed among the original variables through the creation of some derived variables or factors.

FA of authors from 1987 to 1991

The matrix of raw co-citation counts is factor-analysed using a principal component analysis with a varimax rotation (replacing missing values with the mean). The results of FA of IR from 1987 to 1991 are shown in Table 2. Seven factors were extracted, based on a scree test.[1] An eigenvalue less than one has been chosen as the rule for stopping extraction of factors.[2] These seven factors account for 80.6% of the variance, with the first two factors accounting for more than 61.2%. All the seven factors are listed in order of total variance accounted for (with factor 1 being the highest). All authors loading that are above 0.40 are also presented in Table 2.

In using MDS and CT, an author can appear in only one group or one cluster. In contrast, an author can appear in more than one factor in FA. Thus, FA can reveal additional specific facts about an author's research scope that the other methods can not. All the 39 authors load on at least one factor, except Spink (as he was not co-cited with any other author during this period). Thirteen authors load on two factors, five authors load on three factors and one (Fox) loads on four factors. Only authors with loading above 0.7 were used to explain the factor [26].

Authors loading on factor 1 are IR theory researchers. They focused their research on probability of relevance, term dependence, Boolean retrieval and so on. Factor 5 includes the subset of retrieval theory researchers who have furthered their theoretical researches into examining indexing theory, data relevance, or data recall and precision. Factors 3, 4 and 7 or even factor 5 are related to user research, but from different aspects: factor 3 from an online IR aspect; factor 4 from a computerised IR aspect; factor 7 from a cognitive and user behavioural aspect and factor 5 from a theoretical and methodological aspect. Factor 2 is the technical basis for user research. From this point, we can see that user searching behaviour is one of the most important emerging IR research topics. Authors loading on factor 2 have a strong computer science background, e.g. Chen's knowledge-based IR system, Rada's machine learning and Fox's artificial intelligence. Authors loading on factor 6 based their researches on fuzzy-set theory (such as Zadeh, Radecki and Bookstein). Fox loads on four factors (1, 2, 3 and 4), which implies that his research focus is varied (and includes IR theory, computerised IR and user searching behaviour).

Fig. 3 shows the resulting grouping of FA using MDS and CT into three main groups. The first group (factor 1) is separate from the others and located at the right-hand side of the map, since this pertains to the distinct research area of basic IR theories (as similarly shown in Figs. 1 and 2). The second group (factor 6) focuses more on mathematical and probabilistic IR models. The last group's (containing the remaining factors) common research topics include computerised IR and user researches. From Fig. 3, we can see that the result from FA coordinates well with the result from MDS and CT.

1. Scree test is a test advocated by Cattell. The rules direct one to examine the graph of eigenvalues, and stop factoring at the point where the eigenvalues (or characteristic roots) begin to level off, forming a straight line with an almost horizontal slope (where scree is the geological term referring to the debris which collects on the lower part of a rocky slope) [25].
2. This simple criterion seems to work well, in the sense that it generally gives results consistent with researchers' expectations [25].

Table 2
FA of IR from 1987 to 1991

Factor 1 *IR theory*		Factor 2 *Computerised IR*		Factor 3 *User behaviour in online IR*	
Yu	0.94	Chen	0.83	Ingwersen	0.90
Fuhr	0.91	Rada	0.74	Meadow	0.73
Wong	0.89	Bates	0.57	Harter	0.61
Robertson	0.84	Blair	0.57	Bates	0.57
Van Rijsbergen	0.75	Markey	0.54	Fidel	0.55
Smeaton	0.75	Borgman	0.52	Borgman	0.52
Salton	0.74	Lancaster	0.47	Fox	0.50
Maron	0.68	Marchionini	0.47	Markey	0.48
Bookstein	0.67	Harman	0.44		
Losee	0.66	Fox	0.44		
Croft	0.66	Fidel	0.43		
Sparck Jones	0.66				
Cooper	0.65				
Radecki	0.58				
Harman	0.47				
Fox	0.46				
Lancaster	0.45				
Turtle	0.44				
Blair	0.41				

