Conference Proceedings as a Source of Scientific Information: A Bibliometric Analysis

Cynthia Lisée and Vincent Larivière
Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, CP 8888, Succursale Centre-ville, Montréal, Québec, H3C 3P8, Canada. E-mail: c_lisee@yahoo.com; lariviere.vincent@uqam.ca

Éric Archambault
Science-Metrix, 1335A avenue du Mont-Royal E, Montréal, Québec, H2J 1Y6, Canada and Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, Montréal, Québec, Canada. E-mail: eric.archambault@science-metrix.com

While several authors have argued that conference proceedings are an important source of scientific knowledge, the extent of their importance has not been measured in a systematic manner. This article examines the scientific impact and aging of conference proceedings compared to those of scientific literature in general. It shows that the relative importance of proceedings is diminishing over time and currently represents only 1.7% of references made in the natural sciences and engineering, and 2.5% in the social sciences and humanities. Although the scientific impact of proceedings is losing ground to other types of scientific literature in nearly all fields, it has grown from 8% of the references in engineering papers in the early 1980s to its current 10%. Proceedings play a particularly important role in computer sciences, where they account for close to 20% of the references. This article also shows that not unexpectedly, proceedings age faster than cited scientific literature in general. The evidence thus shows that proceedings have a relatively limited scientific impact, on average representing only about 2% of total citations, that their relative importance is shrinking, and that they become obsolete faster than the scientific literature in general.

Introduction

The most frequently used source material in bibliometric studies consists of published scientific papers; however, it has been shown by several researchers that other types of literature are often published and cited by researchers. In the case of the social sciences and humanities (SSH) for example, Larivière, Archambault, Gingras, and Vignola-Gagné (2006) showed that papers published in serials represent only 50% of cited documents. Other researchers such as Hicks (1999) and Glänzel and Schoepflin (1999) noted the extent of coverage difficulties in Thomson Scientific's citation indexes, which do not include sources of scientific information such as grey literature, books, open-access preprints, patent applications, and conference proceedings. In line with Larivière et al. (2006), this article examines how important conference proceedings are in various fields of scientific inquiry by measuring the extent to which they are cited in other papers.

Though the traditional model of the growth of scientific literature considers conference proceedings as prototype papers that do not withstand the test of time, recent literature has shown that in some fields, conference proceedings can be considered as the final product of scientific research. Drott (1995), for instance, showed that the function of proceedings in the scholarly communication system is much more complex than what the *standard* evolutionary model has suggested, and that they are more than just "preliminary material that will later be turned into rigorous, finished works and formally published as journal articles" (p. 299). In fact, Drott suggested that there are three specific functions for proceedings. First, they can help researchers improve their papers by allowing the latter to gather feedback from other researchers before submitting them to a journal. Second, they can stimulate discussion within a field by, for example, allowing researchers to exchange ideas on emerging questions and paradigmatic positioning or to seek peer expertise. Third, they can be a vehicle for information that would otherwise be difficult to include in an article, such as application reports or theories that are either too broadly presented, difficult to confirm, or too short to be published in article form. More recently, Montesi and Mackenzie Owen (2008) discussed the specific roles of proceedings and journal articles in software engineering. They concluded that conference proceedings

Received February 12, 2008; revised April 10, 2008; accepted April 10, 2008

 DOI: 10.1002/asi.20888

can "measure the ability to innovate and propose new ideas, whereas journal publication can more strongly contribute to building a knowledge basis . . . " (p. 828).

In the field of bibliometrics, authors such as Butler and Visser (2006), Glänzel, Schlemmer, Schubert, and Thijs (2006), and Godin (1998), using Thomson's proceedings database, showed that taking proceedings into account in bibliometric studies produces a more complete and hence precise picture of a given discipline's scientific production. Moreover, again according to Glänzel et al. (2006), conference proceedings represent an important part of the published literature in engineering. Similarly, Goodrum, McCain, Lawrence, and Giles (2001), Visser and Moed (2005), and Butler (2008) argued that for computer sciences and other emerging disciplines, proceedings are attributed even greater importance than are articles in the transmission of knowledge.

The first part of this article analyzes the importance of proceedings in all fields of the natural sciences and engineering (NSE) and social sciences and humanities (SSH). More specifically, the relative importance of proceedings among all documents cited by articles indexed in Thomson Scientific databases is measured. This measure shows (a) the extent to which proceedings can be considered as final products of scientific research in a given field and, by extension, (b) the share of the relevant scientific production that is missing when proceedings are not taken into account in bibliometric studies. To evaluate this importance, this article starts from the working hypothesis that if proceedings are indeed important sources of scientific information, they will inevitably be frequently cited in core papers, particularly in areas where this type of document plays a central role. This is a departure from previous studies on proceedings, which postulated that their share of the number of publications in a given field was a measure of their importance. Even though this argument cannot be completely dismissed, there are grounds to believe that the best way to evaluate the scientific importance of proceedings within a field's production is to measure the extent to which they are cited. The second part of this article studies aging characteristics of cited proceedings. It aims at assessing whether proceedings are indeed vectors of recently acquired knowledge, in which case they would be more ephemeral than the literature published in serials, or if their *useful* life is as long as that of cited literature in general.

Methods

Thomson's citation indexes cover a significant part of the world's scientific journals. The objective of the current research project was to locate the largest possible number of proceedings papers cited in articles in these journals. Data were collected using the Thomson Scientific *Science Citation Index* (*SCI*), *Social Sciences Citation Index* (*SSCI*), and *Arts and Humanities Citation Index* (*AHCI*) CD-ROMs, covering the years 1980 through 2005. Only references made by articles, notes, and review articles were considered in the study.

TABLE 1. Keywords used for retrieving cited proceedings.

Actes	Exhibition	Symposium
Abstracts	Meeting	Workshop
Colloquy	Poster	P
Congress	Proceedings	M
Conference	Seminar	S
Convention	Session	C

The fields and subfields classification of journals is based on the scheme used in the Science and Engineering Indicators of the National Science Foundation.

