Bibliometric studies of research collaboration: A review

K. Subramanyam
School of Library and Information Science, Drexel University, Philadelphia, PA 19104, U.S.A.

Received 19 February 1982
Revised version received 3 May 1982

Scientific research is becoming an increasingly collaborative endeavour. The nature and magnitude of collaboration vary from one discipline to another, and depend upon such factors as the nature of the research problem, the research environment, and demographic factors. Earlier studies have shown a high degree of correlation between collaboration and research productivity, and between collaboration and financial support for research. The extent of collaboration cannot be easily determined by traditional methods of survey and observation. Bibliometric methods offer a convenient and non-reactive tool for studying collaboration in research. In this paper, several types of collaboration have been identified, and earlier research on collaboration has been reviewed. Further research is needed to refine the methods of defining and assessing collaboration and its impact on the organization of research and communication in science.

1. Introduction

Scientists do not work in isolation. In a very general sense, all scientists are members of a world-wide community working together to probe and understand the mysteries of nature, and to provide the theoretical basis upon which the superstructure of technology can be built. Although the organizational dynamics of scientific research and technological development are influenced in each country by the political power structure and socio-economic infrastructure, scientific knowledge is (or at least ideally, should be) supra-national in character. This is where science differs from technology. There is no such thing as 'American physics', 'Russian chemistry', or 'Chinese thermodynamics'. In so far as the laws of nature are immutable and uninfluenced by socio-economic and political considerations, science will always be supra-national. But technology, which is the application of scientific knowledge to achieve pragmatic goals can, and often is, coloured by such considerations, partly because of the effect on the practice of technology of environmental variables such as climate and natural resources. Hence in such areas of technology as building construction and agriculture, for example, there are vast differences in the practices of different countries.

This universalism of science and the interdependence of scientists across cultural and geographical interfaces provides us with a reliable framework to study the generation, processing, and communication of scientific knowledge. There are certain norms that are widely accepted and followed by scientists all over the world. One such norm is the established practice of giving credit to earlier records of science when the information contained in them is used in a subsequent investigation. Melvin Weinstock has identified 15 reasons why scholars cite earlier publications [19]. This practice of making references to earlier works has given information scientists and sociologists of science a powerful tool for the analysis and understanding of phenomena concerning the generation, communication, and use of scientific information.

There are no doubt certain pitfalls that are inevitable in the use of citation analysis as a research method. Some of the limitations and assumptions associated with citation analysis have been explicated earlier [7,17]. In this paper, several earlier studies of collaboration will be identified, and the possibility of using bibliometric methods to study the phenomenon of collaboration in scientific research will be discussed.

2. Types of collaboration

No scientific investigation can take place without the use of prior knowledge. The greatest of scientists like Newton and Einstein have been able to advance the frontiers of scientific knowledge only by standing on the shoulders of giants. Even the secluded solo researcher is indebted to his forerunners. In a more pragmatic sense, though,

North-Holland Publishing Company
Journal of Information Science 6 (1983) 33–38

collaboration in research is said to have taken place when two or more investigators work together on a project and contribute resources and effort, both intellectual and physical. Depending on the participants, the following kinds of collaboration can be identified:

(1) *Teacher–pupil collaboration.* This is a very common mode of collaboration in an academic setting. The professor in a university department provides the ideas and guidance, and sometimes also the funds from a research grant, and the research assistant or student does most of the bench work. The resulting project report, conference paper, or journal article usually carries the names of both the professor and the student. It is not uncommon for a professor to be guiding several students in different research projects at the same time.

(2) *Collaboration among colleagues.* It is a very common practice in corporate research centers for a number of colleagues to be working on one or more projects, each contributing expertise in a different aspect of the project. In interdisciplinary fields such as environment, energy, or space research, scientists and engineers from a wide variety of specialities often collaborate. It is not uncommon for chemists, chemical engineers, materials engineers, biophysicists, and other specialists to be working together in an interdisciplinary project. Husband-and-wife teams can also be included in this category.

(3) *Supervisor–assistant collaboration.* Earlier studies on the sociology of science, for example, by Cole and Cole [5], have shown the existence of a stratified structure within the scientific community. In research projects requiring extensive use of laboratory facilities or very specialized equipment, the principal investigator is often assisted by an array of laboratory assistants and technicians.

(4) *Researcher–consultant collaboration.* In large-scale research projects, the individual researcher or the research team can secure the assistance of a consultant or a consulting firm for specialized tasks such as data collection (involving, for eample, the design and administration of questionnaires, or conducting interviews), data processing and analysis.