Factor 4 *User behavioural model (computerised IR)*		Factor 5 *Information seeking and retrieving study (methodology)*		Factor 6 *Probabilistic IR (fuzzy set)*		Factor 7 *Information needs and user research*	
Ellis	0.82	Saracevic	0.67	Zadeh	0.76	Kuhlthau	0.84
Oddy	0.79	Cleverdon	0.64	Radecki	0.64	Dervin	0.81
Belkin	0.72	Swanson	0.64	Bookstein	0.45	Borgman	0.54
Cleverdon	0.53	Cooper	0.63	Harman	0.44	Turtle	0.45
Croft	0.46	Maron	0.52	Van Rijsbergen	0.41		
Fidel	0.46	Markey	0.43				
Sparck Jones	0.45	Harter	0.42				
Fox	0.41	Losee	0.42				
		Lancaster	0.41				

FA of authors from 1992 to 1997

The matrix of raw co-citation counts is factor-analysed as before. However, this time round, varimax rotation failed to converge in 25 iterations with eigenvalue less than one as the cut-off. Thus, an eigenvalue that is less than 1.03 was selected instead, so that each researcher appears in at least one factor while still allowing the factors to be interpreted in most cases. In so-doing, six factors were extracted.

The results of FA of IR from 1992 to 1997 are shown in Table 3. These six factors account for 80.1% of the variance, with the first two factors accounting for more than 62.3%. All authors loading above 0.4 are presented in the table. All 39 authors load on at least one factor, 13 authors load on two factors and one author (Blair) loads on four factors.

Factors 1 and 3 represent the foundations of IR theory research, with factor 3 being more focused on the cognitive aspect. The authors loading on factor 1 seldom load on other factors. In factor 1, the highest loadings come from Robertson (0.93), Van Rijsbergen (0.90) and Bookstein (0.88), whose researches focus on IR theory on probability of relevance, weighting function, query expansion, term dependence, traditional Boolean retrieval, probabilistic indexing and so on. Authors loading on factor 3 have concentrated their efforts on psychological relevance or psychological evaluation of IR systems and so on.

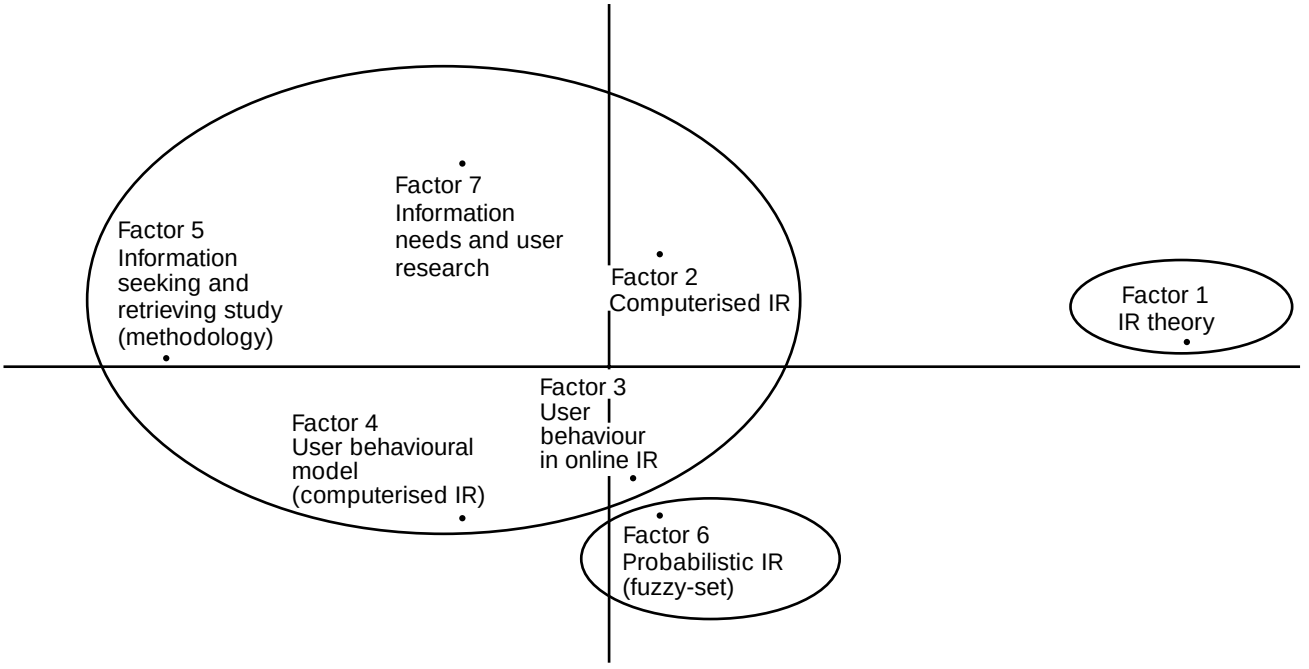

Fig. 3. MDS mapping of factors from 1987 to 1991.