The main methodological challenge of this study was to isolate among all cited material the references made to conference proceedings (refer to the Appendix for a diagram-form synthesis of the approach used). The challenge as always in this type of study is to maximize the signal-to-noise ratio, that is, to retrieve a maximum of proceedings (i.e., signal) while keeping false positives (i.e., noise) to a minimum. The first step of the process consisted of compiling a list of keywords (initially $n = 18$)—along with several truncations and linguistic variations in Spanish, Italian, and German—that represent various ways in which conference proceedings are recorded (Table 1). Data retrieved by each keyword were filtered and stored in their own distinct database. Thomson uses the letters *P*, *S*, *C*, and *M* to designate *proceedings*, *symposiums*, *conferences*, and *meetings*, respectively; however, the use of these letters is not reserved exclusively for this practice; they also can designate *physics*, *supplement*, *control*, or *mechanical*, for example. To filter noise generated by these four letters, only references containing one of the four letters and particles such as *int*, *nat*, *ann*, *day*, *bienn*, *world*, and *joint*, abbreviations in names of months and seasons, and 1 to 4 number numerations and ordinations were retained. The quality of remaining filtered data was further improved by the elimination of references containing the particles *J*, *U*, *acad*, *roy*, *philo-soc*, *report*, and *thesis*. For certain keywords, references with the *I* particle also were considered as noise.

References with the *P* particle to designate *proceedings* constituted the most complex case because of the large number of references that contained the *P* particle without referring to a set of proceedings. As these references alone represented 75% of all retrieved data, further filtering was absolutely essential. Given that using a *P*-* query would have generated an excessive amount of false positives—such as journal names that included the term *proceedings* (e.g., *Proceedings of the National Academy of Sciences*, *Proceedings of the American Mathematical Society*, etc.), all references containing the particles *J*, *I*, *U*, *acad*, *roy*, *philo-soc*, *report*, and *thesis* were filtered out. Particles such as *series*, on the other hand, confirmed that these references were proceedings. Search tools Google and the WorldCat catalog were used to validate the references. For example, *Proceedings of the Institution of Mechanical Engineers. Part D*, *Journal of Automobile Engineering* indexed in WorldCat under the subject heading *Automobiles–Design and construction–Periodicals*

DOI: 10.1002/asi

was rejected whereas *Proceedings of the Combustion Institute* indexed under *Combustion–Congresses* was retained.

Despite all of these efforts, the presence of false negatives and false positives was unavoidable. For instance, some proceedings share a heading that is similar to other monographic publications and are thus impossible to recuperate. To validate our final query, a random sample of 1,000 retrieved proceedings was manually validated. For the SSH, 8.9% were false positives; 4.4% were false positives in the NSE. On the other hand, a sample of 1,000 nonproceeding references also was manually validated with a percentage of false negatives of less than 1% for both the NSE and the SSH. Overall, these very low percentages indicate that the collected data constitute a valid representation of cited proceedings in Thomson's scientific document sources. Even though the study's actual numbers may be imprecise in terms of absolute value, its relative values should reflect the same proportions that would appear if one had been able to retain all cited proceedings, since all scientific fields and years were equally subjected to the same research criteria.

Results

Number and Percentage of Proceedings Cited

Figures 1 and 2 show the annual number of proceedings cited along with their share of all cited literature for both the NSE and the SSH. Of all the documents cited in Thomson Scientific's CD-ROM databases in 2005, approximately 1.7% consists of proceedings in the NSE and 2.5% in the SSH. Unsurprisingly, the absolute number of cited proceedings has increased in the course of the last quarter of a century; however, given the increase in the number of references per paper over the period (Larivière, Archambault, & Gingras, 2008), their share among all cited documents has decreased in both the NSE and the SSH. More specifically, the drop of

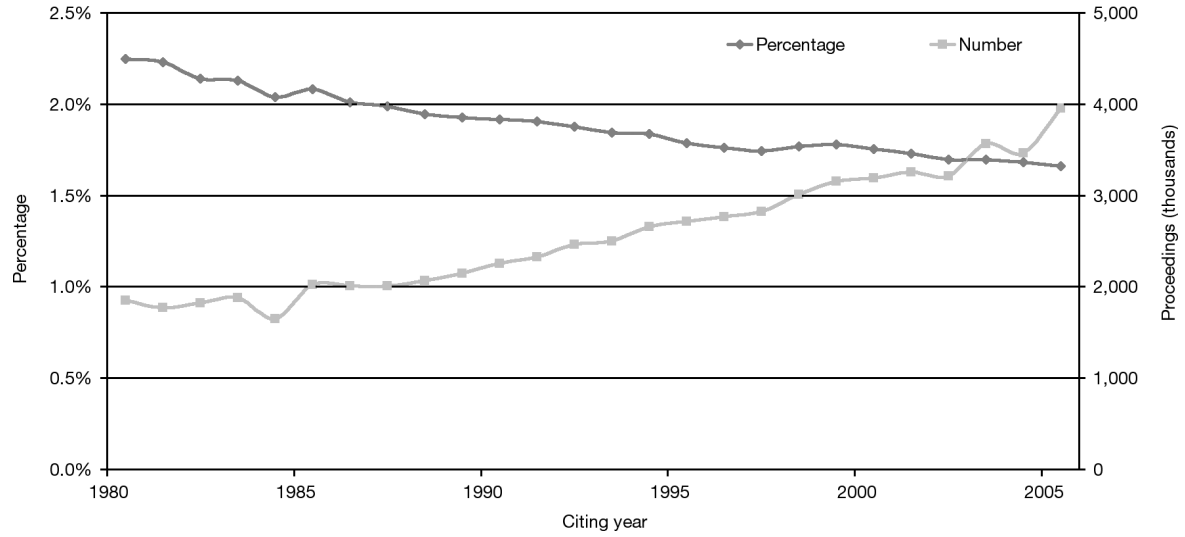


FIG. 1. Number and percentage of proceedings cited in the NSE, 1980–2005.

