Information Processing & Management Vol. 23, No. 1, pp. 1-6, 1987
Printed in Great Britain.

0306-4573/87 $3.00 + .00

CONCEPTUAL CITATION DIFFERENCES IN SCIENCE, TECHNOLOGY, AND SOCIAL SCIENCES LITERATURE

C. D. HURT
Director, Graduate Library School, University of Arizona,
1515 E. First Street, Tucson, AZ 85719

(*Received* 17 *June* 86; *in revised form* 2 *October* 1986)

Abstract—This study examined three disciplines and the literature to determine if there were any differences in conceptual referencing patterns. Physics, engineering, and sociology were examined using the actual literature of the disciplines for the year 1984. Samples were drawn from major abstracting and indexing services. A Dunn planned comparison approach was used to test for differences. The results of the testing indicated that physics, engineering, and sociology all differ in terms of conceptual references. Two sub-tests were performed on the data. Total number of references were examined, indicating no difference between physics and sociology, but distinct differences between physics and engineering and between sociology and engineering. The second sub-test examined the number of references to literature within a 5-year span of publication. This second test indicated differences between physics and sociology and between physics and engineering, but no difference between sociology and engineering. Suggestions for further study are offered.

INTRODUCTION

This study examines the differences in conceptual citation frequency in science, technology, and social sciences literature. Physics, engineering, and sociology were examined for different rates of conceptual referencing behavior. The three areas were chosen because of their perceived placement on a continuum ranging from "hard" to "soft." This is not a longitudinal study but a description of the current state of referencing behavior. It takes a diachronic slice of the literature and examines that slice as a potential model.

Differences in conceptual citation rates are one indication of previous or current use of another's work. Conceptual citations are examined here as opposed to all citations to determine the effect of conceptual versus methodological citations. Previous work with methodological citations suggested differences in the three areas studied[1]. References, however, will be only a formal indication of the use of concepts. No attempt is made here to track the informal patterns of intellectual activity leading to publication.

Using the topology suggested by Moravcsik and Murugesan[2], in which they identify operational and conceptual citations, this study attempts to measure a hypothesized difference in conceptual citation frequency between the three areas. Examination of conceptual citation rate is one method of examining the intellectual linkages between documents on an unobtrusive level.

Moravcsik and Murugesan defined conceptual references in the following manner.

> If a concept or theory of the cited paper is used directly or *indirectly* in the citing paper in order to lay foundations to build on it or to contribute to the citing paper, then the citation is a conceptual one.
>
> By "using directly" we mean that a concept is taken from the cited paper and the material of the cited paper forms the basis of the citing paper (whether the author states this explicitly or not).
>
> By "using indirectly" we mean the situation when in the course of development of his ideas, the author of the citing paper finds it necessary to incorporate certain ideas or concepts which are not strictly necessary to formulate the basic ideas of his paper but may add more insight or help clarify certain key concepts of the citing paper.[3]

Additional discussion on the utilization of this definition is found in the methodology section below.

The definitions used by Price to differentiate between references and citations will be used throughout this paper.

> . . . if Paper R contains a bibliographic footnote using and describing Paper C, the R contains a *reference* to C, and C has a *citation* from R.[4]

The examination of specific literatures is one attempt to begin defining parameters of scientific and nonscientific literature. This study assumes a structuralist approach to the underlying literature. That is, a literature can be described by the common characteristics that comprise its growth and direction. This study is part of a continuing research program examining individual disciplines and specialty areas. As such it contributes to the accumulating knowledge concerning a literature and its subsets. The ultimate aim is to delineate models for the literature in increasingly finer detail and scope.

PREVIOUS RESEARCH

There is not an extensive body of literature which deals directly with the problem addressed here. The major work to date began with Moravcsik and Murugesan[2], who developed a topology of referencing behavior. Chubin and Moitra's paper[5], which extended the general approach of Moravcsik and Murugesan, was also an exploratory paper. Unfortunately, the two studies reported divergent findings.

Murugesan and Moravcsik[3] presented additional results, which they suggest validated the generalizability of the classification system they developed in their first paper. They found differences in citation behavior in the physics community in various countries and between specialties. They were careful, however, to point out the exploratory nature of their findings.

