

A Bibliometric Analysis Of Collaboration In The Field of Information Retrieval

Ying Ding, Schubert Foo, Gobinda Chowdhury

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

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 perspective in detail.

In particular, this study reveals a perceptible upward trend of collaborative IR research with the results of these research efforts been 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 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 cooperate to investigate problems that are almost impossible to solve by an individual working alone.

Scientists do not work in isolation (Cronin, 1982). 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 interdependence 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 (Heffner, 1981). Looking at the dramatic increase of multi-authors articles between individual scientists as well as among research institutions or universities (Melin and Persson, 1998), one is inclined to assume that collaboration has become a prerequisite for modern science. The tremendous growth of collaboration among nations and research institutions witness during the last twenty years is a function of the internal dynamics of science as well as science policy initiatives (Luukkonen et al., 1993).

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 (Subramanyam, 1983). Documented communication may offer important insights into patterns of relationships, research focus, interdisciplinary links, and changes in communication over time. Coauthorship (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 (Melin and Persson, 1996). Hence, it appears appropriate to examine the characteristics of the coauthor relationship that may influence communication structures and to identify characteristics that might be generalized to other bibliometric structures as well.

Information retrieval (IR), a subfield of library and information science, is concerned with the finding and recall of information from a store (Prytherch, 1995). Investigation on collaborative patterns in IR can reflect the nature, dynamism and other characteristics of the discipline. Findings of such studies can usefully be 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 a 11 year period between 1987 to 1997.

Methodology

Authorship is an observable phenomenon reflecting the 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 specialty. 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.

IR is a very broad-based research area. In order to make this IR study manageable, it has been limited to those literature reported in social science, especially in the domain of library and information science. Bruin and Moed (Bruin, 1993) compared the ISI databases with several other databases in studying corporate address and found that the ISI databases have more complete information on affiliations and address than other databases. The data were collected from all the IR papers published from 1987 to 1997 and covered by the *Social Science Citation Index* (SSCI) database. Among them, four articles were excluded because of missing addresses. A total of 1,462 journal articles in 367 different journals were collected from SSCI as the research sample in the current research.

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.

[Figure 1 and Table 1 about here]

Figure 1 and Table 1 show the patterns of inter-personal collaboration among researchers as measured by multiple authorship and the yearly distribution of

collaboration level in the IR field. Among the total 1,462 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). Thus, 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 which reflects the influence of development in inter- or intra-disciplinary research collaborative effort. Authorship per item has been increasing 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 Figure 1 and Table 1 indicate that collaborative research in IR has increased dramatically with a strong possibility that this trend will continue to increase in future.

Journal collaborative distribution

[Figure 2 about here]

In order to analyse the collaborative distribution in IR journals, twenty journals with high IR content were identified based on the research sample. The collaborative distribution in these journals was shown in Figure 2. *Information Processing & Management* and *Journal of the American Society for Information Science* are the first

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 followings: *Journal of the American Medical Informatics 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 cooperations 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 has high potential that more and more collaborative papers will appear in these IR core journals.

Disciplinary collaborative distribution

[Figure 3 about here]

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 reveal the diversity of inter and intra-disciplinary scholarly communication in IR collaboration research as shown in Figure 3. The inter-disciplinary scholarly communication cover very broad research areas including Psychology, Computer Science, Medical Science, Ergonomics, Business, Engineering, Management, Multidisciplinary 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 process. 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 consequence of the fact that the international dimension has become a more essential part of scientific activities. In order to obtain a comprehensive picture of international collaborative works, international co-authored articles are frequently used as a measure of this activity. 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 (Okubo et al., 1992; Luukkonen et al., 1993; Arunachalam et al., 1994; Vinkler, 1993).

