Intl. Inform. & Libr. Rev. (1999), **30**, 367–376
Article No. lr990103
Available online at http://www.idealibrary.com on IDEAL®

A Bibliometric Analysis of Collaboration in the Field of Information Retrieval

YING DING,* SCHUBERT FOO* & GOBINDA CHOWDHURY*

ABSTRACT

Collaboration practices vary greatly per scientific area and discipline and influence the scientific performance and its scholarly communication. In this study, the collaborative pattern of the Information Retrieval (IR) research field is analyzed using co-authored articles retrieved from *Social Science Citation Index* for a period of 11 years from 1987 to 1997. The level of collaboration, journal collaborative distribution, disciplinary collaborative distribution and country collaboration are probed according to IR collaborative research. Findings are discussed from the above perspectives in detail. In particular, this study reveals a perceptible upward trend of collaborative IR research with the results of these research efforts being reported in all major core IR journals. The inter-disciplinary and intra-disciplinary scholarly communications in collaborative researches are very much in evidence and cover broad areas like psychology, and computer and medical sciences, respectively.

INTRODUCTION

Communication and collaboration between researchers are of great importance in the development of subject areas and in the dissemination of research results. As the new results and investigations filter through the network of interested parties, new insights are obtained and people are inspired to work on the same or related research fields. People co-operate to investigate problems that are almost impossible to solve by an individual working alone.

* Division of Information Studies, School of Applied Science, Nanyang Technological University, Nanyang Avenue, Singapore 639798.

1057–2317/98/040367 + 10 $30.00/0

Scientists do not work in isolation.[1] In a very general sense, all scientists are members of a world-wide community working together to probe and understand the mysteries of nature. The universalism of science and the inter-dependence of scientists across cultural and geographical interfaces provide us with a reliable framework to study the generation, processing, and communication of scientific knowledge. Collaboration is an intense form of interaction that allows for effective communication as well as the sharing of competence and other resources.[2] Looking at the dramatic increase of multi-author articles between individual scientists as well as among research institutions or universities,[3] one is inclined to assume that collaboration has become a prerequisite for modern science. The tremendous growth of collaborations among nations and research institutions witnessed during the last 20 years is a function of the internal dynamics of science as well as science policy initiatives.[4]

Investigating the relationships found in the documentation of a subject field is one method of examining the communication taking place in the field. Bibliometrics provides a method for examining communication among scholars in a field through their scholarly publications.[5] Documented communication may offer important insights into patterns of relationships, research focus, inter-disciplinary links and changes in communication over time. Co-authorship (collaboration) appears as a central consideration in investigations of communication patterns linking scholars in a subject field, and along with other social relationships contributes to a "network" facilitating communication among scholars.[6] Hence, it appears appropriate to examine the characteristics of the co-author relationship that may influence communication structures and to identify characteristics that might be generalized to other bibliometric structures as well.

The degree of collaboration varies from one discipline to another. It is generally high in the intensely collaborative scientific and technical fields, but low in the humanities. For example, Garfield[7] reported that multi-authored papers accounted for only 17 to 25% of samples of published

[1] Cronin, B. (1982) Invisible colleges and information transfer: a review and commentary with particular reference to the social sciences. *Journal of Documentation* **38**(3), pp. 212–236.

[2] Heffner, A.G. (1981) Funded research, multiple authorship, and subauthorship collaboration in four disciplines. *Scientometrics* **3**(1), pp. 5–12.

[3] Melin, G. & Persson, O. (1988) Hotel cosmopolitan: a bibliometric study of collaboration at some European universities. *Journal of the American Society for Information Science* **49**(1), pp. 43–48.

[4] Luukkonen, T., Tijseen, R. J. W., Persson, O. & Sivertsen, G. (1993) The measurement of international scientific collaboration. *Scientometrics* **28**(1), pp. 15–36.

[5] Subramanyam, K. (1983) Bibliometric studies of research collaboration: a review. *Journal of Information Science* **6**, pp. 33–38.