Hurt[1] examined methodological citation differences in physics engineering and sociology. He reports that physics differs from both engineering and sociology in terms of their methodological citation frequency.

Bertram[6] also dealt with the topology of citations. Her work deals less with the issue of conceptual versus methodological and more with the problem of internal referencing behavior.

In addition to the work referenced above, there is extensive literature dealing with bibliometrics and literature characteristics. Narin and Moll[7] present an excellent historical and methodological review from a number of perspectives. Edge[8] is representative of those less convinced of the utility or validity of the bibliometric techniques. Virgo[9], Frost[10], and Hurt[11] also deal with the overall literature of bibliometrics.

Price[4] suggested that science and technology differed on a series of variables, including an "immediacy effect." He later acknowledged his "immediacy effect" was not the predictor he once thought[12]. Immediacy, however, was only one factor in Price's differentiation of science and technology. His discussion of papyrocentric vs. papyrophobic literature behavior is one example of the differences other than immediacy.

Recent work by Narin and Carpenter[13] suggests that distinctions such as those Price made might once have been valid. They perceive a trend where science and technology are becoming indistinguishable. Price's distinctions need to be examined in light of this and other research programs investigating the nature of formal patterns of scholarly communication. This is even more imperative if Narin and Carpenter are correct. The majority of formal library structures and organizations are predicated on a clear distinction between science, technology, and the social sciences.

METHODOLOGY

Data were gathered for this study from three major sources. These sources corresponded to the three bodies of literature under examination. *Physical Abstracts* for 1984 was examined for the literature of physics. *Engineering Index* annual cumulation for 1984 was used to gather the engineering literature. Finally, the 1984 volume of *Sociological Abstracts* was used to gather the material for sociology. Items were selected from each of these sources at random.

Sample size was determined using a variation of the Scheffé statistic[14], which balances both Type I and Type II error. Type I error was set at 0.05 and Type II error at 0.90. The form of the statistic is given as:

$$n = \Sigma a_k^2 (Z_1 - Z_2)^2 / d_\sigma^2$$

where a_k^2 = weights squared; $Z_1 = Z$ value associated with $Z(1 - \alpha/c)$ where c = number of contrasts; $Z_2 = Z$ value associated with $Z(\beta)$; and d_σ = standardized contrast of interest.

The standardized contrast was set at 0.5, or one-half a standard deviation. The sum of the weights squared was 2. The resultant value for n, the sample size, was 84.5 for each group.

The Scheffé assumes the researcher can set Type I and Type II errors and assign an appropriate standardized contrast of interest. The procedure is a normal approximation algorithm with the added benefits of balancing experimental error and statistical power on a scale of practical significance.

Eighty-five items were selected at random for each of the three literatures from the sources listed above. Each of the papers corresponding to the listing in the index was examined and the following information extracted:

1. Number of references in each paper.
2. Number of items citing material published in the 5 years previous to the publication date.
3. Number of conceptual citations.

Summary measures were computed for each of the three areas and are found in Table 1.

Table 1. Summary measures

	Physics	Engineering	Sociology
Conceptual References			
Mean	9.964	2.893	8.024
SD	6.047	2.024	4.534
All References			
Mean	17.024	6.190	17.452
SD	10.336	3.965	10.138
Five Yr. References			
Mean	12.143	5.310	5.952
SD	6.864	3.307	3.124

The number of references in each paper needs no explanation in terms of operational definition. The number of references citing material published in the 5 years previous to the publication date was used as a measure of variation in the literature areas of the three areas examined. The 5 year period is an arbitrary distinction, but seems to enjoy considerable support in the literature[3,16]. The use of the 5 year cut-off allows this piece of research to be compared with others, hopefully leading to additional research.

The definition of conceptual literature is the most sensitive part of this study. Conceptual literature in one paper might be used as a methodological reference in another. In recognition of this problem, each paper was examined individually. In all cases, the definition for conceptual citations used by Moravcsik and Murugesan was employed.