[Figure 4 about here]

The dependence on the international scene is proportionately higher for smaller countries. This is a more or less logical consequence of the fact that the smaller a country is the greater is the share of scientists outside it, so that there is a need, hence, a greater chance of international collaboration to achieve recognition. Data on country-to-country co-authorship was generated from the research sample. 32 of the most productive countries were selected and the number of co-authorship for each pair of countries was

counted according to the research sample. The Jaccard measure was chosen to normalize the matrix, which can eliminate the difference of the collaboration of smaller countries with larger countries (Luukkonen et al., 1993). Figure 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 Figure 4, countries with similar coauthorship profiles will be closed 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 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 (1994). We can easily identify the distinct Nordic and the Asian region. USA is located near the origin, linking Africa, Asia, Latin America and Europe, which suggests that USA collaborates on a broad region.

Discussion

The results of coauthorship studies can be used in a research policy perspective. It provide 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 disciplinary and a particular country. Especially, if the coauthorship 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. This is a perceptible increase in the collaborative research in the field of IR from 1987 to 1997. The authorship per paper has been increasing 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 researches cover very broad areas with Psychology, Computer Science and Medical Science as the most distinctive areas.
4. Geographical proximity plays a significant role in IR collaboration. The greater the geographical distance the greater the impediment to IR collaboration. It would be interesting to see if the advent of Internet and its ability to support collaborative group work will close this gap of international collaboration in future.

Acknowledgment

We are grateful to Prof Roderick Cave (Nanyang Technological University in Singapore) for valuable suggestion and kindly supervision.

References

- Arunachalam, S., Srinivasan, R. and Raman, V. (1994). International collaboration in science: Participation by the Asian Giants. *Scientometrics*, 30(1), 7-22.
- Bruin, R. E. & Moed, H. F. (1993). Delimitation of scientific subfields using cognitive words from corporate addresses in scientific publications. *Scientometrics*, 26, 65-80.

Cronin, B. (1982). Invisible colleges and information transfer: A review and commentary with particular reference to the social sciences. *Journal of Documentation*, 38(3), 212-236.

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

Katz, J. S. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, 31(1), 31-43.

Luukkonen, T., Tijssen, R. J. W., Persson, O. and Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics*, 28(1), 15-36.

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

Melin, G. & Persson, O. (1998). Hotel cosmopolitan: A bibliometric study of collaboration at some European universities. *Journal of the American Society for Information Science*, 49(1), 43-48.

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), 321-351.

Prytherch, R. (1995). *Harrod's librarians' glossary: 9,000 terms used in information management, library science, publishing, the book trades and archive management*. UK: Gower Publishing Company Limited.

Subramanyam, K. (1983). Bibliometric studies of research collaboration: A review. *Journal of Information Science*, 6, 33-38.

Vin Vinkler, P. (1993). Research contribution, authorship and team cooperativeness. *Scientometrics*, 26(1), 213-230.