[6] Melin, G. & Persson, O. (1996) Studying research collaboration using co-authorships. *Scientometrics* **36**(3), pp. 363–377.

[7] Garfield, E. (1979) Is citation analysis a legitimate evaluation tool? *Scientometrics* **1**(4), pp. 359–375.

papers in economics, social work and sociology; but in gerontology, psychiatry, psychology and biochemistry, multi-authored papers constituted 47 to 81% of the samples. Studies by Price and Beaver,[8] Zuckerman[9] and Pao[10] have shown a strong association between collaboration and productivity. They noted the existence of a small core of extremely active researchers, surrounded by a large floating population of people who collaborated with leaders on only one or two projects and then disappeared. Price and Beaver[11] suggested that "part of the social function of collaboration is that it is a method of squeezing papers out of the rather large population of people who have less than a whole paper in them."

Information retrieval is often regarded as synonymous with document retrieval, and nowadays with text retrieval, implying that the task of an information retrieval system is to retrieve documents or texts with information content that is relevant to a user's information need.[12] Investigation on collaborative patterns in IR can reflect the nature, dynamism, and other characteristics of the discipline.

Findings of such studies can be usefully employed in research planning and organizing information resources and services more effectively and efficiently. This facilitates and accelerates research activities in IR. The main objective of this study is to investigate the collaborative trends in the research discipline of IR for an 11-year period between 1987 and 1997.

Methodology

Authorship is an observable phenomenon reflecting contemporary practices by clearly showing patterns of communication, productivity, and collaboration and influences among researchers even though their quantities and qualities are not well understood. Thus, the knowledge of this relationship is useful in the study of their strength in communication among scholars. It also helps in assessing scientific productivity and in determining research patterns of a subject or speciality. The bibliometric measures can reflect collaboration at individual (number of authors), organizational (number of countries), media (number of different journals), and disciplinary (subject areas) levels.

[8] Price, D. J. de. S. & Beaver, D. de B. (1966) Collaboration in an invisible college. *American Psychologist* **21**, pp. 1011–1018.

[9] Zuckerman, H.A. (1968) Patterns of name ordering among authors of scientific papers: A study of social symbolism and its ambiguity. *American Journal of Sociology* **74**, pp. 276–291.

[10] Pao, M.L. (1981) Coauthorship as communication measure. *Library Research* **2**, pp. 327–338.

[11] Price & Beaver (1966) *op. cit.*

[12] Sparck Jones, K. & Willett, P. (1997) Overall introduction. *In* Sparck Jones, K. & Willet, P. (Eds) *Readings in Information Retrieval*. San Francisco. Morgan Kaufmann Inc., pp. 1–7.

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 the *Social Science Citation Index* (SSCI) and *Science Citation Index* (SCI). However, a preliminary test conducted on SCI and SSCI confirmed that there are about 4500 publications on the information retrieval field during the 11-year period of study (i.e., between 1987 and 1997). Thus, including both sets of data would prove too much to manage within a stipulated period of time because the total number of citations in those 4500 or so publications would be more than 100 000. Therefore, this study focuses on data in SSCI. All relevant papers were selected in SSCI via Dialog during the period of 1987–1997. Other useful records were added from the *Library and Information Science Abstracts* (LISA) CD-ROM. Among them, four articles were excluded because of missing authors. Finally, a total of 1462 IR-related papers were selected from 367 journals with 44836 citations.

Findings

Level of Collaboration

In IR, the collaboration degree is 0·48 among the research sample, which means that 48% of the papers were written by more than one author. This degree gives a fairly clear idea of the extent of collaboration in IR.

