Teaching Bibliometrics

ALVIN M. SCHRADER

BIBLIOMETRICS, THE SCIENTIFIC STUDY of recorded discourse, offers much promise for enhancing university curricula in the informational domain. This promise involves two dimensions of empirical knowledge, a theoretical dimension and a practical dimension, and so ought to interest not only researchers and educators but professional practitioners as well. This promise issues from the special nature of empirical knowledge, by which ideas about the world can be related to practical activity. The special nature of such knowledge is derived from what might be called a metatheory about the logic of inquiry.¹ This metatheory is outlined below.

Bibliometrics taken as theoretical knowledge is the quantitative characterization of the properties of recorded discourse. Quantitative characterization is the setting forth of probabilistically true ideas about selected phenomena. These ideas express patterns, tendencies and regularities that are said to be inherent in the phenomena. Such ideas, because they describe general qualities, form "empirical theory" or just "theory." Maccia (now Steiner) and Maccia put it this way: "Understanding should lead to explanation, because understanding provides relationships or regularities which make sense of our happenings. To explain is to appeal to regularities, i.e., to appeal to theory."² Thus, the objective of bibliometrics as a scientific study is to produce ideas—that is, theory—about recorded discourse and its various important properties.

Alvin M. Schrader is a doctoral candidate, School of Library and Information Science, Indiana University, Bloomington.
In addition, bibliometrics is considered to have promise in the realm of practical knowledge, because theory permits control. More is involved, however, than simply theory. A developmental bridge is required by which theoretical knowledge is related to both the means and the ends of the proposed practice. It is not only the effectiveness of a practice that must be considered, but also its intrinsic merit, for a practice is a system of human acts devised to bring about an intended condition, and so involves values. This linking process from theory to practice is described as development inquiry, operations research, or systems analysis, though the latter two terms have generally connoted a much narrower perspective of means-oriented research only.

If bibliometrics as seen in the context of metatheory has theoretical and practical dimensions, that it can contribute both to our intellectual understanding and to the control of professional activity, then it is plausible that bibliometrics contains elements of a scientific discipline, or, at least, for undergirding such a discipline within the domain of informational phenomena and problems. But if bibliometrics has so much promise, where is the spark that will inspire curiosity and consensus about this domain, and launch the needed programs of empirical inquiry?

The missing ingredient is the collective imagination and commitment of our community of educators and researchers. True, under the disciplinary umbrella of information and library science, one can identify a small (and growing) constituency of enthusiasts who take as self-evident the power of quantitative research to enhance thinking about informational phenomena and problems. Unfortunately, however, most members of this amorphous scholarly community have proceeded through graduate school and on to professional practice and teaching and research without even seeing the term bibliometrics in print. They still speak of universal bibliographic control as though it were a meaningful concept, and do not accept the notion that recorded discourse consists of a set of many overlapping literatures, each of which exhibits a statistical structure.

This unsatisfactory condition is exacerbated by library school doctoral programs which, with few exceptions, are still very weakly committed to quantitative research in general—and even more weakly committed to bibliometrics in particular. There are many impediments within the graduate library schools to the attainment of scholarly excellence in mainstream academia. These impediments add up to an inventory of neglect and intellectual confusion. Among the most relevant to teaching bibliometrics are present library school curricula, research methods textbooks, and the professional literature.
Teaching Bibliometrics

With respect to curricula, only a few library schools offer a bibliometrics course, and almost always on an ad hoc basis; some individual faculty have inserted isolated components into traditional courses. The directory of the Association of American Library Schools for 1980 did not list bibliometrics in its classification of teaching areas. This is an important indication of scholarly attitudes toward it.

A second illustration of impediments to bibliometrics concerns research methods textbooks. In the one most recently published for graduate library school students, Busha and Harter devote only one-half page to bibliometrics, while other methodologies receive much greater priority: five pages for content analysis, a 20-page chapter for operations research, and a 30-page chapter for historical method. Such a long discussion of historical method, enigmatic in the context of graduate education for information professionals in the 1980s world of scientific advance and managerial accountability, reflects persistence of the old library school ideology, an ideology of 100 percent book collections, scholar-librarians, parochial history essays, and white gloves.

