Working Paper Series
ISSN 1170-487X

Obsolescence of computing literature.

by Sally-Jo Cunningham, David Bocock

Working Paper 95/8
March 1995

Department of Computer Science
The University of Waikato
Private Bag 3105
Hamilton, New Zealand

Obsolescence of computing literature

Sally Jo Cunningham, David Bocock
Department of Computer Science
University of Waikato
Hamilton
New Zealand
email: sallyjo@waikato.ac.nz

Abstract: A multisynchronous obsolescence study has been performed on two computing journals that publish on technical aspects of computer system management (networks and operating systems). This area of computer science is found to have a relatively high obsolescence rate (a median citation rate of four years). This rate is similar to that of fields in engineering and the technology-dependent "hard" sciences.

1. Introduction

How long does the information in a research article remain current? How old can a document be before it is likely to be judged out of date? In the field of bibliometrics, obsolescence studies attempt to answer these questions by measuring the amount of "usage" (citations) that a document receives after it is published. The quantitative measurements most often describing usage are the median citation age and publication half life for a given field (described in Section 2). Typically, documents receive their greatest number of citations immediately after publication, and the frequency of citation falls rapidly as time passes. If, that is, the document is referenced at all—it has been estimated that approximately one-third of all scientific articles are never cited [Computer Horizons, 1976].

The term "obsolescence" can be misleading, since it implies that the document's information has been discovered to be incorrect or no longer applicable. However, another reason that a document may cease to be cited is that its contents have been so thoroughly incorporated into a field that the document's findings are held to be common knowledge! Further, the characteristics of the users may be as important as the characteristics of a document in determining a document's rate of aging, when only reference dates rather than reference intentions are analyzed [Griffith, 1979]. For example, an active researcher in a field will likely require up to the minute materials; a textbook author will range back through earlier work to present more fundamental and generally accepted results to students; and a historian of science may reference texts that are centuries old and whose scientific findings have long been discredited. A simple analysis of the dates upon which these references occur will not distinguish between the different types of usage that the document sees.

Despite this blurring of reference intentions, quantitative rather than qualitative citation analysis is the preferred method for determining obsolescence, since quantitative studies are more quickly performed and can be automated to a greater extent. These quantitative models of obsolescence, while approximate, are of interest and importance in understanding the nature of scientific information and the dissemination of knowledge in a given field.

Studies of obsolescence have concentrated primarily on the physical and natural sciences, and to a lesser extent on the social sciences and humanities. The computer science, information systems, and information technology literatures have scarcely been examined by bibliometricians and scientometricians [representative studies include Evaristo, 1993; Prechelt, 1994; and Hall, 1985]. To our knowledge, no obsolescence

studies have been conducted in these subjects. This study examines one aspect of the computer science literature—specifically, technical issues of computer system management (networks and operating systems)—as a preliminary attempt to quantify the obsolescence rate of that subject.

This paper is organized as follows: Section 2 describes our research methodology in measuring obsolescence for the chosen sample literature; Section 3 discusses our analysis of the data; and Section 4 presents our conclusions.

2. Research Methodology

Two types of measures are in common use: synchronous obsolescence and diachronous obsolescence. For the latter metric, a sample of documents is selected from a point in the past, and the rate of citations to these documents are calculated on a year by year basis (generally through use of a citation index). Typically a "half life" is calculated as the time period in which half of all observed citations are received by the document set. A diachronous study answers the question, "how long will a publication continue to be used after it has been published?" [Wallace, 1986].

In contrast, a synchronous obsolescence rate is determined by proceeding from the present to the past: the publication dates of documents *referenced* by a sample set of documents are examined to determine how far back in time the more recent half of those references were published. This measure is called the median citation age, and answers the question, "what is the age of a publication likely to be at the time it is cited by another publication?" [Wallace, 1986].

Obviously the synchronous obsolescence rate is easier to calculate; it requires access to the sample documents alone, without recourse to a citation index. Diachronous studies are generally more resource-intensive and time-consuming to conduct. Whether synchronous and diachronous obsolescence are statistically equivalent is debatable [Stinson, 1987; Gupta, 1990]. They are largely complementary, since secondary calculations for the two measures give different insights into the subject literature.

Information systems/information technology literature has a notoriously wide "spread" through computing, mathematics, management, science, and the social sciences, making it particularly difficult to trace citations through the relevant indices. For that reason, a synchronous methodology was chosen. The specific body of literature used for the present study is the 1987-1993 volumes of *Computer Networks and ISDN Systems* and *Operating Systems Review*. This study is, moreover, "multisynchronous" in the sense that the journals selected for analysis are drawn from seven consecutive years. This approach attempts to eliminate the bias that may occur when a single base year is chosen.