Factors 2, 4 and 6's research topics are related to information seeking and retrieving behaviour from the behavioural aspect, the seeking model aspect and the online searching aspect respectively. Factors 2 and 6 have a very close relationship because of online seeking and searching behaviour. There is only one author loading on factor 6: Spink. The reason for the emergence of this is that, being a relatively new researcher when compared with the others, his publications on user search-term selection during online searching appear only after 1993. His loading was below 0.4 in the first period of study (1987–1991) and therefore is not shown in Table 2.

Factor 5 is computerised IR and Rada (0.76) is the highest-loading author in this factor. Blair seems to have broad research areas, because he is loaded on four factors (l, 2, 3 and 5) and his research topics include IR theory, online information seeking and retrieving behaviour and computerised IR.

Fig. 4 shows the MDS map of these six factors. Factors 1, 2, 4 and 6 are located together on the top of the map. Factors 3 and 5 were located alone on the left and bottom half of the map. From this, it is apparent that user searching behaviour, particularly online information seeking and retrieving behaviour, is an emerging important research topic. Likewise, IR system theoretical research continues to be another important area, since it is located in the middle of the map. These results indicate that ACA has the ability to detect intellectual shift or emergence of new research areas if the timescale is divided into appropriate periods for study. A related study [20] based on the whole eleven-year period failed to detect this change in the IR field.

Conclusion

ACA has been applied successfully to map the intellectual structure of the IR field over two consecutive time periods of 1987–1991 and 1992–1997. This partitioning of the period of study has demonstrated that more information can be inferred from ACA in contrast to studying the whole eleven-year time period.

From the results of this study, it is evident that ACA is indeed a useful tool for studying the intellectual structure of scholarly disciplines. It provides one view of the organisation of scholarly specialities, representing the consensus of a very large number of scholars concerning the representatives among important writers in the field. It is very useful in complementing and cross-validating analyses of data representing scholars' perceptions of the field, social relationships, and formal and informal communication networks. Due to the time required for research results to be transferred into the formal literature and to be cited, and for the cumulating of several years' citation data, co-cited author maps are

Table 3
FA of IR from 1992 to 1997

Factor 1 *IR theory*		Factor 2 *Online information seeking and retrieving behaviour*		Factor 3 *IR theory research (cognitive aspect)*	
Robertson	0.93	Borgman	0.91	Swanson	0.92
Van Rijsbergen	0.90	Bates	0.87	Harter	0.77
Bookstein	0.88	Marchionini	0.87	Lancaster	0.63
Wong	0.85	Fidel	0.86	Cleverdon	0.61
Salton	0.85	Meadow	0.80	Saracevic	0.53
Smeaton	0.82	Markey	0.76	Blair	0.53
Radecki	0.82	Saracevic	0.67	Cooper	0.52
Yu	0.81	Belkin	0.63		
Croft	0.81	Ingwersen	0.60		
Cooper	0.78	Spink	0.58		
Losee	0.75	Kuhlthau	0.56		
Fuhr	0.74	Lancaster	0.54		
Maron	0.73	Dervin	0.46		
Turtle	0.72	Harter	0.45		
Sparck Jones	0.71	Blair	0.41		
Harman	0.68				
Zadeh	0.66				
Fox	0.61				
Blair	0.40				

Factor 4 *Information seeking behavioural model*		Factor 5 *Computerised IR*		Factor 6 *User searching strategies (online searching)*	
Ellis	0.83	Rada	0.76	Spink	0.57
Dervin	0.67	Chen	0.68		
Kuhlthau	0.66	Fox	0.62		
Ingwersen	0.56	Croft	0.51		
Belkin	0.55	Blair	0.44		
Oddy	0.55	Salton	0.42		
Cleverdon	0.51				

most useful for providing a general historical view of the intellectual structure of a research area.

The intellectual base of the field, as displayed by an author co-citation map, appears to have strong validity. The tradition of IR seems to be subdivided into one 'hard' part working on IR theory and retrieval algorithms and one 'soft' part concentrating on the user-system relation.