Fig. 1 and Table I show the patterns of inter-personal collaboration among researchers as measured by multiple authorship and the yearly

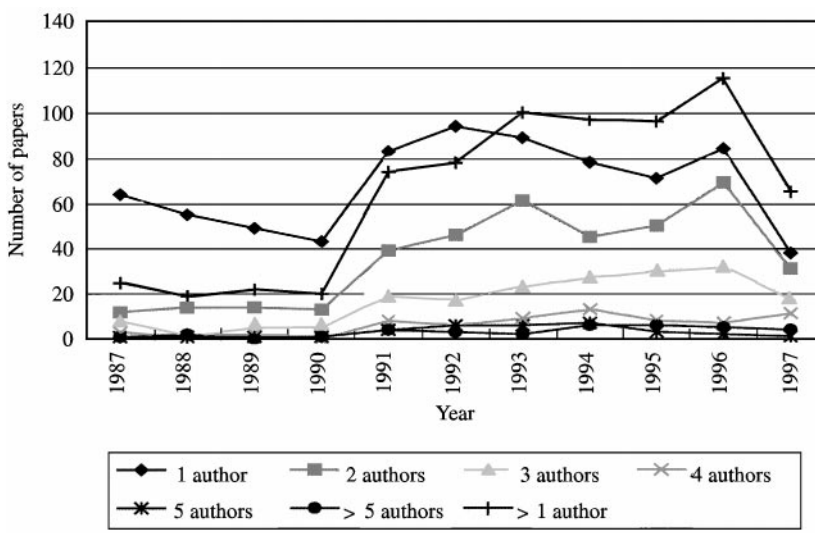

Figure 1. Distribution of collaborative research in IR

TABLE I

Distribution of IR papers according to collaboration of authors

		One author		Two authors		Three authors		Four authors		Five authors		> Five authors			
Period	Total items	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	Authorship	Authorship per items
1987	89	64	71·9	12	13·5	8	9·0	3	3·4	1	1·1	1	1·1	135	1·52
1988	74	55	74·3	14	18·9	1	1·4	1	1·4	1	1·4	2	2·7	110	1·49
1989	71	49	69·0	14	19·7	5	7·0	2	2·8	1	1·4	0	0·0	105	1·48
1990	63	43	68·3	13	20·6	5	7·9	0	0·0	1	1·6	1	1·6	95	1·51
1991	157	83	52·9	39	24·8	19	12·1	8	5·1	4	2·5	4	2·5	295	1·88
1992	172	94	54·7	46	26·7	17	9·9	6	3·5	6	3·5	3	1·7	309	1·80
1993	190	89	46·8	61	32·1	23	12·1	9	4·7	6	3·2	2	1·1	359	1·89
1994	176	78	44·3	45	25·6	27	15·3	13	7·4	7	4·0	6	3·4	376	2·14
1995	168	71	42·3	50	29·8	30	17·9	8	4·8	3	1·8	6	3·6	344	2·05
1996	199	84	42·2	69	34·7	32	16·1	7	3·5	2	1·0	5	2·5	386	1·94
1997	103	38	36·9	31	30·1	18	17·5	11	10·7	1	1·0	4	3·9	233	2·26
Total	1462	748	51·2	394	26·9	185	12·7	68	4·7	33	2·3	34	2·3	2747	1·88

distribution of collaboration level in the IR field. Among the total 1462 items, 748 appeared as single-authored items; 394 as two-authored items; 185 as three-authored items; 68 as four-authored items; 33 as five-authored items; with the remaining 34 items with more than five joint authors. Among collaborative papers, two-author and three-author papers account for 81% (with two-author papers accounting for 55% of this figure). There is a perceptible increase in the collaborative research in the field of IR over this period from 1987 to 1997. The findings show that there is a decreasing trend of single-authored items from 71.9% in 1987 to 36.9% in 1997. The most popular collaboration types are notably two-author and three-author collaborations.

The reductions in single-authored papers during the period 1990–1991, 1992–1993, and 1996–1997 are very sharp and reflect the influence of development in inter- or intra-disciplinary research collaborative effort. Authorship per item has increased from an average of 1.52 per item in 1987 to 2.26 per item in 1997. The collaborative and authorship trends as shown in Fig. 1 and Table I indicate that collaborative research in IR has increased dramatically with a strong possibility that this trend will continue to increase in the future.