Technique for calculating median citation age

The median citation age for an article is calculated using the discrete analysis method: publication years are treated as discrete units, rather than as a continuum of quarters, months, and weeks. To determine the median citation age for a single article, the publication dates of its references are first listed in reverse chronological order. For example, the reference dates for this paper are:

1994, 1993, 1993, 1988, 1987, 1986, 1985, 1982, 1979, 1976, 1970

The median year is 1986 (the sixth of eleven dates). Subtracting 1986 from the year of publication yields a median citation age of none years (1995 - 1986). To determine the median citation age for a given subject, this procedure is applied to a sampling of articles from the field.

The selection of a body of literature and range of years as a focus for a bibliographic study is often dependent on pragmatic as well as theoretic factors—availability of documents, ease of access, etc. While these reasons played a part in the choice of subject for this study, the journals selected for analysis are also appropriate because of their relatively narrow focus.

Data collection

This study analyzed a total of 9,741 references: 5,066 from Computer networks and ISDN Systems (volumes 13-26), and 4,675 from Operating Systems Review (volumes 21-27). The volumes were published in the years 1987-1993.

3. Results

Citation ages for networks/operating systems literature varied from 0-23 years, with a median citation age of 4 years. The relatively narrow range of ages is to be expected, given the field's basis in computer hardware (a subject heavily dependent on recent advances, and in which older algorithms and technology quickly become irrelevant). Furthermore, the median citation ages for both journals are reasonably stable over the seven years analyzed by this study (Figure 1). Note that while the number of articles in each volume (and by extension, the number of references in each volume) has increased, the median citation age has not changed significantly over the time period studied.

Year	Computer Networks and ISDN Systems		Operating Systems Review		Both Journals
	number of references	median citation age	number of references	median citation age	number of references
1987	451	4	368	3	819
1988	476	4	551	4	1027
1989	546	3	658	4	1204
1990	635	5	512	3	1147
1991	745	3	667	3	1412
1992	1153	4	898	3	2051
1993	1060	3	1021	4	2081
Total Med. Cit. Age	5066	4	4675	3	9741 4

Figure 1: median citation ages by year and by journal

As might be expected, the median citation age for *Computer Networks and ISDN Systems*—a hardware-oriented journal—tends to be lower than the median for *Operating Systems Review* (a software-oriented publication). Hardware platforms tend to change more rapidly than software, and software must necessarily be constructed for existing hardware.

To place the median citation age in context, consider the following table:[1]

Discipline	median citation age (years)
metallurgical engineering	3.9
genetics	4.0 - 6.0
physics	4.4-4.9
chemical engineering	4.8
information systems	5
mechanical engineering	5.2
desalination	5.6
chemistry	8.1
archaeology	9.54
botany	10
music education	10.0 - 15.0
music theory	10.0 - 15.0
mathematics	10.5
geology	11.8
biblical criticism	21.63

Figure 2: synchronous median citation ages from other fields

As we might expect, the median citation age for this area of computer science is closest to values reported for engineering fields and other "hard" sciences. This result provides evidence for the oft-stated (but rarely empirically supported) belief that computer science is a rapidly changing field.

A graph of the ages of articles referenced by the journals in this study is presented in Figure 3. The graph shows an exponential curve commonly found in obsolescence studies, and includes the final climb described by Price as an "immediacy effect" [Price, 1970]. The networks/operating systems literature exhibits typical obsolescence patterns, without any obvious peculiarities.

4. Conclusions

This study examines a commonly held belief about computer science—that it is a rapidly changing field with a relatively high obsolescence rate for its documents. This hypothesis is confirmed for the field of computer operating systems and network management, in that the measured median citation rate is closer to that of engineering and the "hard" sciences than to the humanities and social sciences. Of course, computer science is a broad subject, and it is likely that the sub-fields of computing will have different obsolescence rates reflecting their "closeness" to other fields. This type of scatter has been measured in other disciplines [for example, physics; see Gupta, 1990]. Analysis of other areas of computer science remains a topic for further research.

---

[1] All studies are synchronous. With the exception of the information systems study, these median citation ages are taken from Diodato (1992). The information systems results are derived from an unpublished study, available from the authors.