Researchers can find much value in co-cited author mapping. The co-citation data are easily obtainable and amenable to analysis using a variety of well-established techniques to explore hidden patterns within large matrices. They are derived from a much larger body of literature than is likely to be manageable using subjective interviews and surveys. They represent an aggregate consensus that is likely to influence information use and communication patterns. Finally, using the author as the unit of analysis allows the researcher to study a number of complementary aspects of scholarly activity: the intellectual structure through links among bodies of published work and the cognitive structure deriving from citing authors' perceptions of their field.

As for as ACA is helpful in delivering much useful information, it does, however, have a number of drawbacks. First, the interpretation of results from ACA is difficult and subjective. It requires the interpreter to have some basic background knowledge about the research field in order to distinguish the differences between clusters. Sometimes, it is difficult to generalise the research topics of one group because the authors of such a group may also do different kinds of researches so that the inner links between them are weak. Different

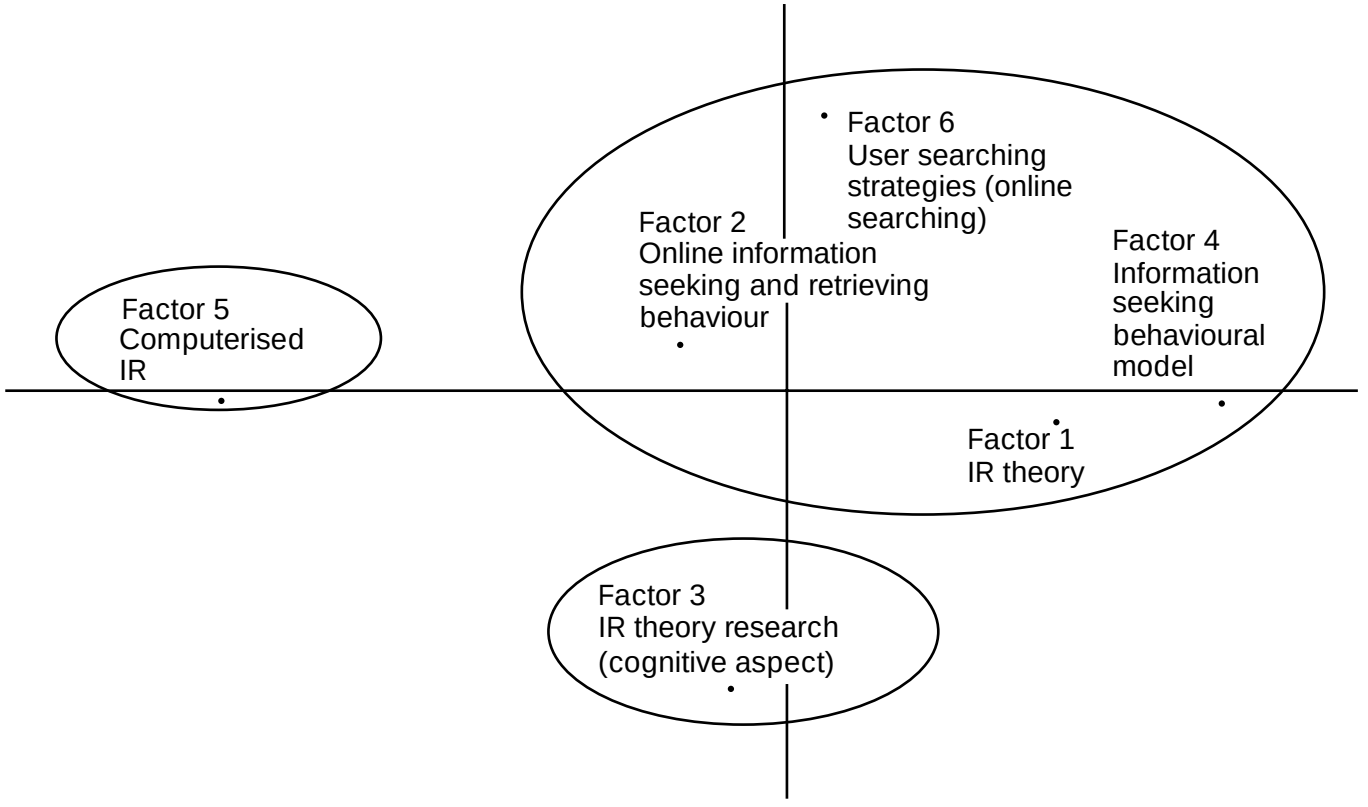

Fig. 4. MDS mapping of factors from 1992 to 1997.