Another impediment to bibliometrics in library schools concerns the professional literature and its bibliographic control. The *Journal of Education for Librarianship*, for example, has published over the past twenty years something less than a handful of articles which employed a bibliometric analysis, and none at all which investigated a bibliometric methodology and its assumptions. Another similar indicator of the absence of interest among educators and researchers in bibliometrics is the fact that only one comprehensive review article, by Narin and Moll, has appeared in the *Annual Review of Information Science and Technology* since its inception in 1966—despite their confident prediction in that review that future issues would treat bibliometrics in greater depth. No general reviews at all have appeared in *Advances in Librarianship* since it began in 1970, though for the record it should be noted that it did publish a review of one type of bibliometric application to library collection building, by Broadus.

With respect to bibliographic control of the literature of bibliometrics, Ferrante has indicated that fifty-two synonymous and semisynonymous search descriptors were required to retrieve the relevant publications during the period from 1969 (when Pritchard first introduced the term *bibliometrics* in place of *statistical bibliography*) until 1977. She noted that: "While *Library and Information Science Abstracts* and *Library Literature* both picked up the term 'bibliometrics' by 1971, *Information Science Abstracts* vacillated until 1973....Neither ERIC nor L.C. Subject Headings include the term among their subject headings...."
These illustrations of impediments to the introduction of bibliometrics into graduate library school curricula can be placed in the larger perspective of major weaknesses in the knowledge base of educators and researchers. The major weaknesses are seen to be their atheoretical approach to problem-solving and their elementary descriptive approach to quantification.

The atheoretical approach to problem-solving is illustrated pointedly by the semantic confusion in the literature between theory and philosophy, in that pleas for a philosophy of library science are taken to be pleas for theory, and the terms are used interchangeably. Philosophy, however, is value theory and is sorted out in logic and epistemology from empirical theory, so that ideas about what ought to be and what ought to be done are differentiated from ideas about what exists in the world. Value theory is not a substitute for empirical theory, but rather, as has been demonstrated already, is a necessary complement in development inquiry which links theory to practice. In any event, pleas for a philosophy of library science have usually boiled down to weak attempts to rationalize the genteel empiricism in which educators and researchers have functioned since the 1870s.

A second major weakness concerns educators' and researchers' traditionally elementary approach to quantification. The charge is frequently made that librarians are hostile to numeracy and quantitative research, but this charge seems inadequate as a description of practitioners' attitudes toward quantitative expression. In fact, numbers as quantifiers of library activity and library services are not merely simple-mindedly avoided or despised, but on the contrary are universally employed to describe such variables as library holdings, book circulation and salaries. The problem is not professional hostility, fear, anxiety, or other psychoanalytic peculiarities brought by students to graduate library schools. The problem is that educators and researchers have left the professional community innumerate and deficient in dealing adequately with quantification. How can graduates go beyond elementary description of data if they have not been educated to do so? How are they to learn that mere data collection is not the complete act of research if their educators teach that it is? How are they to come to an understanding of what Cole and Eales meant in 1917 by a "statistical analysis of a literature"? Or what Hulme meant in 1923 by "statistical bibliography of scientific literature" for documenting the history of science? Or what Lotka meant in 1926 by the "logarithmic frequency distribution" of scientists' productivity to the progress of science as indicated by publications? Or what Bradford meant in 1934 by the
"law of distribution of papers on a given subject in scientific periodicals"? Or what Gosnell meant in 1944 by treating book collections as "populations" with averages and general trends, one of which was that book obsolescence rates correspond to an "exponential curve"?

The quantitative literature—though sparse—has always been there. Library school educators and researchers have not. Presumably, security of institutionalization in university graduate departments has lulled them into complacency with the status quo. However, it is altogether probable that the intellectual confusion which has resulted from this complacency will not satisfy the academic demands posed by an information-consuming world. If the informational community eventually attains a higher-order social role, its emergence from atheoretical empiricism and innumeracy may well turn out to emulate the history of the medical profession, described succinctly by Thomas:

For century after century, all the way into the remote millennia of its origins, medicine got along by sheer guesswork and the crudest sort of empiricism. It is hard to conceive of a less scientific enterprise among human endeavors. Virtually anything that could be thought up for the treatment of disease was tried out at one time or another, and, once tried, lasted decades or even centuries before being given up. It was, in retrospect, the most frivolous and irresponsible kind of human experimentation, based on nothing but trial and error, and usually resulting in precisely that sequence. Bleeding, purging, cupping, the administration of infusions of every known plant, solutions of every known metal, every conceivable diet including total fasting, most of these based on the weirdest imaginings about the cause of disease, concocted out of nothing but thin air—this was the heritage of medicine up until a little over a century ago. It is astounding that the profession survived so long, and got away with so much with so little outcry.