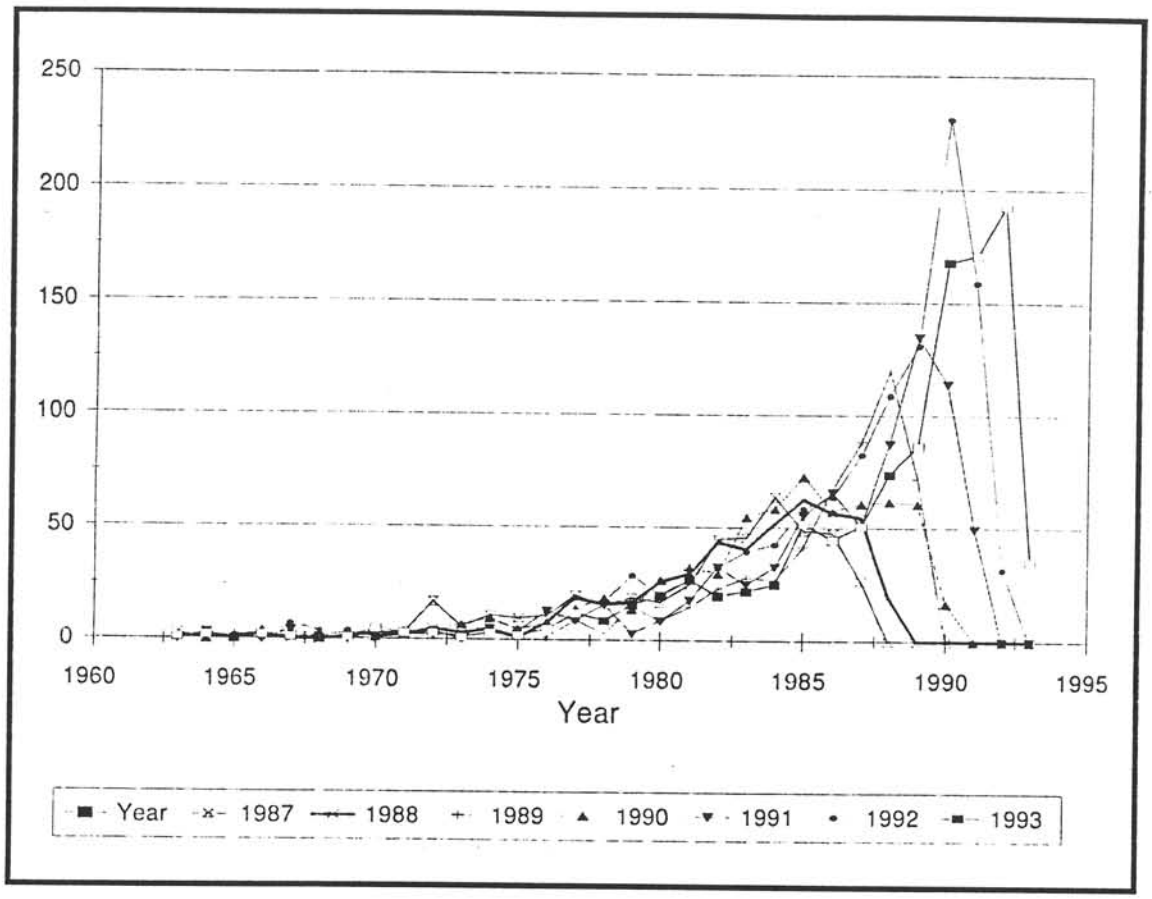

Figure 3. Distribution of reference dates

References

Computer Horizons, Inc. *Evaluative Bibliometrics: The Use of publication and citation analysis in the evaluation of scientific activity*. Reproduced by the US Dept. of Commerce National Technical Information Service, PB-252 339, 1976.

Diodato, V., and Smith, F. "Obsolescence of Music Literature." *Journal of the American Society for Information Science 44(2)*, 1993, pp. 101-112.

Evaristo, J.R., and Karahanna-Evaristo, E. "An empirical investigation of differences between North American and European MIS research." *Proceedings of the 14th International Conference on Information Systems* (Orlando, FL, USA), 1993, pp. 387-388.

Griffith, B.C., Servi, P.N., Anker, A.L., and Drott, M.C. "The aging of scientific literature: A citation analysis." *Journal of Documentation 35(3)*, 1979, pp. 179-196.

Gupta, U. "Obsoloscence of physics literature: Exponential decrease of the density of citations to *Physical Review* articles with age." *Journal of the American Society for Information Science 41(4)*, 1988, pp. 282-287.

Hall, R.P., and Kibler, D.F. "Differing methodological perspectives in artificial intelligence research." *AI Magazine 6*, 1985, pp. 166-167.

Khawam, J.J. "The AI interdisciplinary context: Single or multiple research bases?" *Library and Information Science Research 14,* 1982, pp. 57-74.

Prechelt, L. "A study of experimental evaluations of neural network learning algorithms: Current research practice." *Technical Report 19/94*, Fakultat fur Informatik, Universitat Karlsruhe, 76137 Karlsruhe, Germany, 1994. Anonymous ftp: /pub/papers/techreports/1994/1994-19.ps.Z on ftp.ira.uka.de.

Price, D. J. de Solla. "Citation measures of hard science, soft science, technology, and nonscience." Nelson, CE., and Pollock, D.K., eds. *Communication Among Scientists and Engineers*. Lexington, MA: Heath Lexington, 1970.

Stinson, E.R., and Lancaster, F.W. "Synchronous and diachronous methods in the measurement of obsolescence by citation studies." *Journal of information Science 13,* 1987, pp. 65-74.

Wallace, D. "The relationship between journal productivity and obsolescence." *Journal of the American Society for Information Science 37(3)*, 1986, pp. 136-145.