researchers potentially can explain the results in different ways or even in opposite ways. Thus, the interpretation of results is subjective, even though the data and method of ACA are standardised and objective. On the other hand, no fixed rules exist for the interpreter to follow. The good and knowledgeable interpreter can see much to explicate the fine structure of the results. However, not all interpreters are knowledgeable. As such, there is a need for some form of theory or heuristics for the interpreter to follow in order to reduce the subjective difference of interpretation.

Second, most of ACA researches are based on *SCI* or *SSCI* databases. The question arises as to how collaborative relationships between author and co-authors can be studied, since these databases include only first authors. For example, if Belkin and Croft are collaborators in IR technique research, we would expect them to be located in one group on the map. In fact, they are not, but are separated into two groups. The same applies for other authors such as Blair and Maron. Thus, if co-citation includes second or third authors, the resulting ACA maps are likely to be different. At the same time, ACA cannot distinguish collegial ties through common institutional affiliations between authors, since this information is not considered in ACA. In order to disentangle such difficulties, the ACA methodology needs to be refined.

Third, ACA by itself is insufficient, but it can be enhanced when combined with FA. FA is an important complementary aspect of ACA research, since it aids the fuller understanding of the MDS results. The FA technique is able to distinguish the important author in a cluster. As a result, this author's research topics become the contenders for the group's research topics. This is in contrast to MDS which can identify only the four or five authors in the group, but not the group's research topics. In addition, FA can show the relationship among different factors as a result of the load values obtained. For instance, a high loading (above 0.70) author will appear in only one factor. By having such values, interpretation and explanation become more obvious. At the same time, FA has the ability to detect small transitions and changes in research patterns.

Appendix 1

This Appendix lists the most highly cited articles of each author in the author sample over the two consecutive time periods of study.

M.J. Bates, The design of browsing and berrypicking techniques for the online search interface, *Online Review* 13(5) (1989) 407–424.

M.J. Bates, Information search tactics, *Journal of the American Society for Information Science* 30(4) (1979) 205–214.

M.J. Bates, Subject access in online catalogs: a design model, *Journal of the American Society for Information Science* 37 (1986) 357–376.

N.J. Belkin, R.N. Oddy and H.M. Brooks, ASK for information retrieval: Part 1. Background and theory, *Journal of Documentation* 38(2) (1982) 61–71.

N.J. Belkin, R.N. Oddy and H.M. Brooks, ASK for information retrieval: Part 2. Results of a design study, *Journal of Documentation* 38(3) (1982) 145–164.

D.C. Blair, An evaluation of retrieval effectiveness for a full-text document retrieval system, *Communications of the ACM* 28 (1985) 289–299.

A. Bookstein, Probability and fuzzy-set applications to information retrieval, *Annual Review of Information Science and Technology* 20 (1985) 117–151.

C.L. Borgman, Why are online catalogs hard to use? Lessons learned from information retrieval studies, *Journal of the American Society for Information Science* 37 (1986) 387–400.

V. Dhar and H. Chen, Cognitive process as a basis for intelligent retrieval system design, *Information Processing and Management* 27(5) (1991) 405–432.

C.W. Cleverdon and M. Keen, *Factors Determining the Performance of Indexing Systems* (Aslib Cranfield Research Project, Cranfield, Bedford, 1966).

W.S. Cooper, A definition of relevance for information retrieval, *Information Storage and Retrieval* 7(1) (1971) 19–37.

W.S. Cooper, Indexing documents by gedanken experimentation, *Journal of the American Society for Information Science* 29(3) (1978) 107–119.

W.S. Cooper, On selecting a measure of retrieval effectiveness: I. The 'subjective' philosophy of evaluation, *Journal of the American Society for Information Science* 24(2) (1973) 87–100.

W.B. Croft and R.H. Thompson, I^3R: a new approach to the design of document retrieval system, *Journal of the American Society for Information Science* 38(6) (1987) 389–404.

B. Dervin and M. Nilan, Information needs and uses, *Annual Review of Information Science and Technology* 21 (1986) 3–33.