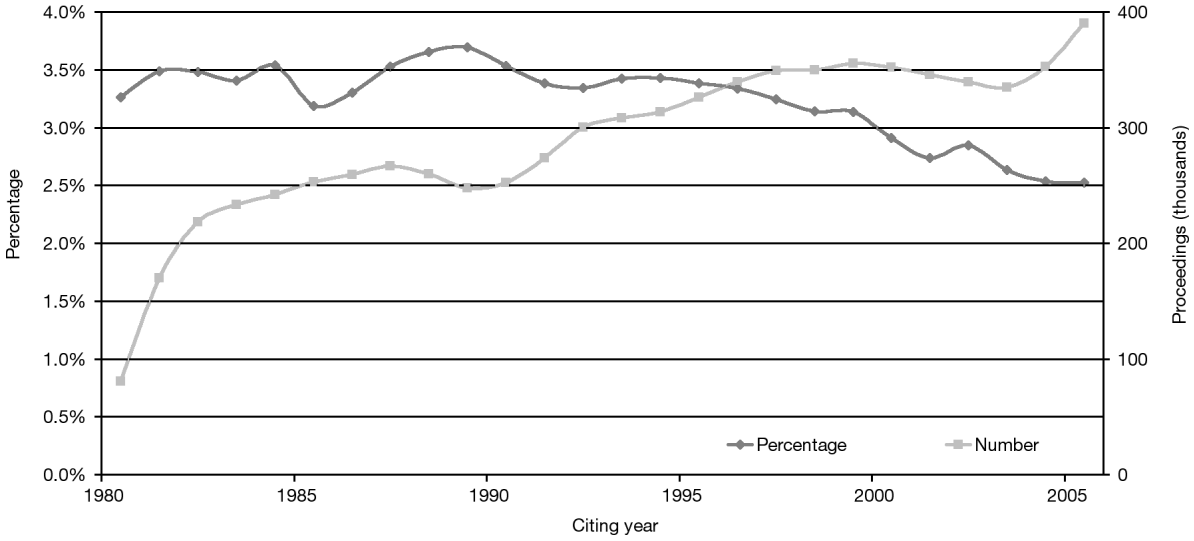


FIG. 2. Number and percentage of proceedings cited in the SSH, 1980–2005.

DOI: 10.1002/asi

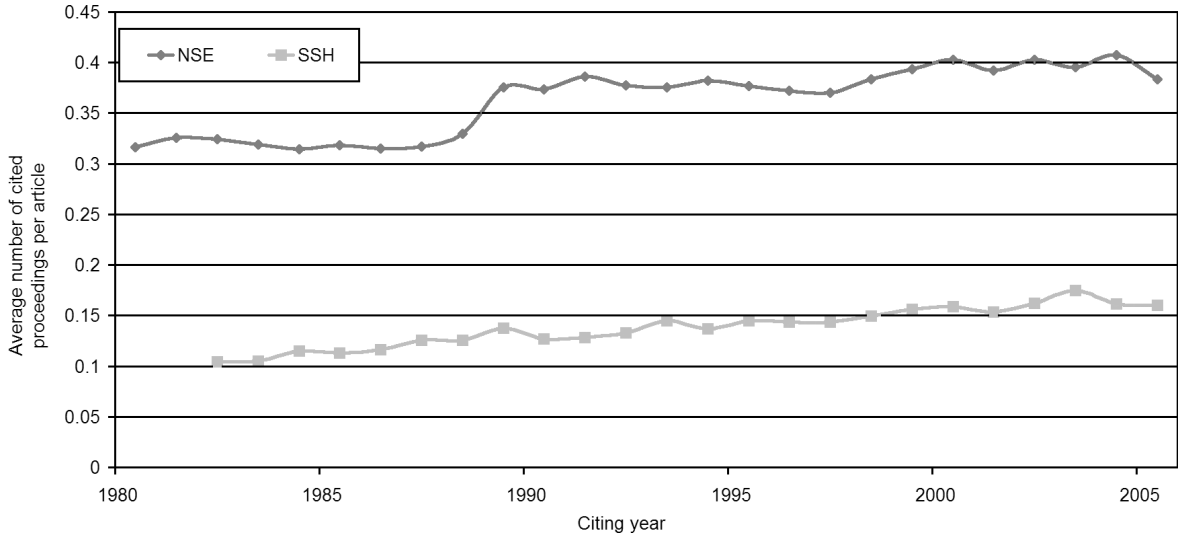


FIG. 3. Average number of proceedings cited per paper in both the NSE and the SSH, 1980–2005.

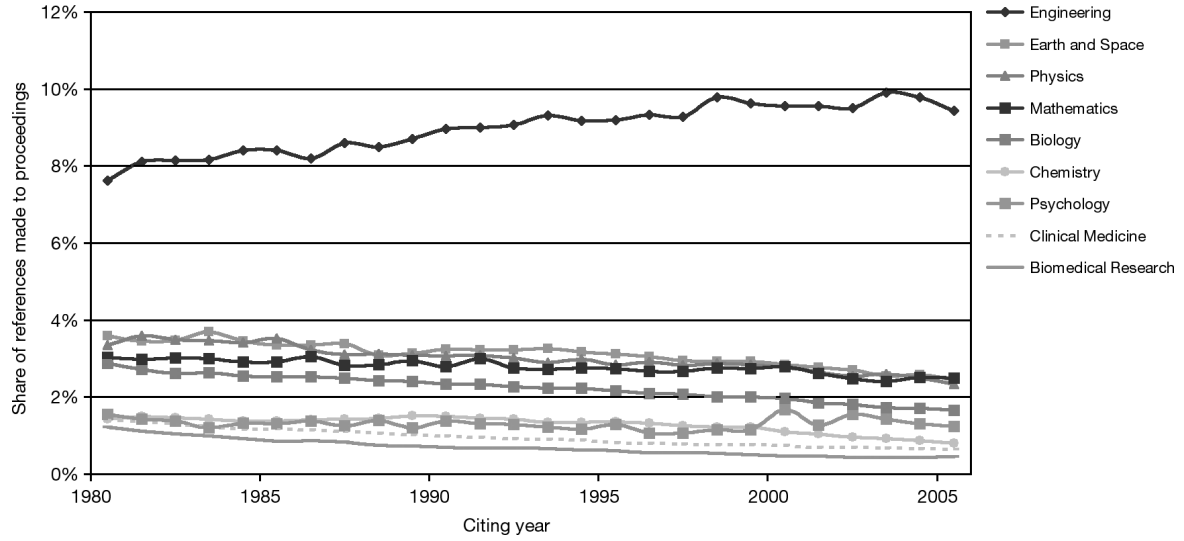


FIG. 4. Percentage of proceedings cited in the NSE, by field, 1980–2005.