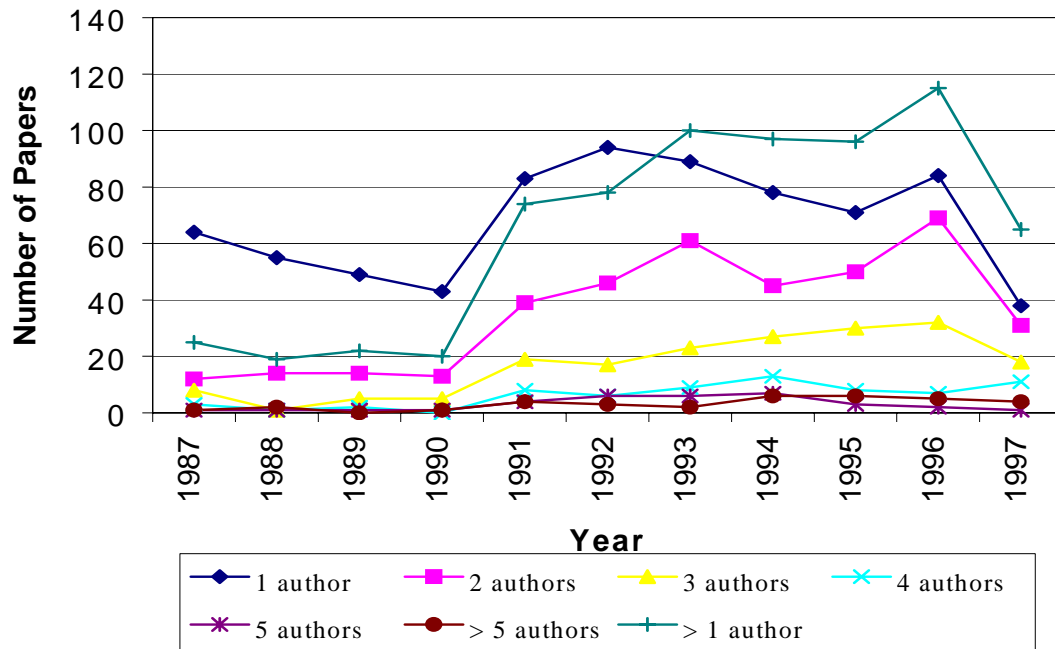
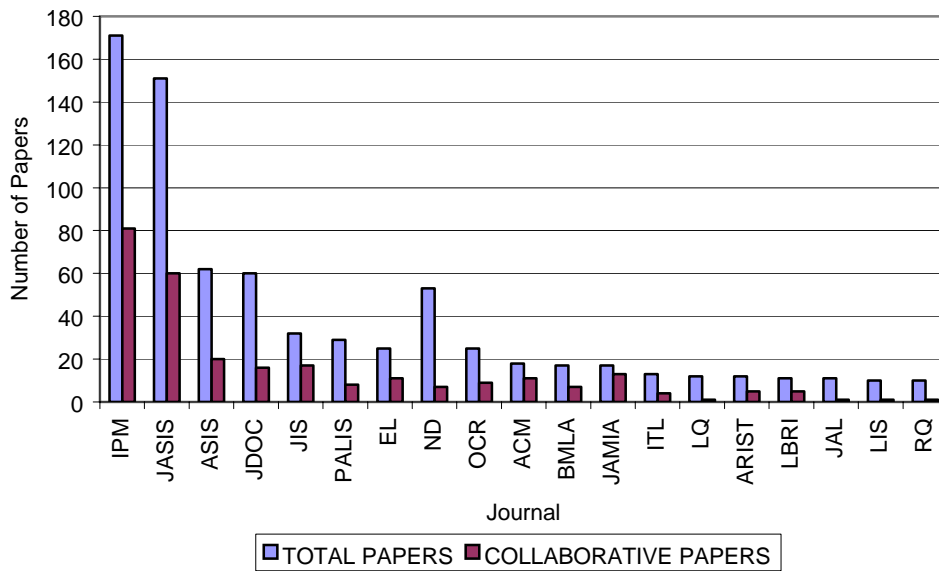
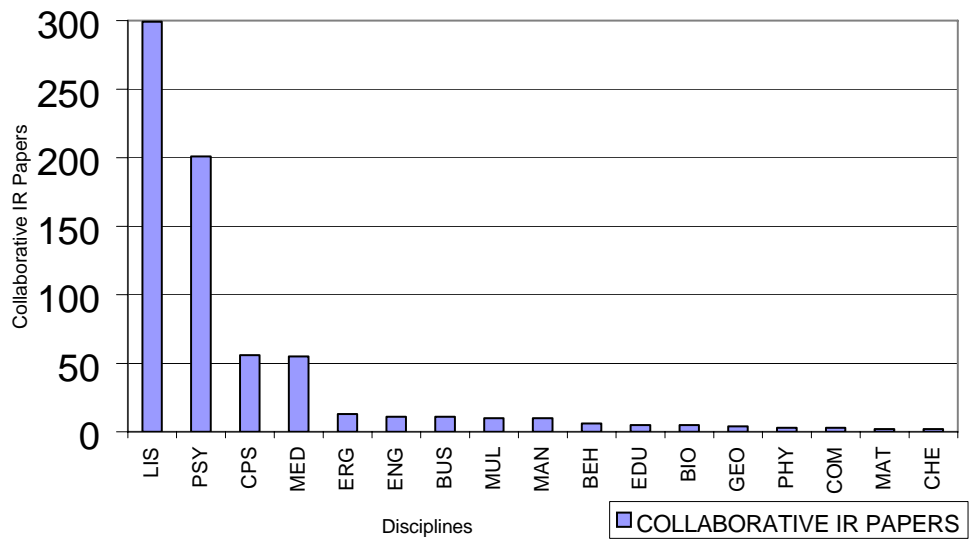


