Information Processing & Management Vol. 19, No. 3, pp. 151–157, 1983
Printed in Great Britain.

0306–4573/83/030151–07$03.00/0
Pergamon Press Ltd.

A COMPARISON OF A BIBLIOMETRIC APPROACH AND AN HISTORICAL APPROACH TO THE IDENTIFICATION OF IMPORTANT LITERATURE

C. D. HURT

Graduate School of Library Science, McGill University, 3459 McTavish Street, Montreal, Quebec, Canada H3A 1Y1

(*Received for publication* 15 *October* 1982)

Abstract—This study examines the problem of identification of important literature in a specific scientific area, quantum mechanics. An examination was conducted using two literature sets. The first set was identified using a bibliometric approach and the second was identified using an historical approach. A gamma test of association was employed, resulting in a finding of no significant association between the two files of important literature. Particular attributes of the literature under examination were also studied. These tests supported the finding of no association. Validation testing was done to insure the integrity of the results. The major conclusion in the study was that the use of citation analysis alone or historical analysis alone will not result in the same set of literature being produced. Use of one method singly appears to be risky since the second method of selection produces an entirely different literature. Areas for further investigation of this problem are suggested.

INTRODUCTION

The purpose of this study was to test for association between two independent means of selecting important literature. The testing involved taking historical treatments of a scientific specialty and matching these against the actual citation records for the specialty. The major test was for overall association between ranks of individual pieces of literature. Ranks were determined by frequency. The scientific specialty examined was quantum mechanics, particularily during the years 1900–35. Data were collected from a series of histories regarding the rise of quantum mechanics as well as from research papers published during the years 1900–35. Additional testing was done to insure validity of the results.

PREVIOUS RESEARCH

There is a great deal of research which deals with bibliometrics and literature characteristics. This literature is not confined to the library and information science sources but is a topic in sociology and history of science as well. Narin and Moll[1] present a good historical and methodological review of bibliometrics from a number of perspectives. Edge[2] represents those who are not convinced of the utility or validity of bibliometrics. The dissertations of Bertram[3], Virgo[4], Frost[5] and Hurt[6] deal with the overall literature of bibliometrics in a more extensive fashion.

The subset of the general literature of bibliometrics dealing with comparisons of methodologies for determining important literature is not as extensive. It is this subset of the literature, and less the general literature of bibliometrics, which is of interest in this study.

One of the first studies to compare methodologies was Garfield *et al.*'s [7] examination of the DNA structure discovery. Their study took Asimov's[8] chronicle of the discovery and initiated literature searches beginning with "nodal events" approved by Asimov. Although reporting success, the authors failed to explain why Watson and Crick's first two articles in *Nature* were missing from the list of important papers.

Virgo's[9] work with prediction of important scientific papers used both citation frequency ranks and expert judges' ranks of the same literature. She concluded that citation frequency was able to predict the important papers. The use of expert judges presents methodological problems. Mulkay and Edge[10] found that the participants in major discoveries recounted

those discoveries in different ways. Retesting their subjects, they found that the same respondents recounted the discovery in yet another order. Most experts tend also to be authors and it is unclear if methodologies employed for the selection of important literature by authors are different from those employed by judges who are also authors.

In the most direct test, Studer and Chubin[11] examined endocrinology using citation analysis and historical records. They determined from a descriptive analysis that when different techniques were used to define important literature, different sets of important literature were produced. This conclusion was based on visual inspection and not tested hypotheses.

The literature referenced above is by no means a complete enumeration of the available literature. It is, however, the material considered crucial to this study.

DATA COLLECTION

Data were collected using two methods and fell into two data files, Bibliometric or Historical. The Bibliometric File was built using the base year 1932. This particular year was chosen because the historical accounts of the quantum mechanics problem indicate that, while as far as principles were concerned quantum mechanics was complete by 1927, the theory was incomplete in terms of mathematical formalism. Von Neumann's formalism, published in 1932, supplied an appropriate end date for the development of quantum mechanics and signaled an acceptance of the theory as a working tool of physics[12].