A rationale for moving bibliometrics into the mainstream of graduate library school curricula has been set forth based on the logic of inquiry. Indeed, bibliometric knowledge ought to be integrated into existing courses and, at the same time, specialized programs ought to be offered at both the MLS and Ph.D. levels for advanced study of both theory and methodology. There is a growing body of researchers and educators who are utilizing and extending bibliometrics, and some scholarly community will no doubt lay claim to this domain in the near future. If that scholarly community is not the library schools as presently constituted, then there are other plausible claimants, including (but not limited to) academic programs of information science, sociology of knowledge, computer science, public policy, education, and history and philosophy of science. Indeed, the pioneering advances in
relevant theory have so far come from scholars outside the library schools, scholars such as Merton in the sociology of science, Kuhn in the history of science, and Price in the history of science and medicine.

If none of the foregoing arguments for teaching bibliometrics has been convincing, the only remaining appeal is to an observation attributed by Pritchard to Fairthorne: "Numerical data may or may not be dull, but they are the only alternative to thumping the table and affirming one's intuitions."¹⁵

Proposal for an MLS Course in Bibliometrics

The proposal for a course in bibliometrics set forth here is notably tentative and pertains to the MLS level; doctoral work in bibliometrics should focus on theory construction and testing, and on advancing the methodology and statistical techniques. The only previous discussion in the literature of teaching bibliometrics was by Aiyepeku,¹⁶ but he did not furnish an exemplar syllabus, which is the intention of this article.

Proposed course objectives are: (1) to teach students the basic principles of bibliometrics as related to scholarly literature; (2) to work toward the construction of adequate theory of bibliometrics; and (3) to review the practical applications of bibliometric methods for information retrieval systems. The emphasis of the course will be on the theoretical aspects of bibliometrics within the framework of compatible research traditions such as epistemology, sociology of knowledge, scientific communication theories, and history and philosophy of science. Students will familiarize themselves with the seminal papers and landmark literature of bibliometrics; examine major problem areas for definitions, key assumptions, methodological procedures, and statistical distributions; and formulate theoretical statements.

No course prerequisites are assumed, but much of the substance of bibliometrics involves the logic of inquiry and techniques of quantification; hence math anxiety should be avoided. Since standard parametric statistics are generally not utilized in describing and evaluating bibliometric distributions, there is no reason to require advanced familiarity with them; an understanding of nonparametric statistical tests (e.g., Siegel¹⁷) and lognormal distributions (e.g., Pratt¹⁸) would be very helpful, but unrealistic to require of MLS students. At the doctoral level, however, learning these nontraditional statistical procedures and distributions should be a major priority, so that a core of numerate researchers can be developed for advancing the theory and methodology of bibliometrics.
Teaching Bibliometrics

A suggested range of student assignments for the MLS course follows:

1. A citation analysis of a library and information science journal with respect to core journals, journal-to-journal citation, core of authors, journal scatter, or subject dispersion.
2. Using the Sweaney interpretation of Bradford’s law, plotting two sets of data and calculating possible estimates for the parameters of journal variables, articles per zone, and multiplier. Alternate projects are plots for Lotka’s law or for Pratt’s measure of class concentration.
3. A bibliographical analysis of the literature of one of the following subjects: referencing theories; typologies of citations; citation errors; bibliographic coupling; cocitation analysis; author collaboration; corporate authorship; author institutional affiliation; author discipline affiliation; obsolescence of literature; and referencing in non-scientific literatures.

Minimum expectations in papers would include the provision of a theoretical framework, definition of terms, explication of assumptions, and a review of related research. Of course, it is anticipated that this issue of Library Trends will also stimulate a variety of ideas that could become the focus of student assignments.