B. Dervin and M. Nilan, Information needs and uses, *Annual Review of Information Science and Technology* 21 (1986) 3–33.

D. Ellis, Theory and explanation in information retrieval research, *Journal of Information Science* 8(1) (1984) 25–38.

D. Ellis, A behavioural approach to information retrieval system design, *Journal of Documentation* 45(3) (1989) 171–212.

D. Ellis, The physical and cognitive paradigms in information retrieval research, *Journal of Documentation* 48(1) (1992) 45–64.

R. Fidel, Toward expert systems for the selection of search keys, *Journal of the American Society for Information Science* 37(1) (1986) 37–44.

R. Fidel, Searchers' selection of search keys: 2. Controlled vocabulary or free-text searching, *Journal of the American Society for Information Science* 42(7) (1991) 501–514.

R. Fidel, Searchers' selection of search keys: 1. The selection routine, *Journal of the American Society for Information Science* 42(7) (1991) 490–500.

E.A. Fox, Development of the CODER system: a testbed for artificial intelligence methods in information retrieval, *Information Processing and Management* 23(4) (1987) 341–366.

N. Fuhr, Models for retrieval with probabilistic indexing, *Information Processing and Management* 25(1) (1989) 55–72.

D. Harman and G. Candela, Retrieving records from a gigabyte of text on a minicomputer using statistical ranking, *Journal of the American Society for Information Science* 41(8) (1990) 581–589.

S.P. Harter, *Online Information Retrieval: Concepts, Principles and Techniques* (Academic Press, Orlando, 1986).

S.P. Harter, Psychological relevance and information science, *Journal of the American Society for Information Science* 43(9) (1992) 602–615.

P. Ingwersen, A cognitive view of three selected online search facilities, *Online Review* 8(5) (1984) 465–492.

P. Ingwersen, Search procedures in the library – analysed from the cognitive point of view, *Journal of Documentation* 38(3) (1982) 165–191.

P. Ingwersen, *Information Retrieval Interaction* (Taylor Graham, London, 1992).

K. Sparck Jones, *Information Retrieval Experiment* (Butterworths, London, 1981).

K. Sparck Jones, *Automatic Keyword Classification for Information Retrieval* (Butterworths, London, 1971).

C.C. Kuhlthau, B.J. Turock, M.W. George and R.J. Belvin, Validating a model of the search process: a comparison of academic, public and school library users, *Library and Information Science Research* 12(1) (1990) 5–31.

C.C. Kuhlthau, Inside the search process: information seeking from the user's perspective, *Journal of the American Society for Information Science* 42(5) (1991) 361–371.

F.W. Lancaster, *Vocabulary Control for Information Retrieval* (Information Resources Press, Arlington, 1986).

F.W. Lancaster, *Information Retrieval Systems: Characteristics, Testing, and Evaluation* (Wiley, New York, 1968).

R.M. Losee, Parameter estimation for probabilistic document-retrieval models, *Journal of the American Society for Information Science* 39(1) (1988) 8–16.

G. Marchionini and B. Shneiderman, Finding fact vs. browsing knowledge in hypertext systems, *IEEE Computer* 21 (1988) 70–80.

G. Marchionini, Information seeking strategies of novices using a full-text electronic encyclopedia, *Journal of the American Society for Information Science* 40(1) (1989) 54–66.

K. Markey, Levels of question formulation in negotiation of information need during the online research interview: a proposed model, *Information Processing and Management* 17(5) (1981) 215–225.

K. Markey, Searching and browsing the Dewey Decimal Classification in an online catalog, *Cataloging and Classification Quarterly* 7(3) (1987) 37–68.

M.E. Maron, On relevance, probabilistic indexing and information retrieval, *Journal of the ACM* 7 (1960) 216–244.

C.T. Meadow, Online access to knowledge: system design, *Journal of the American Society for Information Science* 40(2) (1989) 86–98.

C.T. Meadow, OAKDEC, a program for studying the effects on users of a procedural expert system for database searching, *Information Processing and Management* 24(4) (1988) 449–457.

R.N. Oddy, Information retrieval through man-machine dialogue, *Journal of Documentation* 33(1) (1977) 1–14.

R.N. Oddy, PTHomas: an adaptive information retrieval system on the connection machine, *Information Processing and Management* 27(4) (1991) 317–335.