approximately 0.5 percentage points amounts to a decline of one fourth of the presence of proceedings in the references in NSE papers. Similarly, the drop of about 0.75 percentage point amounts to a decline of about one fifth of the presence of conference proceedings in references in SSH papers.

Figure 3 shows that the average number of proceedings cited per article has increased slightly in these 26 years, growing from 0.32 per paper in 1980 to 0.38 in 2005 in the NSE and from 0.10 to 0.16 in the SSH. However, given that the total number of references also increased during the same period, the relative importance of proceedings among all cited material has been decreasing steadily, as seen in Figures 1 and 2.

Figures 4 and 5 present the evolution of the percentage of cited proceedings by field over the 1980–2005 period. The evolution shows that engineering is the only field in which proceedings account for an increasingly large share of cited references (Figure 4). Indeed, this share increased from more than 7% to almost 10% over the period. All other NSE and SSH fields show a decline in the share of cited proceedings. In 2005, the share of proceedings was below 3% in all NSE fields except engineering and below 1.5% in all SSH fields.

The share of proceedings among all cited literature varies considerably between subfields. Of 109 NSE subfields, 5 subfields had a share of proceedings of more than 10% of citations; 11 subfields had a share between 5 and 10%; 57 subfields had a share of citations to proceedings between 1 and 5%, and 36 subfields cited less than 1% of proceedings. Of 77 SSH subfields, 1 subfield cited more than 5% of proceedings; 48 subfields cited proceedings between 1 and 5% of the time, and 28 subfields cited less than 1% of proceedings.

Figures 6 and 7 present the main subfields in which proceedings account for a substantial share of the cited literature.

DOI: 10.1002/asi

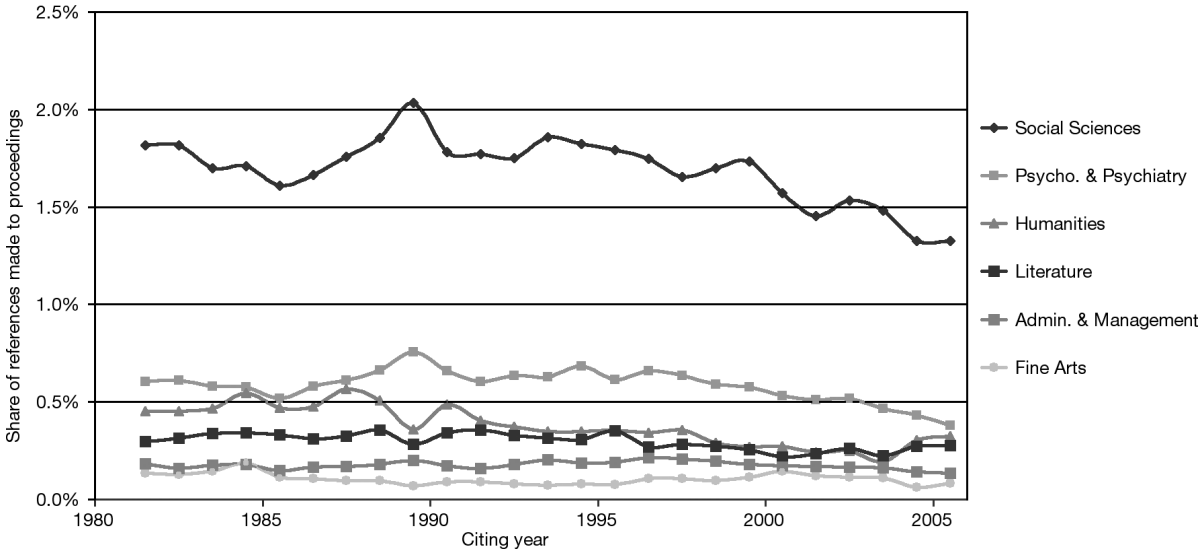


FIG. 5. Percentage of proceedings cited in the SSH, by field, 1981–2005.

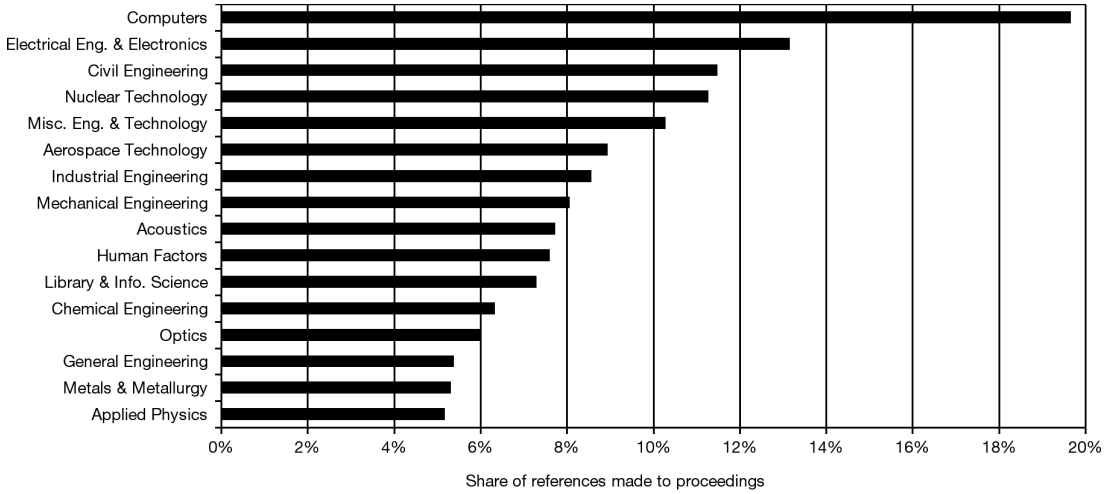


FIG. 6. Natural sciences and engineering subfields with the highest percentage of references made to proceedings, 1980–2005.