Because particular differences were the object of the study, no omnibus tests of difference, such as an ANOVA or Chi-Square test, were used. A Dunn test was performed examining each of the pair-wise comparisons[15]. The form of the Dunn test is

$$t_d = d_k/[\mathrm{VAR}(d_k)]^{.5}$$

where d_k is the particular contrast under examination; $\mathrm{VAR}(d_k) = MS_w * \Sigma(a_k^2/n_k)$.

Overall, the Dunn test is a form of a t test, but has the additional property of being able to be used repeatedly throughout an experiment assuming the Type I error level has been adjusted for the number of comparisons.

A contrast matrix was generated to determine the particular weights attached to each of the tests of interests. The matrix is given in Table 2.

HYPOTHESES

The overall purpose of the study is to test for differences in conceptual referencing behavior between three sets of literature. The hypothesis under test (null hypothesis) was defined as: There is no significant difference between the conceptual citation rate of physics, engineering, or sociology. The alternative hypothesis simply stated that at least one statistically significant difference was present. Statistically, these hypotheses were stated as:

H_0: $d_1 = d_2 = d_3$
H_1: H_0 is false.

Given the Type I error level established previously (0.05), a critical value for rejection of the hypothesis under test can be determined. The critical value was determined to be 2.45. Using this value, a decision rule was written to govern rejection of the hypothesis under test.

Decision Rule: With Type I error = 0.05, reject the hypothesis under test, H_0, if the absolute value obtained for t_d is less than 2.45.

Two additional sub-tests were identified. First, a test was conducted to determine if there were differences in the total number of references in each of the three literature groups. A second sub-test examined differences between literature cited in the previous 5 years in each of the literature sets. The same hypotheses and decision rule was used in each of the two sub-tests.

RESULTS

The contrast matrix generated in Table 2 was used to obtain the difference or contrast values for each of the individual tests. Table 3 displays the contrast values and Dunn values for the major test of the study, differences in conceptual referencing behavior. There are significant differences among all three sets of literature in terms of conceptual citations.

One secondary test was to examine the total citations in each area for differences. The results of this test, together with the contrast values are in Table 4. Significant differences exist between physics and engineering and between sociology and engineering. The test for difference between physics and sociology in terms of overall references is not significant.

Table 2. Contrast matrix

d_1 = 1(Physics) +	0(Engineering) +	−1(Sociology)
d_2 = 1(Physics) +	−1(Engineering) +	0(Sociology)
d_3 = 0(Physics) +	1(Engineering) +	−1(Sociology)

Table 3. Conceptual references: Contrast and Dunn values

	Contrast Value	Dunn Value	Difference
Physics–Sociology	1.94	2.799	significant
Physics–Engineering	7.07	10.204	significant
Engineering–Sociology	5.13	7.405	significant

Table 4. Total references: Contrast and Dunn values

	Contrast Value	Dunn Value	Difference
Physics-Sociology	−0.428	−0.228	
Physics-Engineering	7.314	3.89	significant
Sociology-Engineering	11.262	5.989	significant

Table 5. Previous 5-year literature: Contrast and Dunn values

	Contrast Value	Dunn Value	Difference
Physics-Sociology	6.191	8.49	significant
Physics-Engineering	6.833	9.369	significant
Sociology-Engineering	0.642	0.880	

The final test was to examine the three areas for differences arising from their frequency of use of literature published within five years of the referencing paper. The results are given in Table 5. Significant differences between physics and sociology and between physics and engineering appear. There is no significant difference found between sociology and engineering in this test.

CONCLUSIONS

The results of the tests above point to the following conclusions:

1. Physics differs from engineering in all areas tested.
2. Sociology and physics differ from each other in terms of frequency of conceptual references and in terms of use of literature five or fewer years old. There appears to be no significant difference in terms of the overall number of references.
3. Sociology and engineering show no difference in terms of the number of references to literature five or fewer years old. Differences are apparent between the two in their frequency of conceptual references and the number of references overall.

DISCUSSION

The most straightforward methodology for defining a literature is to define its characteristics. The literature may then be examined from the standpoint of similarity or difference. This study indicates some clear similarities and differences in the three literature areas studied. The area of examination of particular literature-based similarities or differences needs expansion. It is through the delineation of such differences and similarities that the best models for utilization of the literature will be developed. Clearly, this is an area of concern to information providers, such as libraries.