(5) *Collaboration between organizations.* Scientists and engineers employed in different organizations often collaborate on research projects of mutual interest. Such collaboration may be spurred by informal contacts or prior acquaintance of the researchers. It is also possible that when a scientist leaves an organization and joins another, he or she may carry on an unfinished research project in the new organization with the continued collaboration of former colleagues. Inter-organizational collaboration may also be necessitated by a community of concerns (as between two government agencies) or by the complexity of a research project, or when researchers in one organization may need to use expensive equipment or specialized service available at another organization. According to recent data published by the U.S. National Science Foundation, research collaboration between industries and academic institutions has been gradually increasing [11].

(6) *International collaboration.* International collaborative behaviour among scientists has been studied by Frame and Carpenter [6]. The degree of collaboration was found to be higher in basic fields of science (such as physics, mathematics, and chemistry) than in applied fields (such as engineering and technology, clinical medicine and biomedical research). Frame and Carpenter further found that

(a) the extent of international collaboration was inversely proportional to the size of a country's scientific enterprise, and

(b) extra-scientific factors such as geography, politics, and language, played a strong role in determining who collaborates with whom in the international scientific community.

An interesting but rather rare type of collaboration can be seen in the work of the imaginary polycephalic mathematician named Nicolas Bourbaki. A group of young French mathematicians collaborated and wrote many volumes of an extraordinary treatise on mathematics under this whimsical pseudonym derived from the name of a general in the Franco–Prussian War [9].

3. Levels of collaboration

Collaboration in research can take many forms of activity ranging from offering general advice and opinion to active and sustained participation and contribution of physical and intellectual resources. Scientists employed in different organizations may collaborate by sharing data or ideas through correspondence or at conferences, visiting each other's research facilities, or actually performing parts of a project separately and then integrating the results.

Heffner has characterized collaboration as being 'theoretical' (rendering advice, ideas, or criticism), or 'technical' (providing tangible assistance in a research endeavour) [10]. He has also distinguished between 'coauthors' (i.e. those who share authorship of a publication), and 'subauthors' (i.e. those who are not coauthors, but whose theoretical or technical assistance in the research project is acknowledged by the author or authors in a publication).

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 in which the lonely scholar, working without the trappings of 'big science' still produces much of the scholarly literature. In an unpublished study by D. Lindsey and G.W. Brown, quoted by Garfield [7], multi-authored papers accounted for only 17–25% of samples of published papers in economics, social work, and sociology; but in gerontology, psychiatry, psychology, and biochemistry, multi-authored papers constituted 47–81% of the samples.

4. Bibliometric method

Collection of data in studies of research collaboration is a difficult problem. The precise nature and magnitude of collaboration cannot be easily determined by the usual methods of observation, interviews or questionnaire because of the complex nature of human interaction that takes place between or among collaborators over a period of time. Both the nature and magnitude of contribution of each collaborator are likely to change during the course of a research project. While some of the tangible aspects of a scientist's work can be quantified and expressed in input units (e.g. number of hours spent in the laboratory on a task) or in output units (e.g. number of samples analysed), the less tangible aspects (Heffner's 'theoretical' contribution) cannot be quantified easily. In any case, qualitative assessment of the contribution of each collaborator is extremely complex, if not impossible, because of the indeterminate relationship between quantifiable (technical) activities and intangible (theoretical) contributions. For example, a brilliant suggestion made by a scientist during casual conversation may be more valuable in shaping the course and outcome of a research project than weeks of labour-intensive activity of a collaborating scientist in the laboratory.

In view of these difficulties in the direct assessment of the contribution of collaborators, an unobtrusive indicator such as the number of coauthors of a research paper can be conveniently used as a measure of research collaboration. The principal advantages of using the number of coauthors as a measure of collaboration are that it is

(a) invariant,

(b) easily and inexpensively ascertainable,

(c) quantifiable, and

(d) non-reactive (i.e. the process of ascertaining collaboration does not affect the process of collaboration itself).

The bibliometric method facilitates the investigation of the relationship between research collaboration and variables pertaining to the research problem and the research environment, by applying statistical techniques such as regression, correlation, and factor analysis.

Every research method has some disadvantages, and the bibliometric method is no exception. In a study of manuscripts submitted for publication in an astronomy journal, Gordon identified three assumptions that should be considered while using the bibliometric method [8]:

(1) The number of papers produced by a group of scientists is proportional to (and hence an index of) their research activity.