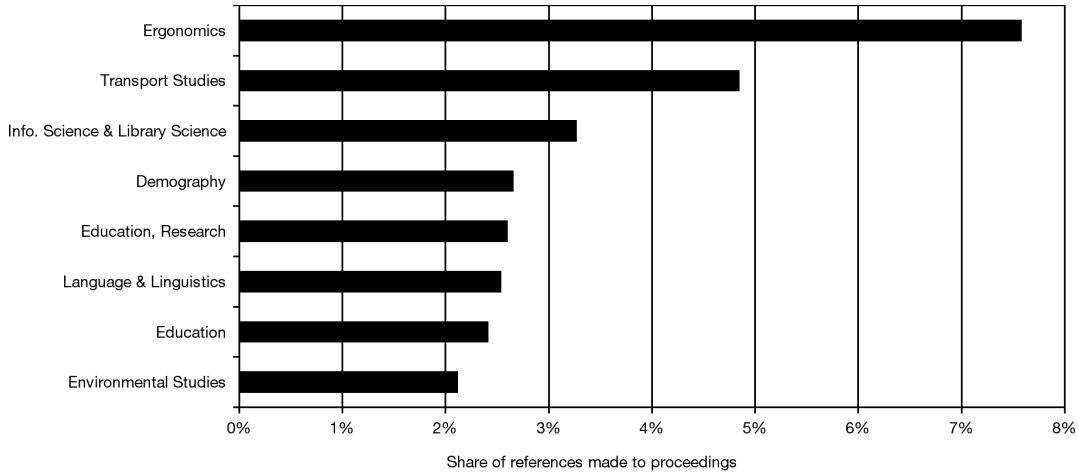


FIG. 7. Social sciences and humanities subfields with the highest percentage of references made to proceedings, 1980–2005.

DOI: 10.1002/asi

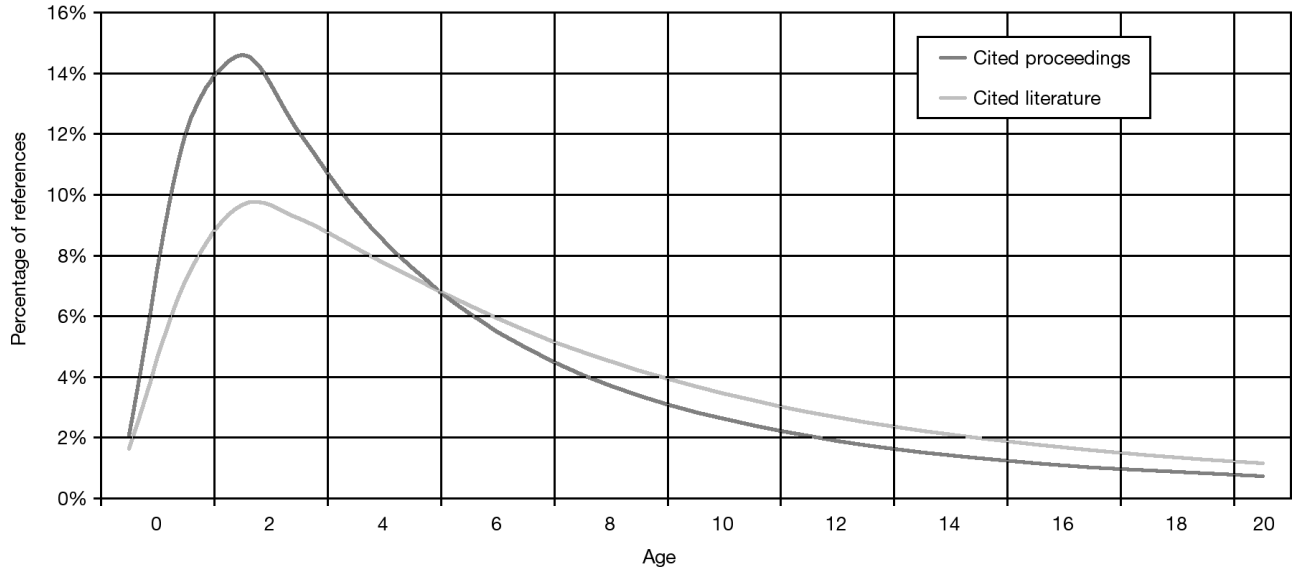


FIG. 8. Distribution of the age of cited proceedings and of cited literature, all NSE and SSH fields, 1980–2005.

As one might expect and as previous literature on the topic has suggested, most of these subfields are related to the field of engineering. In NSE (Figure 6), these subfields are computers (19.6%), followed by other engineering subfields such as electrical engineering & electronics (13.1%), civil engineering (11.5%), nuclear technology (11.2%), miscellaneous engineering & technology (10.3%), aerospace technology (8.9%), industrial engineering (8.5%), and mechanical engineering (8%). Other subfields in which proceedings account for more than 5% of the references—unrelated to engineering—are acoustics (7.7%), psychology–human factors (7.6%), library & information science (7.3%), and optics (6%).

In SSH (Figure 7), the only subfield with a share of proceedings among all referenced material that is above 5% is ergonomics (7.6%). Other subfields that cited more than 2% of proceedings include transport studies (4.8%), information science and library science[1] (3.3%), demography (2.6%), education–research (2.6%), language and linguistics (2.5%), education (2.4%), and environmental studies (2.1%).