Journal Collaborative Distribution

In order to analyze the collaborative distribution in IR journals, 20 journals with high IR content were identified based on the research sample. The collaborative distribution in these journals is shown in Fig. 2. *Information Processing & Management* and *Journal of the American Society for Information Science* are the top two journals with high IR content, but their collaborative degrees (collaborative IR papers over total IR papers in one particular journal) are not more than others. The journals with highest collaborative degrees are the following: *Journal of the American Medical Informatrics Association* (76.5%), *ACM Transaction on Information Systems* (61.1%), *Journal of Information Science* (53.1%). In contrast, the journals with the lowest collaborative degrees are *Library Quarterly* (8.1%) and *Journal of Academic Librarianship* (9.1%).

Scientific co-operation can be investigated from the point of view of journal collaborative distributions. Almost all the core IR journals published IR collaborative papers. At the same time, the collaborative degree of each journal is not lower than expected. There is a high potential that more and more collaborative papers will appear in these IR core journals.

Disciplinary Collaborative Distribution

In this study, disciplinary classification is based on the SSCI journal category. An examination of the lists of the journals in which collaborative IR papers were published reveals the diversity of inter- and intra-disciplinary

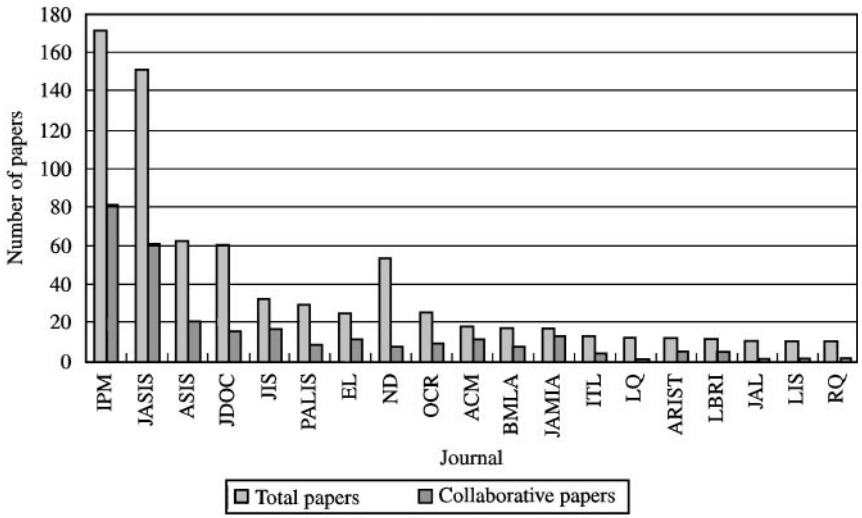

FIGURE 2. Collaborative research distribution among journals. Journal codes: IPM=Information Processing & Management; JASIS=Journal of the American Society for Information Science; ASIS=Proceedings of the ASIS annual meeting; JDOC=Journal of Documentation; JIS=Journal of Information Science; PALIS=Program-Automated Library and Information Systems; EL=Electronic Library; ND=Nachrichten Fur Dokumentation; OCL=Online & CD ROM Review; ACM=ACM Transactions on Information Systems; BMLA=Bulletin of the Medical Library Association; JAMIA=Journal of the American Medical Informatics Association; ITL=Information Technology and Libraries; LQ=Library Quarterly; ARIST=Annual Review of Information Science and Technology; JAL=Journal of Academic Librarianship; LIS=Library and Information Science; RQ=Research Quarterly

scholarly communication in IR collaboration research as shown in Fig. 3. The inter-disciplinary scholarly communication covers very broad research areas including psychology, computer science, medical science, ergonomics, business, engineering, management, multi-disciplinary science, behavioral science, educational science, geography, biology, communication science, physics, chemistry, material science, agriculture, and so on. In particular, psychology, computer science, and medical science have vital effects in IR collaboration research during inter-disciplinary scholarly communication processes. In intra-disciplinary scholarly communication, some subfields of IR play very important roles in IR collaboration research, such as computer science in IR, computer application in IR, and medical library science in IR.