(2) The relative frequency of coauthorship within such groups is proportional to (and hence an index of) the degree of scientific collaboration within the group.

(3) The relative frequency of *production* of research journal papers with different levels of multiple authorship (i.e. 1 author, 2 authors, 3 authors, etc.) is proportional to (if not equivalent to) the relative frequency of appearance of papers by groups of each size in research journals.

Gordon investigated the third assumption and found that in astronomy, there was a significant relationship between the number of authors per manuscript submitted to a journal and the rate of acceptance of the manuscripts for publication. Manuscripts with a large number of authors had, in general, a higher rate of acceptance for publication. Gordon also conjectured that similar relationships could exist in other areas of research which use large-scale, highly complex experimental or observational equipment.

The first two assumptions stated by Gordon can be amplified and restated as follows:

(1) All collaborative research effort results in one or more published papers.

(2) All the collaborators are mentioned as coauthors in the publications.

(3) All the coauthors mentioned in the publication have actually collaborated in the research effort.

None of the above assumptions can be verified easily. The practice of naming coauthors varies widely and depends on the attitudes and perceptions of the individuals involved, and the policies and traditions of the organization in which the research endeavour takes place. In some research papers, all collaborators, including laboratory assistants and statisticians who rendered technical help, are mentioned as coauthors. In papers based on extensive laboratory tests or field work, it is not uncommon to find the names of ten or more coauthors. In others, the principal investigator and only those scientists who provided substantial and sustained collaboration are named as coauthors.

The question of ordering the names of coauthors is highly complex and elusive. While it is generally true that the name of the principal investigator is almost always mentioned first, the order in which the remaining coauthors are named in the paper does not necessarily reflect the degree of collaboration. The names of coauthors are sometimes arranged in alphabetical order, except for the principal investigator's name which might be placed at the beginning or at the end [21].

The interpretation of bibliometric research studies on research collaboration should be tempered by the above assumptions and shortcomings.

5. Research trends and prospects

In 1963, Derek J. de Solla Price noted that the proportion of multiple-authored papers had accelerated steadily since the beginning of the 20th century, and that if the same trend continued, there would be no single-authored paper by 1980 [15]. Obviously, this has not happened. Although the extent of collaboration in research has steadily increased, individual research effort is nowhere near extinction. Beverly Clarke in 1964 challenged Price's contention of rapidly decreasing single-authorship, and produced data on authorship in biomedical literature from 1934 to 1963 to show that the average number of authors per paper had remained almost steady at about 2.3 during that period [4].

In a series of three articles in *Scientometrics*, Beaver and Rosen studied the history of research collaboration from 17th century onwards. These studies have shown that collaboration in scientific research is related to professionalization of the scientific community, and that collaboration generally leads to greater productivity in research and enhances the mobility and visibility of scientists [1–3].

The relationship between collaboration and financial support for research in four disciplines (political science, psychology, biological science, and chemistry) was studied by Heffner [10]. In all these disciplines, financial support for research was associated with an increase in the total number of persons (including coauthors and subauthors) involved in the production of knowledge per research paper; the association was particularly strong in biological science and chemistry. A similar correlation between collaboration and financial support for research had been reported in 1965 by W. Hirsch and J.F. Singleton in an unpublished report quoted by Price and Beaver [16].

Studies by Price and Beaver, Zuckerman, and Pao have shown a strong association between collaboration and productivity. In their study of collaboration in Information Exchange Group No. 1 (on oxidative phosphorylation and terminal electron transport, organized by the National Institutes of Health), Price and Beaver found that productivity increased as the number of collaborators increased. They noted the existence of a small core of extremely active researchers, surrounded by a large floating population of people who collaborated with the leaders on only one or two projects and then disappeared. Price and Beaver 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" [16].

Zuckerman's study of 41 Nobel laureates also showed a high degree of correlation betweeen collaboration and productivity. In general, laureates published more and were more apt to collaborate than a matched sample of scientists. Also, in arranging names of coauthors, laureates exercised a certain *noblesse oblige* and allowed their junior collaborators to be the senior authors [20].

Miranda Lee Pao has investigated the relationship between collaboration and productivity in musicology, a 'humanistic subject' [13,14]. Though only 15% of the literature of musicology was the result of collaborative authorship (as compared with 80% in scientific disciplines), the most collaborative musicologists were also the most productive. Applying a normalized diversity measure to study the productivity of authors, Pao found a high degree of correlation between productivity and collaboration in computational musicology.