Average Age of Cited Proceedings

Figure 8 provides aging curves of cited proceedings and cited literature generally that show that proceedings are cited more rapidly than literature as a whole. These data show that proceedings' peak life cycle is shorter than that of literature in general and that they obsolesce faster. For instance, their half-life (i.e., median age) is 4.0 while that of literature in general is 6.1. Hence, it seems that proceedings may provide researchers with—possibly preliminary—results at the forefront of scientific research, but they certainly obsolesce faster.

Figure 9 compares the average age of cited proceedings with that of the whole cited literature in the NSE and the SSH. One can readily see that—at least in the case of NSE—the age of both cited proceedings and cited literature is steadily increasing, a trend similar to that observed by Larivière et al. (2008). Interestingly, proceedings do not behave differently than the cited literature in general; researchers rely on an increasingly old body of conference proceedings. In SSH, the tendency is not as clear, although it does seem that researchers have been relying to a greater extent on older proceedings and literature in general since the mid-1990s.

For both the NSE and the SSH, cited proceedings are clearly younger than the cited literature in general, which confirms the pattern observed in the citation curves presented in Figure 8. However, in NSE, the difference between the two curves has been diminishing steadily over time. Indeed, while the average age of cited proceedings was 6.3 and that of cited literature in general was 9.3 in 1980, those values were 8.4 and 10.1, respectively, in 2005. In the SSH, there is an even larger difference between the two values: the age of cited proceedings in 2005 was 10.3 while that of cited literature as a whole was 14.2. Also note that in contrast to the large difference in the age of cited literature generally in the SSH compared to that in the NSE, the difference in age of the cited proceedings is not all that great between the SSH and the NSE.

Figures 10 (NSE) and 11 (SSH) illustrate how the age of proceedings compares with that of all cited scientific literature in broad fields such as physics and chemistry. These data confirm the pattern noted earlier, whereby the age of cited proceedings is younger than that of cited literature in general.

[1]The differences between the percentage of proceedings cited in library & information science in the *SCI* and in information science and library science in *SSCI*/*AHCI* are caused by the fact that different journals are included in each of the databases. *SCI*'s journals are more oriented toward quantitative research and information sciences while those in *SSCI*/*AHCI* encompass all aspects of LIS.

DOI: 10.1002/asi

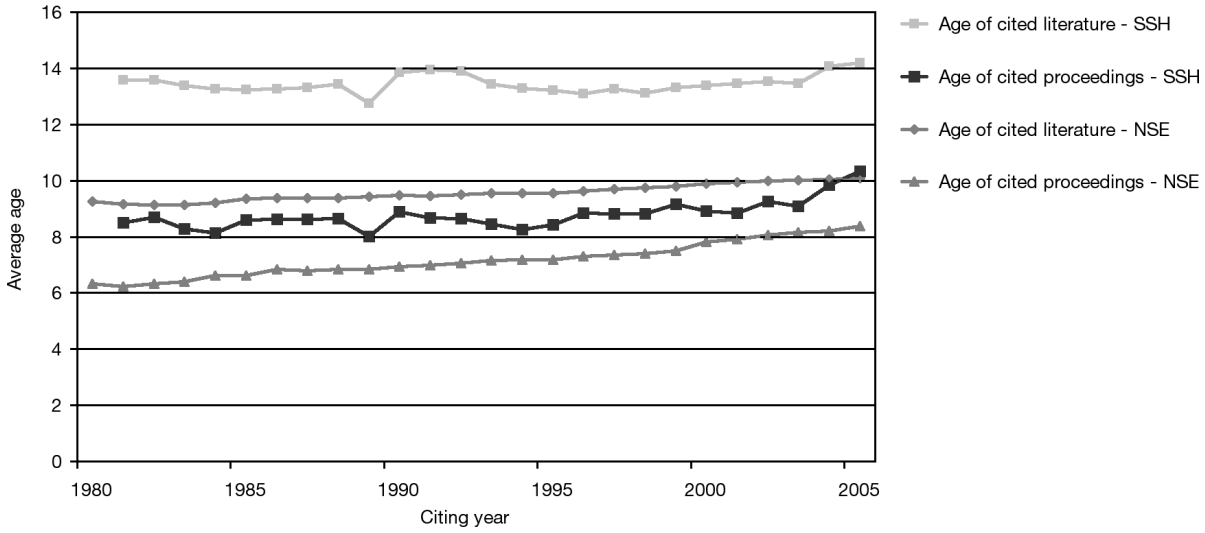


FIG. 9. Average age of cited literature and average age of cited proceedings, 1980–2005.

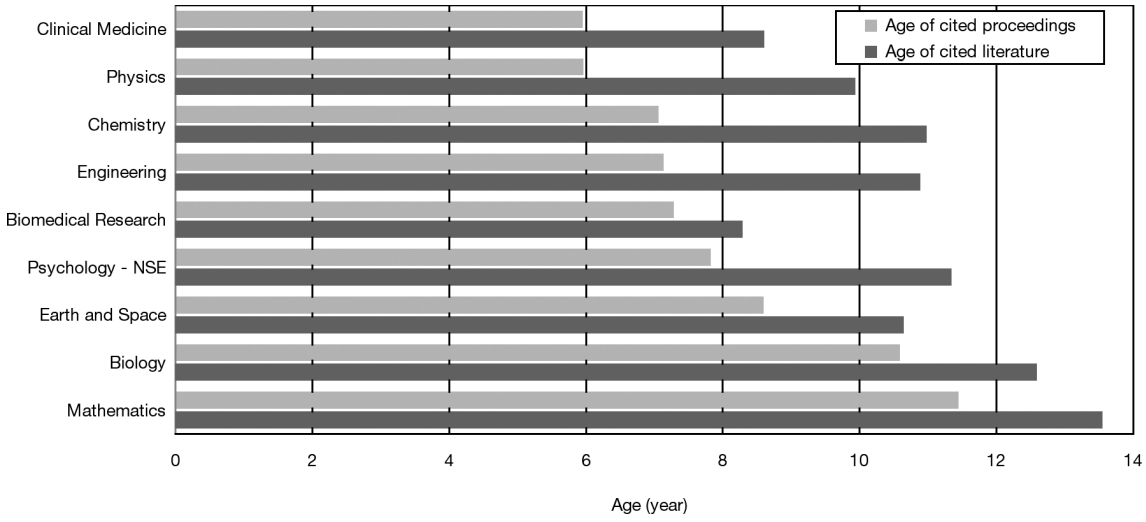


FIG. 10. Average age of cited literature and average age of cited proceedings, by NSE field, 1980–2005.