Country Collaborative Distribution

Studies on international collaboration have developed in the last few years as a consequence of the fact that the international dimension has become a

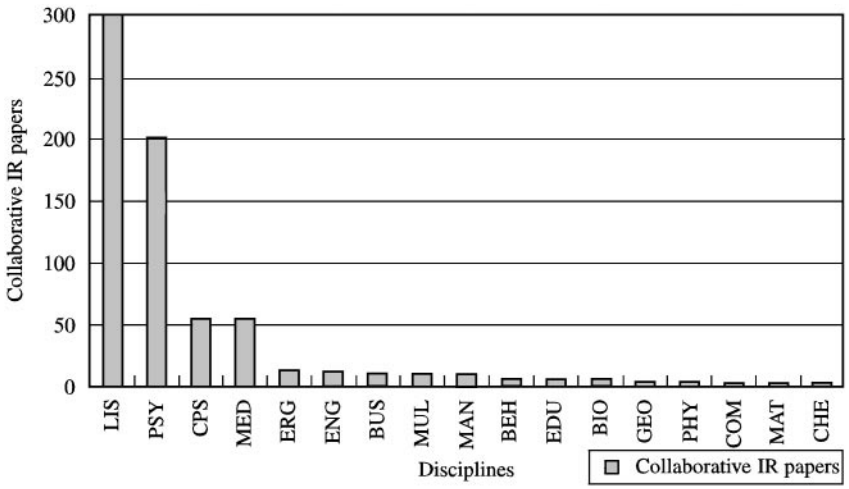

FIGURE 3. Collaborative research distribution among disciplines. Discipline codes: LIS=Library and Information Science; PSY=Psychology; CPS=Computer Science; MED=Medical Science; ERG=Ergonomics; ENG=Engineering; BUS=Business; MUL=Multidisciplinary Science; MAN=Management; BEH=Behavioral Science; EDU=Education; BIO=Biology; GEO=Geography; PHY=Physics; COM=Communication Science; MAT=Material Science; CHE=Chemistry

more essential part of scientific activities. International co-authored articles are frequently used as a measure to obtain a comprehensive picture of international collaborative works. Co-authorship is a final result of diverse sequences of scientific exchanges that facilitate the acquisition of science undertaken within a community of facts and ideas. It can be used to study the country-interactions in science as a whole or within major science fields.[13–16]

The dependence on the international scene is proportionately higher for smaller countries. This is a logical consequence of the fact that the smaller a country is the greater the share of scientists outside it, so that there is a need, and hence, a greater chance of international collaboration to achieve recognition. Data on country-to-country co-authorship was generated from the research sample. Thirty-two of the most productive countries were selected

[13] Okubo, Y., Miquel, J.F., Frigoletto, T. & Dore, J.C. (1992) Structure of international collaboration in science: typology of countries through multivariate techniques using a link indicator. *Scientometrics* **25**(2), pp. 321–351.

[14] Luukkonen *et al.* (1993) *op. cit.*

[15] Arunachalam, S., Srinivasan, R. & Raman, V. (1994) International collaboration in science: Participation by the Asian Giants. *Scientometrics* **30**(1), pp. 7–22.

[16] Vinkler, P. (1993) Research contribution, authorship and team cooperativeness. *Scientometrics* **26**(1), pp. 213–230.