R. Rada, Knowledge-sparse and knowledge-rich learning in information retrieval, *Information Processing and Management* 23(3) (1987) 195–210.

R. Rada and E. Bicknell, Ranking documents with a thesaurus, *Journal of the American Society for Information Science* 40(5) (1989) 304–310.

T. Radecki, Fuzzy set theoretical approach to document retrieval, *Information Processing and Management* 15(5) (1979) 247–259.

T. Radecki, Trends in research on information retrieval – the potential for improvements in conventional Boolean retrieval systems, *Information Processing and Management* 24(3) (1988) 219–227.

S.E. Robertson, Relevance weighting of search terms, *Journal of the American Society for Information Science* 27(3) (1976) 129–146.

S.E. Robertson, The probability ranking principle in IR, *Journal of Documentation* 33(4) (1977) 294–304.

S.E. Robertson, Probability of relevance: a unification of two competing models for document retrieval, *Information Technology: Research and Development* 1(1) (1982) 1–21.

G. Salton and M.J. McGill, *Introduction to Modern Information Retrieval* (McGraw-Hill, New York, 1983).

T. Saracevic, P. Kantor, A.Y. Chamis and D. Trivison, A study of information seeking and retrieving. 1. Background and methodology, *Journal of the American Society for Information Science* 39(3) (1988) 161–176.

T. Saracevic and P. Kantor, A study of information seeking and retrieving. 2. Users, questions, and effectiveness, *Journal of the American Society for Information Science* 39(3) (1988) 177–196.

T. Saracevic and P. Kantor, A study of information seeking and retrieving. 3. Searchers, searches, and overlap, *Journal of the American Society for Information Science* 39(3) (1988) 197–216.

A.F. Smeaton, The retrieval effects of query expansion on a feedback document retrieval system, *Computer Journal* 26 (1983) 239–246.

A. Spink and T. Saracevic, Sources and use of search terms in online searching. In: D. Shaw (ed.), *Proceedings of the 55th Annual Meeting of the ASIS, Pittsburgh, 26–29 October 1992* (Learned Information, New Jersey, 1992). D.R. Swanson, Historical note: information retrieval and the future of an illusion, *Journal of the American Society for Information Science* 39(2) (1988) 94–98.

H. Turtle and B. Croft, Evaluation of an inference network-based retrieval model, *ACM Transactions on Information Systems* 9 (1991) 187–222.

C.J. Van Rijsbergen, *Information Retrieval* (Butterworths, London, 1975).

C.J. Van Rijsbergen, A theoretical basis for the use of co-occurrence data in information retrieval, *Journal of Documentation* 33(2) (1977) 106–119.

S.K.M. Wong and Y.Y. Yao, A probability distribution model for information retrieval, *Information Processing and Management* 25(1) (1989) 39–53.

S.K.M. Wong and Y.Y. Yao, Query formulation in linear retrieval models, *Journal of the American Society for Information Science* 41(5) (1990) 334–341.

S.K.M. Wong and Y.Y. Yao, A generalized library probabilistic independence model, *Journal of the American Society for Information Science* 41(5) (1990) 324–329.

C.T. Yu, Precision weighting – an effective automatic indexing method, *Journal of the ACM* 23 (1976) 76–88.

C.T. Yu, C. Buckley, K. Lam and G. Salton, A generalized term dependence model in information retrieval, *Information Technology: Research and Development* 2(4) (1983) 129–154.

L.A. Zadeh, Fuzzy set, *Journal of Information and Control* 8 (1965) 338–353.

Appendix 2

This Appendix lists the psychological researchers that are found in *SSCI* but not included in the study. (*Note* that the numbers in the brackets denote the cited frequency of the related author.)

Tulving, E. (186), Anderson, J.R. (169), Ratcliff, R. (144), Hintzman, D.L. (124), Murdock, B.B. (118), Jacoby, L.L. (106), Squire, L.R. (89), Craik, F.I.M. (87), Brainerd, C.J. (77), Schacter, D.L. (74), Humphreys, M.S. (70), Roediger, H.L. (70) and Baddeley, A.D. (70).

Acknowledgements

The authors are grateful to Professor Roderick Cave of Nanyang Technological University, Singapore, and Dr E.C.M. Noyons of Leiden University, Netherlands, for their valuable advice on the work and paper. The authors are also grateful to Dr Eugene Garfield and other anonymous referees for their important comments and suggestions on this paper.