This study points to some differences that add to the store of knowledge concerning bodies of literature in science, technology, and the social sciences. It extends a previous study dealing with methodological literature and a study based solely in physics[16].

The results here indicate that there are significant differences in the frequency of conceptual references in the three areas. Only the formal references to published literature were examined. It is possible an informal mechanism may be operating that circumvents the formal structure. This study would not have found such a phenomenon.

Differences in areas of total references and in references to recent literature were also uncovered. These differences are, for the most part, consistent with previous findings[1]. Only the test for difference in total number of references between physics and engineering is at variance. Additional work will be necessary to explore potential reasons for the variation.

At this point, only general hypotheses can be generated to explain the results. The first hypothesis proposed is that there are still major differences between science and technology. Narin and Carpenter have evidence to suggest otherwise[17]. Their study was based on patent literature and, as such, the contradictory results are not directly comparable. At the minimum, we can hypothesize differences in the formal communication patterns other than patents between physics and engineering. What appears clear from this study is that physics and engineering are very different. This is a surprising finding in light of the perceived natural connection between science and technology.

The lack of difference between sociology and physics in terms of overall number of references may be the result of sociology becoming more like the hard sciences in terms of its publications. The sociometric advances of the past 30 years have increased the ability of the field to conceptualize.

SUGGESTIONS FOR FURTHER RESEARCH

Replication of this study is a necessity. Additional work both with these three areas and with other literature areas is a fruitful and necessary research program. The study of formal transfer of information, as viewed through citation analysis, is one vehicle for turning good theory into good practice. Further research should be done in terms of applying the findings here and elsewhere to libraries and information centers. The assumption that modeling the library or information center on the structure and use of the literature is in need of robust testing.

REFERENCES

1. Hurt, C. D. Methodological citation differences in science, technology, and social sciences literatures. Library and Inf. Sci. Res. 7:345-355; 1985.
2. Moravcsik, M. J.; Murugesan, P. Some results on the function and quality of citations. Social Studies of Sci. 5:86-92; 1975.
3. Murugesan, P.; Moravcsik, M. J. Variation in the nature of citation measures with journals and scientific specialties. J. Am. Soc. Inf. Sci. 29:141-147; 1978.
4. Price, D. Citation measures of hard science, soft science, technology, and nonscience. In: Nelson, C. E.; Pollock, D. K., editors, Communication among scientists and engineers. Lexington, MA: Heath; 1970.
5. Chubin, D. E.; Moitra, S. D. Content analysis of references: Adjunct or alternative to citation counting? Social Studies of Sci. 5:423-441; 1975.
6. Bertram, S. K. The relationship between intra-document citation location and citation level. Chicago: University of Chicago, 1970. Dissertation.
7. Narin, F.; Moll J. K., Bibliometrics. Annu. Rev. Inf. Sci. & Tech. 12:35-58; 1976.
8. Edge, D. Quantitative measures of communication in science: A critical review. History of Sci. 17:102-134; 1979.
9. Virgo, J. A statistical procedure for evaluating the importance of scientific papers. Chicago: University of Chicago, 1974. Dissertation.
10. Frost, C. O. Use of scholarly materials in German literary research: An analysis of reference citations in publications about literary works in the German language. Chicago: University of Chicago, 1977. Dissertation.
11. Hurt, C. D. Important literature identification in science: A critical review of the literature. Adv. in Libr. 13:239-258; 1984.
12. Price, D. A general theory of bibliometric and other cumulative advantage processes. J. Am. Soc. for Inf. Sci. 27:292-306; 1976.
13. Narin, F.; Carpenter, M. Are technology and science becoming indistinguishable? Am. Soc. for Inf. Sci. Proc. 21:199-203; 1984.
14. Scheffé, H. A method for judging all contrasts in the analysis of variance. Biometrika 40:87-104; 1953.
15. Dunn, O. J. Multiple comparisons among means. J. Am. Stat. Assoc. 56:52-46; 1961.
16. Hurt, C. D. Conceptual citation frequency-Quantum mechanics and elementary particle physics. Czech. J. Phys. B6:62-64; 1986.