A five year time lag was built into the collection process to allow for publication and entry into the indexing and abstracting tools[13]. *Physics Abstracts* was used to generate the base literature for the Bibliometric File. The file was cumulated by moving backwards through time to the year 1900. The base literature was generated by examining items found under the subject heading, "Quantum Mechanics", in *Physics Abstracts* for the year 1937. This base literature was then used to generate the remainder of the file on a year-by-year basis. All references within a given year were analyzed. However, the stepping backward through time on the year-by-year basis was done only to 1900. *Physics Abstracts* was not used beyond the identification of the literature in the base year due to variance in indexing and the non-availability of the heading, "Quantum Mechanics", in earlier volumes.

Following the generation of the Bibliometric File, items were ranked in descending order according to citation frequency. There was no weighting of items.

The Historical File was generated by examining historical accounts of the quantum mechanics problem. The references used by each of the historians to buttress their cases were noted and collected. Histories were chosen in consultation with a historian of science familiar with the subject and the method. A list of the sources used is included as Appendix 1. Items in the file were ranked in descending order according to citation frequency. There was no weighting of items.

METHODOLOGY REJECTED

In the course of beginning this investigation, a major false start was encountered. The prime objective of the study was to determine the relative merits of two techniques of identifying important literature. The most direct method of accomplishing this seemed to be the use of a criterion variable or file against which to measure the merits of the other files. Three files were planned: a Bibliometric File, an Historical File and an Expert File. The Expert File was to have been generated using reading lists and course syllabi from graduate physics courses at major universities. It quickly became apparent that such reading lists and syllabi dealt with the application of quantum mechanics as a tool in physics and not with the emergence of quantum mechanics. Original papers in the quantum mechanics literature were simply not being cited.

As a result of not finding original papers in quantum mechanics literature in course reading lists and syllabi, the creation of the Expert File was abandoned. The lack of a criterion file changed the methodology involved in testing for differences between the two files. A ranked testing procedure was chosen to replace the more direct test of the two files against an Expert File.

SELECTION OF ATTRIBUTES

The objective of the selection of the particular attributes was to produce a set of characteristics potentially meaningful on a theoretical as well as a practical level. The attributes

were chosen to reflect different facets of the literature. Additionally, these attributes could be compared with the findings of previous studies.

The attributes chosen were culled from a list of potential attributes suggested by previous research. The Fussler study is a prime example[14]. Fussler studied the "temporal span" of the literature in chemistry and physics. He also investigated the principal literature used in the two areas, the national origin of the literature used (although only for that literature used in the United States), as well as attempting to determine the important serial titles for each field. The importance of the Fussler study to this and other research into the utilization of scientific information is significant.

Another influence on the selection of the attributes was the work of Hagstrom in dealing with different modes of publishing in different scientific fields[15]. Hagstrom's work paid particular attention to the presence or absence of review articles, monographs, and textbooks in subject fields. The present study attempted to investigate similar attributes in a smaller and more cohesive group than those Hagstrom studied.

Finally, this study is indebted to the work of Sullivan, White and Barboni, who investigated the specialty area of weak interaction physics[16]. Their study examined the productivity of the theoretical and the experimental physicist. In doing so, they investigated the number of authors per article, "transient" authorship, "professional" authorship, the number of references per article, references to non-articles, as well as a number of other variables not used here such as reference half-life.

DATA ANALYSIS

The overall test of this investigation was to either substantiate or invalidate the expectation that the two methods, Historical and Bibliometric, produce different (nonassociated) sets of important literature for the same field. The test chosen was the gamma test of association as suggested by Goodman and Kruskal[17]. The gamma statistic has the same basic interpretation as the Kendall tau: a probability difference for the same versus different ordering on the underlying variables given a randomly selected pair. The gamma statistic was suggested by Goodman and Kruskal to improve the interpretability of the Kendall tau when ties are present in either or both the rankings, as was the case here. The form of the gamma statistic is:

$$\gamma = P - Q/P + Q$$

where P = any cell with non-zero frequency and, ignoring its row and column in a ranked joint-frequency table, summing the number of entries to the right and below that cell, Q = any cell with a non-zero frequency and, ignoring its row and column in a ranked joint-frequency table, summing the number of entries to the left and below that cell[18].