References

[1] W. Schramm, The beginnings of communication study in the United States. In: D. Nimmo (ed.), *Communication Yearbook 4* (Transaction Books, New Brunswick, NJ, 1980).

[2] C.L. Borgman (ed.), *Scholarly Communication and Bibliometrics* (Sage, Newbury Park, CA, 1990).

[3] B. Cronin, *The Citation Process: The Role and Significance of Citations in Scientific Communication* (Taylor Graham, London, 1984).

[4] C.L. Borgman, Editor's introduction. In: C.L. Borgman (ed.), *Scholarly Communication and Bibliometrics* (Sage Publications, London, 1990).

[5] H. Small, Cocitation in the scientific literature: a new measure of the relationship between two documents, *Journal of the American Society for Information Science* 24 (1973) 265–269.

[6] M.J. Culnan, C.A. O'Reilly III and J.A. Chatman, Intellectual structure of research in organizational behavior, 1972–1984: a cocitation analysis, *Journal of the American Society for Information Science* 41(6) (1990) 453–458.

[7] W. Paisley, The future of bibliometrics. In: C.L. Borgman (ed.), *Scholarly Communication and Bibliometrics* (Sage Publications, London, 1990).

[8] O. Persson, The intellectual base and research fronts of JASIS 1986–1990, *Journal of the American Society for Information Science* 45(1) (1994) 31–38.

[9] A.E. Bayer, J.C. Smart and G.W. McLaughlin, Mapping intellectual structure of a scientific subfield through author cocitations, *Journal of the American Society for Information Science* 41(6) (1990) 444–452.

[10] P. Kärki, Searching for bridges between disciplines: an author co-citation analysis on the research into scholarly communication, *Journal of Information Science* 22(5) (1996) 323–334.

[11] P. Lenk, Mappings of fields based on nominations, *Journal of the American Society for Information Science* 34(2) (1983) 115–122.

[12] H.D. White and B.C. Griffith, Author cocitation: a literature measure of intellectual structure, *Journal of the American Society for Information Science* 32 (1981) 163–171.

[13] H.D. White, A cocitation map of the social indicators movement, *Journal of the American Society for Information Science* 34(5) (1983) 307–312.

[14] K.W. McCain, Longitudinal author cocitation mapping: the changing structure of macroeconomics, *Journal of the American Society for Information Science* 35(6) (1984) 351–359.

[15] K.W. McCain, Cocited author mapping as a valid representation of intellectual structure, *Journal of the American Society for Information Science* 37(3) (1986) 111–122.

[16] S.J. Pierce, Disciplinary work and interdisciplinary areas: Sociology and Bibliometrics. In: C.L. Borgman (ed.), *Scholary Communication and Bibliometrics* (Sage Publications, London, 1990) 46–58.

[17] B.C. Griffith, Understanding science: studies of communication and information. In: C.L. Borgman, (ed.), *Scholary Communication and Bibliometrics* (Sage Publications, London, 1990) 31–45.

[18] R.R. Braam, Foci of intellectual interest in scientific literature. In: R.R. Braan (ed.), *Mapping of Science: Foci of Intellectual Interest in Scientific Literature* (DWSO Press, Leiden, Netherlands, 1991) 267–295.

[19] E. Garfield, *Citation Indexing – Its Theory and Application in Science, Technology and the Humanities* (ISI, Philadelphia, 1979).

[20] Y. Ding, Visualization of intellectual structure in information retrieval: author cocitation analysis, *International Forum on Information and Documentation* 23(1) (1998) 25–36.

[21] J.B. Kruskal and M. Wish, *Multidimensional Scaling* (Sage Publications, California, 1978).

[22] I. Borg, *Modern Multidimensional Scaling: Theory and Applications* (Springer, New York, 1997).

[23] M.S. Aldenderfer and R.K. Blashfield, *Cluster Analysis* (Sage Publications, California, 1984).

[24] Y. Ding, *Bibliometric Analysis of Information Transfer in the Field of Information Retrieval* (First-year report) (Nanyang Technological University, 1998).

[25] M.S. Lewis-Beck, *Factor Analysis and Related Techniques* (Sage Publications, London, 1994).

[26] K.W. McCain, Mapping authors in intellectual space: a technical overview, *Journal of the American Society for Information Science* 41(6) (1990) 433–443.