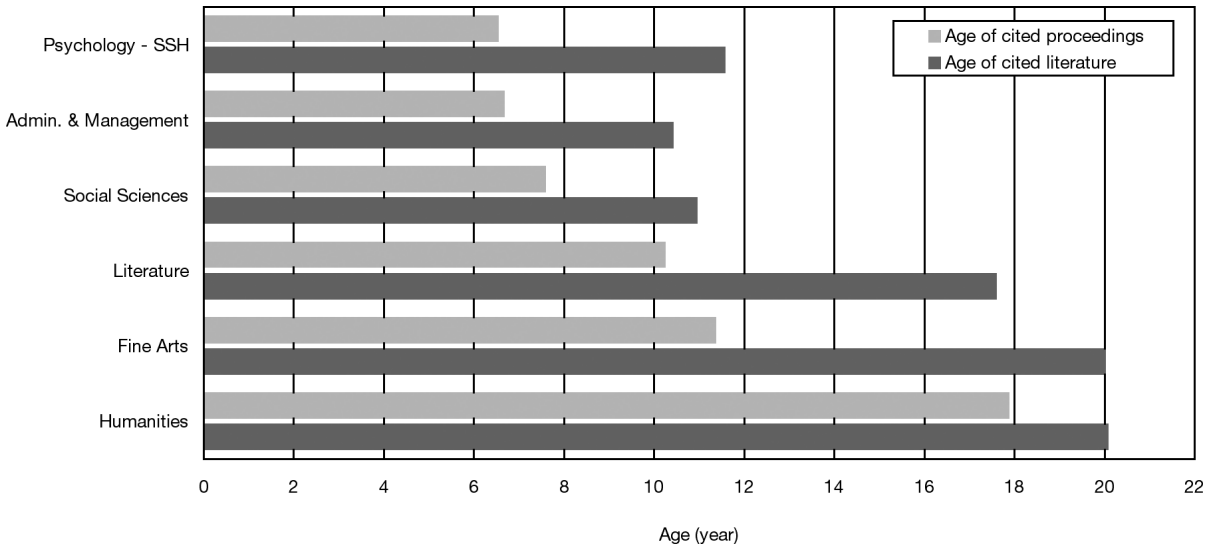


FIG. 11. Average age of cited literature and average age of cited proceedings by the SSH field, 1980–2005.

DOI: 10.1002/asi

While in some NSE fields such as physics (−40%), chemistry (−36%), engineering (−34%), clinical medicine (−31%), and psychology (−31%), this difference is quite large, in others such as earth and space sciences (−19%), mathematics (−16%), biology (−16%), and biomedical research (−12%), their aging characteristics are similar to those of cited literature in general.

In all SSH fields but the humanities (Figure 11), one can observe a considerable difference between the aging process of proceedings and that of the literature in general. Indeed, cited conference proceedings are 43% younger in fine arts and psychology, 42% in literature, 36% in administration and management, and 31% in social sciences, while they are only 11% younger in humanities. This suggests that conference proceedings serve different functions and have different life cycles depending on the community they serve. In some fields, they can expect to have a useful life which is as long as that of literature in general; in some others, they appear to have a more "traditional" function and obsolesce faster. At the field level, there is no correlation between the extent of their use—as measured by their share of all referenced material—and their aging characteristics.

Discussion and Conclusion

The small percentage of references made to proceedings among all cited documents—about 2% in the NSE and the SSH combined—indicates that conference proceedings do not account for a significant share of the relevant scientific literature in general. Their share among all cited material is decreasing and while the number of references per paper increased steadily over the period, the number of proceedings cited per paper increased only very slightly. However, proceedings account for a relatively large share of the references made in engineering, with a "conservative" percentage of all cited material that increased from 7.6% to almost 10% over the last 25 years. Within the specialties in engineering, five subfields have a share of cited proceedings that is above 10% of total citations: computers (19.6%), electrical engineering & electronics (13.1%), civil engineering (11.5%), nuclear technology (11.2%), and miscellaneous engineering & technology (10.3%).

These findings are consistent with those of Butler and Visser (2006), Glänzel et al. (2006), and Godin (1998), who suggested that conference proceedings account for a significant part of published literature in engineering, as well as with those of Moed (2005), who revealed that Thomson's coverage of this discipline is good, but not excellent. These findings also support those of Goodrum et al. (2001), Visser and Moed (2005), and Montesi and Mackenzie Owen (2008), who argued that conference proceedings are important in computer sciences. All in all, the data presented in this article provide evidence that in addition to being published, conference proceedings are indeed cited by researchers in those fields. Though proceedings account for a very small share of relevant literature in science as a whole, they do account for a nonnegligible part of cited literature in engineering and computer sciences and hence should, in addition to scientific articles, be considered for inclusion in bibliometric studies and evaluations.

Importantly, the data in this article contrast markedly with those of Butler (2008) as to the importance of proceedings in the construction of scientific knowledge. Whereas Butler found that about 63% of publications in computer sciences by Australian universities consisted of conference proceedings, overall, only 20% of the references in this field are made to proceedings. If one uses theses figures as ballpark estimates, it shows that the 60% of the output translates into a mere 20% of citations, suggesting that computer scientists wishing to maximize their *scientific* impact may prefer using other media than proceedings. Thus, although proceedings may be extremely important as diffusion media, their scientific impact does not seem to be all that important. This does not mean that proceedings are not important overall, as they may be a better way of reaching practitioners in the field who are more inclined to transfer the knowledge they learn through proceedings into technology. As such, it would be interesting to study the relative frequency of references to proceedings, and also to scientific papers and books, in patents.