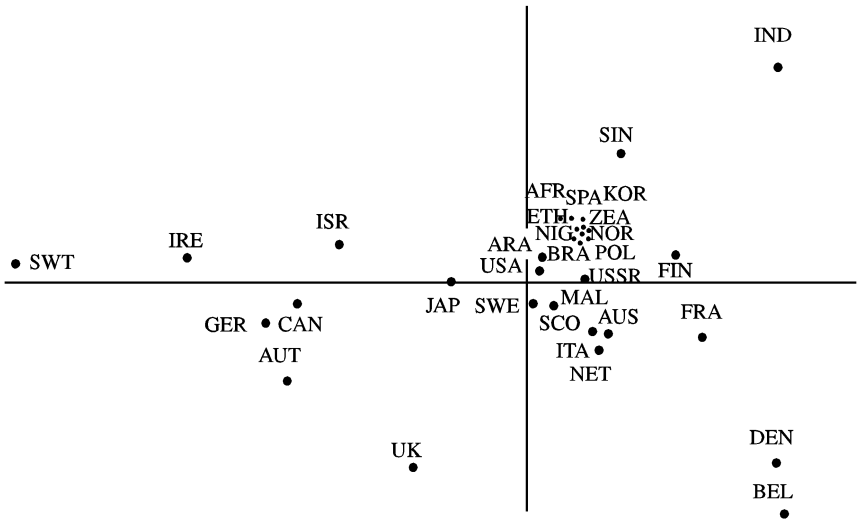

FIGURE 4. Map of international co-authorship relationships among 32 countries. Country codes: AFR=South Africa; ARA=Saudi Arabia; AUS=Australia; AUT=Austria; BEL=Belgium; BRA=Brazil; CAN=Canada; DEN=Denmark; ETH=Ethiopia; FIN=Finland; FRA=France; GER=Germany; IND=India; IRE=Ireland; ISR=Isarel; ITA=Italy; JAP=Japan; KOR=South Korea; MAL=Malaysia; NET=Netherlands; NIG=Nigeria; NOR=Norway; POL=Poland; SCO=Scotland; SIN=Singapore; SPA=Spain; SWE=Sweden; SWT=Switzerland; ZEA=New Zealand

and the number of co-authorships for each pair of countries was counted according to the research sample. The Jaccard measure[17] was chosen to normalize the matrix, which can eliminate the difference of the collaboration of smaller countries with larger countries. Fig. 4 presents a map based on the Pearson's correlation coefficients. The coordinates of the map were found by using a multi-dimensional scaling program with 81.2% fit.

In Fig. 4, countries with similar co-authorship profiles will be close to each other and countries with low correlations will be located far apart. Based on the map, we can assume that geographical distance is the major force at work, since there is apparently a west–east and a north–south dimension that separates the countries. This provides strong evidence to support one claim that factors such as greater geographical distance with the additional travel or telecommunication cost and time involved, are impediments to IR collaboration. This result coincides with the research conducted by Katz.[18] We can easily identify the distinct Nordic and Asian regions. The USA is located near the origin, linking Africa, Asia, Latin

[17] Luukkonen *et al.* (1993) *op. cit.*

[18] Katz, J.S. (1994) Geographical proximity and scientific collaboration. *Scientometrics* **31**(1), pp. 31–43.

America and Europe, which suggests that the USA collaborates over a broad region.

Discussion

The results of co-authorship studies can be used in a research policy perspective. It provides an overview of the main features of the scholarly communication process of one research area; the collaboration can be seen from the perspective of one paper, a specific journal, a particular discipline and a particular country. Especially if the co-authorship patterns are studied over time, there is a possibility to test or evaluate various assumptions and science policies, in so far as they relate to scientific collaboration.

The following general conclusions on IR collaborative research in the scholarly communication process can be drawn from the findings of this study:

(1) There is a perceptible increase in collaborative research in the field of IR from 1987 to 1997. The authorship per paper has increased from 1.52 per paper in 1987 to 2.26 per paper in 1997.
(2) IR collaborative papers appear in almost all the core IR journals.
(3) The inter- and intra-disciplinary scholarly communications on IR collaborative research cover very broad areas with psychology, computer science, and medical science being the most distinctive.
(4) Geographical proximity plays a significant role in IR collaboration. The greater the geographical distance the greater the impediment to IR collaboration. It will be interesting to see if the advent of the Internet, and its ability to support collaborative group work, will close this gap in international collaboration in the future.

Acknowledgements

We are grateful to Professor Roderick Cave (Nanyang Technological University, Singapore) for his kind supervision of this
work and valuable suggestions.