If there are no tied ranks, then

$$\gamma = \tau.$$

HYPOTHESES

The purpose of this study was to investigate whether or not there are statistically significant differences in the identification of important literature resulting from two approaches of examining a specific field. The hypotheses testing for overall association or difference, Ho, and its alternative, H1, were:

Ho: There is no significant association between the ranks of literature identified by means of historical accounts and the ranks of literature identified by means of citation analysis.
H1: There is a significant association between the ranks of literature identified by means of historical accounts and the ranks of literature identified by means of citation analysis.

The statistical form of the hypotheses were:

$$\text{H0: } \gamma = 0$$
$$\text{H1: } \gamma \neq 0.$$

A decision rule for rejection of the hypothesis under test, Ho, can be formed directly from the gamma statistic. Gamma may take values ranging from -1.00 to $+1.00$. The sign of the value obtained for gamma is an indication of the direction of association. Using the absolute value of gamma as a probability value, an appropriate decision rule can be written where the Type I error level or $\alpha = 0.05$ for a two-tailed test of association:

Decision Rule: Reject Ho if the absolute value obtained for the gamma statistic is greater than 0.975.

If this test of association is negatively significant, the value obtained for gamma is interpreted as the probability of obtaining different ranks if given a pair of items chosen at random, one from each file. If the test of association is significant in the positive direction, the gamma value is interpreted as the probability of obtaining the same ranks if given a pair of items chosen at random, one from each file. If the test of association is not significant, the gamma value is interpreted as there is a statistically significant difference or nonassociation between the ranks of items in the two files and statistical independence can be inferred.

The hypothesis under test was formulated under the expectation that lack of association would be discovered and independence or difference was present. Building on this expectation, an examination was conducted of the ways in which the two files might differ. A set of particular attributes (see Table 1) was examined in an effort to identify differences. Rather than use an omnibus Chi-Square test followed by phi or Cramer statistics, Goodman and Kruskal suggest an index of predictive association[19]. In the context of this study, it is desirable to have a symmetric measure of the power to predict: (1) the file from knowledge of an attribute or, (2) an attribute from knowledge of the file.

The form of the symmetric measure of predictive association as defined by Goodman and Kruskal is given as:

$$\lambda_{AB} = \frac{\sum_i \max_k f_{jk} + \sum_k \max_j f_{jk} - \max_k f_k - \max_j f_j}{2N - \max_k f_k - \max_j f_j}$$

Table 1.

ATTRIBUTE	LAMBDA VALUE
author	.07540
number of authors	.00000
year	.06667
title	.13437
type of material	.00563
location of author	.13701
recognized Author	.08419
number of journal references	.01618
number of monographic references	.04462
number of conference reports	.00000
number of handbook references	.00000
number of preprint references	.00000
number of references to lectures	.00000
number of festschrift references	.00000
number of references to local material	.00000
number of references to technical reports	.00000
number of references to letter journals	.00000
number of references to dissertations	.00000
number of references to explanatory material	.04167
total number of references	.03566

where f_{jk} is the frequency observed in cell (A_j, B_k), $\max_k f_{jk}$ is the *largest* frequency in column A_j, and $\max_k f_k$ is the largest *marginal* frequency among the rows B_k.

The interpretation of the lambda value is that the probability of predictive error is reduced by the value of lambda given knowledge of: (1) from which file the item came or, (2) the particular attribute. The object is prediction of the other category.

It should be pointed out that this measure of predictive association is not equivalent to the Chi-Square, phi, or Cramér tests of independence and association. It is possible for statistical association to exist even though the value of lambda is zero. The measure of predictive association is an index of the change in probability given certain information and is not a direct measure of independence. However, if the Cramér or phi statistics equal zero, the lambda value must also equal zero. The converse is not true.

VALIDITY

One of the techniques of the historian is to divide chronologies of events into periods. To investigate the possibility of this phenomenon in the historical accounts, the files were divided into two periods. The first period represented what was called the "Old Quantum Theory" period and was defined by items published prior to 1916[22]. The second period, the "New Quantum Theory" was also examined. This second period was defined by items published after 1915. Specifically, the files were tested for overall differences using the same analysis and interpretation used on the composite files.