McCauley has described the problem posed by multiple-authored papers in naming taxonomic species, and suggested that in taxonomic papers, the number of authors should be limited. Referring to a paper with 15 authors describing the species *Mycoplasma pneumoniae*, he commented: "The fact that each man contributed to the research... does not necessarily mean that all should be authors... the man who first recognizes a species as unique should be entitled to describe it" [12].

In a preliminary study of research collaboration in biochemistry and chemical engineering, the degree of collaboration in a discipline was defined as the ratio of the number of collaborative research papers to the total number of research papers published in the discipline during a certain period of time. This definition of collaboration can be expressed thus:

$$C = \frac{N_m}{N_m + N_s}$$

where

C = degree of collaboration in a discipline,
N_m = number of multiple-authored research papers in the discipline published during a year,
N_s = number of single-authored research papers in the discipline published during the same year.

This value of C, along with the weighted average number of authors per paper, gives a fairly clear idea of the extent of collaboration in a discipline. This preliminary study showed that

(a) the degree of collaboration is higher in biochemistry than in chemical engineering;

(b) the proportion of research papers supported by grants is significantly higher in biochemistry than in chemical engineering; and

(c) in both biochemistry and chemical engineering, collaborative research papers were supported by grants to a larger extent than were single-authored papers [18].

These investigations have indicated that

(a) there is variation from one discipline to another in the nature and magnitude of research collaboration, and

(b) collaboration is affected by various factors including availability of financial support, nature of the research problem, and the research environment.

Collaboration has also been found to affect the visibility and productivity of scientists. Further research is needed to refine the methods of defining and assessing collaboration at different levels, and to ascertain the impact of collaboration on the organization of research and communication in science.

References

[1] D. de B. Beaver and R. Rosen, Studies in scientific collaboration. Part I. The professional origins of scientific coauthorship, *Scientometrics 1* (1) (1978) 65–84.

[2] D. de B. Beaver and R. Rosen, Studies in scientific collaboration. Part II. Scientific coauthorship, research productivity and visibility in the French scientific elite, *Scientometrics 1* (2) (1979) 133–149.

[3] D. de B. Beaver and R. Rosen, Studies in scientific collaboration. Part III. Professionalization and natural history of modern scientific coauthorship, *Scientometrics 1* (3) (1979) 231–245.

[4] Beverly L. Clarke, Multiple authorship trends in scientific papers, *Science 143* (3608) (1964) 822–824.

[5] Jonathan R. Cole and Stephen Cole, *Social Stratification in Science* (University of Chicago Press, Chicago, 1973).

[6] J. Davidson Frame and Mark P Carpenter, International research collaboration, *Social Studies of Science 9* (1979) 481–497.

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

[8] M.D. Gordon, A critical reassessment of inferred relations between multiple authorship, scientific collaboration, the production of papers and their acceptance for publication, *Scientometrics 2* (3) (1980) 193–201.

[9] Paul R. Halmos, Nicolas Bourbaki, *Scientific American 196* (5) (1957) 88–99.

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

[11] Wil Lepkowski, NSF study gives global technology picture, *Chemical & Engineering New 60* (3) (1982) 49–52.

[12] James E. McCauley, Multiple authorship, *Science 141* (3580) (1963) 579.

[13] Miranda Lee Pao, Coauthorship as communication measure, *Library Research 2* (1981) 327–338.

[14] Miranda Lee Pao, Collaboration in computational musicology, *Journal of the American Society for Information Science 33* (1) (1982) 38–43.

[15] Derek J. de Solla Price, *Little Science Big Science* (Columbia University Press, New York, 1963).

[16] Derek J. de Solla Price and Donald de B. Beaver, Collaboration in an invisible college, *American Psychologist 21* (1966) 1011–1018.

[17] K. Subramanyam, Citation and significance, *New Library World 76* (905) (1975) 227–228.

[18] K. Subramanyam and Elsie M. Stephens, Research collaboration and funding in biochemistry and chemical engineering, *International Forum on Information and Documentation 7* (4) (1982) 26–29.

[19] Melvin Weinstock, Citation indexes, in: *Encyclopedia of Library and Information Science, Vol. 5* (Marcel Dekker, New York, 1971) 16–40.

[20] Harriet A. Zuckerman, Nobel laureates in science: Patterns of productivity, collaboration, and authorship, *American Sociological Review 32* (3) (1967) 391–403.

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