The findings of this article also indicate that in all fields, cited proceedings are younger than cited literature in general, suggesting that proceedings are a medium of more recent knowledge than are all types of literature in general. Along the same lines, citations received by conference proceedings decline faster than do those received by scientific literature in general. An obvious explanation for this faster rate of obsolescence is the fact that in some fields—such as software engineering (Montesi & Mackenzie Owen, 2008)—proceedings are transformed into published articles, whereas in others they are the final form for the diffusion of scientific knowledge. Though this practice varies between fields (Drott, 1995), this is certainly a partial explanation. That being said, the present article's data also suggest that the extent to which conference proceedings are later converted into scientific articles in a given field is independent of the percentage of references that are made to conference proceedings. Even in fields such as engineering in which proceedings account for a large share of referenced material, proceedings have a shorter life—and half-life—than does cited literature in general. In fact, it seems that even in engineering, proceedings also serve to provide access to new, more recent literature at the forefront of scientific research.

The extent to which conference proceedings are cited in engineering and computer sciences strongly suggests that scientists in these fields consider these documents as more than just prototypes, but rather as final products of scientific research. The transfer rate of proceedings into scientific articles is also likely to be lower in these fields. This is in line with a function of proceedings proposed by Goodrum et al. (2001), which is a substitution of articles with proceedings. However, the fact that the age difference between cited proceedings and cited scientific literature is similar in both engineering and science as a whole suggests that the function of proceedings in engineering is to not only replace articles

DOI: 10.1002/asi

but also provide, as in other fields, access to more recent discoveries.

Acknowledgments

We thank Claude Taillefer for translating the manuscript into English, Johanna Kratz for its revision, François Vallières for constructing the bibliometric database, and Yves Gingras and Jean-Pierre Robitaille for providing comments on an earlier version of the article.

References

Butler, L. (2008). ICT assessment: Moving beyond journal outputs. Scientometrics, 74(1), 39–55.

Butler, L., & Visser, M.S. (2006). Extending citation analysis to non-source items. Scientometrics, 66(2), 327–343.

Drott, C.M. (1995). Reexamining the role of conference papers in scholarly communication. Journal of the American Society for Information Science, 46(4), 299–305.

Glänzel, W., Schlemmer, B., Schubert, A., & Thijs, B. (2006). Proceedings literature as additional data source for bibliometric analysis. Scientometrics, 68(3), 457–473.

Glänzel, W., & Schoepflin, U. (1999). A bibliometric study of reference literature in the sciences and social sciences. Information Processing and Management, 35(1), 31–44.

Godin, B. (1998). Measuring knowledge flows between countries: The use of scientific meeting data. Scientometrics, 42(3), 313–323.

Goodrum, A., McCain, K., Lawrence, S., & Giles, C.L. (2001). Scholarly publication in the Internet age: A citation analysis of computer science literature. Information Processing and Management, 57(2), 661–675.

Hicks, D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. Scientometrics, 44(2), 193–215.

Larivière, V., Archambault, É., & Gingras, Y. (2008). Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004). Journal of the American Society for Information Science and Technology, 59(2), 288–296.

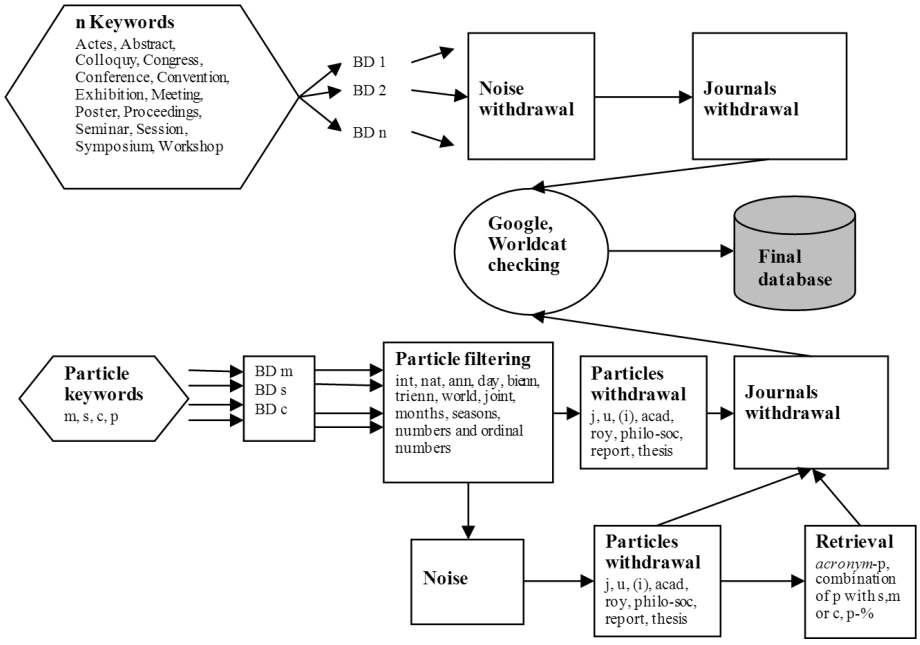
Larivière, V., Archambault, É., Gingras, Y., & Vignola-Gagné, E. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. Journal of the American Society for Information Science and Technology, 57(8), 997–1004.

Moed, H.F. (2005). Citation analysis in research evaluation. Dordrecht, The Netherlands: Springer.

Montesi, M., & Mackenzie Owen, J. (2008). From conference to journal publication: How conference papers in software engineering are extended for publication in journals. Journal of the American Society for Information Science and Technology, 59(5), 816–829.

Visser, M.S., & Moed, H.F. (2005). Developing bibliometric indicators of research performance in computer science. In P. Ingwersen & B. Larsen (Eds.), Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics (pp. 275–279). Stockholm: Karolinska University Press.

Appendix



Schematic representation of the retrieval process

DOI: 10.1002/asi