The construction of the two major files has the potential for ceiling and floor effects on the data. In the case of the Bibliometric File, only items with a publication date of pre-1938 could be included. There was no such restriction on the Historical File. In the same way, the Bibliometric File included references to literature, regardless of date, as long as the referencing paper was published between 1900 and 1937 inclusive. The results of tests on these two sets of literature could well result in an artifact being produced; an artifact which could either indicate lack of association when, in fact, association existed or the converse. In order to make the two sets of literature more comparable, a cutoff range of 1900 through 1935 was made on both files and the two restricted files then subjected to the same tests of association as the unrestricted files.

RESULTS

The primary test in this study was a test for association between the literature identified as important by the Historical File vs. the literature identified as important by the Bibliometric File. The obtained value for the test statistic, gamma, was -0.11472. Under the decision rule established above, the hypothesis under test, Ho, could not be rejected.

Therefore, the interpretation of the results of the test for overall association was that there is no statistically significant association between the two files of literature. As a result, independence between the two files of literature may be inferred. As a further test of association between the two files, only the 146 items which were in common to both files were subjected to the gamma test. (Table 2 is a joint frequency table for items in the Historical File and items in the Bibliometric File.) The results of this second test were consistent with the

Table 2. Joint frequency table

Bibliometric File	Historical File: contained	Historical File: not contained
contained	146	1000
not contained	323	unknown

results just reported. A gamma value of −0.2675 was produced, still well within the critical value for rejection of the hypothesis under test.

In addition to testing the primary hypothesis for overall association, this study also examined the predictive association of a number of attributes in the two files of literature. A symmetric measure of the power to predict either the file from knowledge of the attribute or the attribute from knowledge of the file was used. The lambda statistic was used to measure the predictive association. The lambda statistic has a range from 0 to 1, where 0 indicates no predictive ability while 1 indicates perfect prediction. Table 1 lists the lambda value associated with attributes in the overall file. The results of the lambda tests were consistent with the findings of no overall association between the Historical and the Bibliometric Files.

In order to test for historians' period propensity, the Composite File was partitioned into the Old Quantum Theory File and the New Quantum Theory File. The gamma statistic, testing for overall association, was employed. The results of the two tests indicated no association. The New Quantum Theory File test produced a gamma value of −0.15600 and the Old Quantum Theory File test produced a gamma value of −0.04340. Both values were within the critical value, therefore the hypothesis of no association could not be rejected.

Data were gathered for this study with great care. Nonetheless, the possibility for ceiling and floor effects was present. To control for this possibility, the Composite File was restricted only to the years 1900 through 1935. The rationale for these dates was that they included the major dates for both Old and New Quantum Theory. The upper limit of the file, 1935, coincides with the formalism of the quantum theory.

Following the pattern of the testing done previously, the gamma test was employed to test for association. The same critical value as the original overall test for association, 0.975, was used. The results of the test produced a gamma value of 0.485, well within the critical value, forcing a rejection of the hypothesis of association. The value obtained can be interpreted as there being no statistical association between the two files of literature even when restricted to common dates. Again, statistical independence is indicated by lack of association.

CONCLUSIONS

This investigation lead to three major conclusions:

(1) Association between the ranking of cited items using a citation analysis approach and the ranking of cited items using an approach which pools the references used by historians in their reconstruction of the past is statistically nonsignificant. From no association, statistical independence can be inferred.

(2) Association between the attributes chosen to investigate particular areas of potential difference in the two distributions is nonsignificant and supports the conclusion above.

(3) The use of citation analysis alone or historical analysis alone will not result in the same set of literature being produced. Using citation analysis to predict important literature in a scientific area appears to be risky, since another method which should produce much of the same literature does not.

SUGGESTED FURTHER RESEARCH

This study points to the need for further research in a number of areas. There is a need for replication and validation of the present study. Beyond this, there is a need for replication and validation in areas both in the "hard" and the "soft" sciences. Such research might indicate that quantum mechanics is an anomoly in the structure of science and that differences between it and other areas of science are significant.

Additional work needs to be done in validating the journal article analysis technique and the interpretation of such analyses. As this study has shown, citation analysis is an approximate measure rather than an exact measure. More work needs to be done to narrow and define the limits of the approximation.