Figure 1. Distribution of Collaborative Research in IR



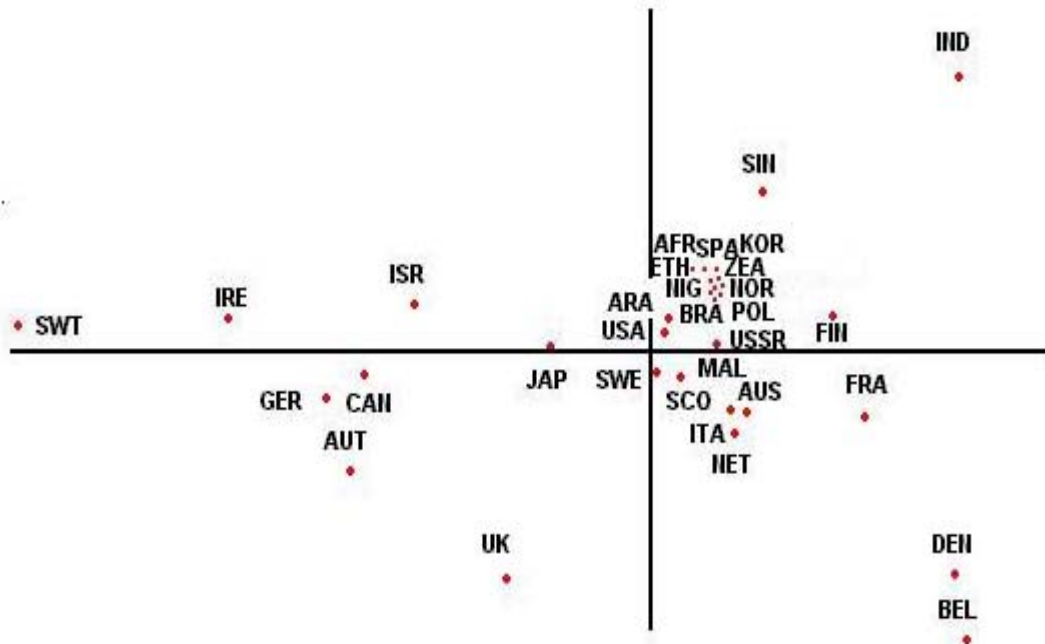
Journal codes: IPM= Information Processing & Management; JASIS= Journal of the American Society for Information Science; ASIS = Proceedings of the ASIS annual meeting; JDOC= Journal of Documentation; JIS= Journal of Information Science; PALIS= Program-Automated Library and Information Systems; EL= Electronic Library; ND= Nachrichten Fur Dokumentation; OCL= Online & CDROM 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.

Figure 2. Collaborative Research Distribution Among Journals.



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.

Figure 3 Collaborative Research Distribution Among Disciplines



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= Israel; 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; SWI= Switzerland; ZEA= New Zealand.

Figure 4. Map Of International Co-Authorship Relationships Among 32 Countries.

Table 1. Distribution of IR Papers According to Collaboration of Authors

PERI OD	TOTAL ITEMS	ONE AUTHOR NO. %	TWO AUTHORS NO. %	THREE AUTHORS NO. %	FOUR AUTHORS NO. %	FIVE AUTHORS NO. %	> FIVE AUTHORS NO. %	AUT HOR- SHIP	AUTHOR -SHIP 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