Finally, this study found that the two methods employed gave different results in terms of important literature. Additional work should be done to determine if combining the two approaches might produce better results than each singly. Other approaches should be

examined as well for their potential in combination with different approaches. A necessary prelude to the investigation of combinatorial approaches is the determination of an acceptable criterion by which to judge the results of identification of important literature.

REFERENCES

[1] F. NARIN and J. K. MOLL, Bibliometrics. *Ann. Rev. Inform. Sci. Tech.* 1976, **12**, 35–58.
[2] D. EDGE, Quantitative measures of communication in science: a critical review. *History of Science* 1979, **17**, 102–134.
[3] S. J. K. BERTRAM, The Relationship between intra-document citation location and citation level. 1970, Ph. D. Dissertation, Univ. of Illinois.
[4] J. VIRGO, A statistical procedure for evaluating the importance of scientific papers, 1974, Ph. D. Dissertation, Univ. of Chicago.
[5] C. O. FROST, Use of scholarly materials in German literary research: an analysis of reference citations in publications about literary works written in the German language, 1977, Ph. D. Dissertation, Univ. of Chicago.
[6] C. D. HURT, Identification of important literature in quantum mechanics: investigation of a bibliometric and an historical approach, 1981, Ph. D. Dissertation, Univ. Wisconsin-Madison.
[7] E. GARFIELD, I. H. SHER and R. J. TORPIE, *The Use of Citation Data in Writing the History of Science.* Institute for Scientific Information, Philadelphia (1964).
[8] I. ASIMOV, *The Genetic Code.* New American Library, New York (1963).
[9] J. VIRGO, See Ref. [4].
[10] M. J. MULKAY and D. EDGE, Cognitive, technical and social factors in the growth of radio astronomy. *Soc. Sci. Inform.* 1973, **12**, 25–61.
[11] K. E. STUDER and D. CHUEIN, The heroic age of endocrinology. Paper presented at the *1st Annual Meeting of the Society for the Social Studies of Science*, Ithaca, New York, 4–6 Nov. 1976.
[12] J. VON NEUMANN, *Mathematische Grundlager der Quatenmechanik.* Springer-Verlag, Berlin (1932).
[13] For a discussion of the issue of time lag in publication, see W. D. GARVEY, N. LIN and C. E. NELSON, Some comparisons of communication activities in the physical and social sciences. In *Communication Among Scientists and Engineers* (Edited by C. E. NELSON and D. K. POLLOCK). Heath, Lexington, Mass. (1970).
[14] H. H. FUSSLER, Characteristics of the research literature used by chemists and physicists in the United States. *Library Quart.* 1949, **19**, 19–35.
[15] W. O. HAGSTROM, Factors related to the use of different modes of publishing research in four scientific fields. In *Communication Among Scientists and Engineers* (Edited by C. E. NELSON and D. K. POLLOCK). Heath, Lexington, Mass. (1970).
[16] D. SULLIVAN, D. H. WHITE and E. BARBONI, The state of a science: indicators in the speciality of weak interactions. Paper presented at the *70th Meeting of the American Sociological Assoc.*, San Francisco, 27 Aug. 1975.
[17] L. A. GOODMAN and W. H. KRUSKAL, Measures of association for cross classifications. *J. Am. Statist. Assoc.* 1954, **49**, 733–764.
[18] For a more complete description and discussion of the gamma statistic, see J. H. MUELLER, K. F. SCHUESSLER and H. L. COSTNER, *Statistical Reasoning in Sociology*, 2nd Edn. Houghton Mifflin, Boston (1970).
[19] L. A. GOODMAN and W. H. KRUSKAL, See Ref. [17].

APPENDIX 1

List of historical sources

Hund, Friedrich, THE HISTORY OF QUANTUM THEORY, (London: Harrap, 1974).
Jammer, Max, THE CONCEPTUAL DEVELOPMENT OF QUANTUM MECHANICS, (New York: McGraw-Hill, 1966).
ter Haar, D., THE OLD QUANTUM THEORY, (New York: Pergamon, 1967).
HANDBUCH DER PHYSIK, Second Edition, (Berlin: Springer, 1955-).