A Syllabus for Teaching Bibliometrics

The appendix to this paper suggests tentative content and emphases for an MLS course, together with (currently) desirable readings. It is noted that few (if any) students will have the time to read everything listed, and so the onus is on the professor to map out a manageable program based on local institutional objectives and priorities. Introductory remarks are presented for each major segment of the proposed course in an attempt to identify progress and problems to date. The remarks might furnish a starting point for lectures, or they might be revised and distributed to students for reference.

The major course segments given in the syllabus are: (1) overview of the field, one unit; (2) theoretical framework, two units; (3) research traditions: laws and models, five units; (4) research traditions: empirical descriptions, five units; and (5) applications for professional practice, two units; for a total of fifteen units.
ALVIN SCHRADER

Future Prospects for Teaching Bibliometrics

The literature of bibliometrics is a rapidly growing one. In 1977 Voos estimated there were 1400-2400 publications on the subject from the nineteenth century to date. Pritchard published a 700-item interim bibliography on bibliometrics for the period 1881-1969, and announced in 1979 that he is compiling a far more extensive one of 3000-4000 items as a byproduct of a research degree. Hjerppe has published a bibliography of bibliometrics and citation indexing and analysis. This work indicates the growth of the literature and the international activity in the field. It also suggests the need by any professor teaching bibliometrics to keep abreast of new research and to be prepared to discard any of the above suggested readings as advances in theory and methodology are made.

In evaluating the literature of bibliometrics and in helping to shape future directions of bibliometric research, educators and researchers are encouraged to emphasize the following problem areas: (1) theoretical formulations to link social communication processes and cognitive structures in a field to its literature; (2) research into information exchange patterns, multiple and overlapping channels, and information demands; (3) citation behavior and citing theory; and (4) research into the properties of varying fields within science and social science, and between them and nonscience. Finally, it is suggested that less priority be placed on mathematical modeling with limited variables, and instead that more emphasis be directed to underlying multivariate conceptual dimensions in order to construct a more adequate theory of bibliometrics in the context of information transfer processes and systems.

References


Teaching Bibliometrics

19. Many of the suggested projects are from the list of assignments for Dr. L. Houser's bibliometrics course, University of Toronto, spring 1981.
21. Lotka, "Frequency Distribution."
26. The author wishes to thank Prof. L. Houser of the University of Toronto and Prof. A. Pratt of the University of Arizona, Tucson (formerly of Indiana University) for stimulating and supporting my interest in bibliometrics.

SUMMER 1981 159
ALVIN SCHRADER

Appendix

BIBLIOMETRICS COURSE SYLLABUS*

1. Overview of the Field (1 unit)

This unit focuses on terminology, major concepts and reviews of the literature.

Uncertainty about a variety of variables and their interconnections with respect to scientific literatures was the impetus for bibliometric study. Some of the initial questions were: Does the literature of a field represent the field? How does the growth of a literature relate to the growth of scientific knowledge? What are the essential characteristics constituting the structure of a literature? How do various literatures compare with respect to structure? Who are the producers of a literature? Who are its users? How are quantity and quality of literature production related? These and later, more complex questions have attracted the attention of increasing numbers of researchers and theoreticians in a wide spectrum of academic disciplines. Among current difficult problems are: the functions of referencing (intellectual property recognition, persuasion or window dressing); the relationship between the cognitive structure of a discipline and its social structure, particularly as manifested in communication and publishing patterns; and the theoretical validity of bibliometrics in scholarly nonscientific fields.

The rapidly advancing status of bibliometrics as a scholarly specialty is indicated by its large body of literature, now well over 2000 publications, by the recent appearance of at least three journals, and by the attendant review literature. Particularly exciting is the international makeup of the research front, comprising social scientists not only in the United States but also Russia, Europe and England. Although bibliometric study began with the literatures of the natural and biological sciences, social science literatures have also been examined bibliometrically from time to time. In addition, there have been a handful of attempts to apply the various techniques to some of the literatures of the humanities disciplines.

Although there does not appear to be a consensus in the literature on the use of the term bibliometrics, the various other descriptions represent subspecialty thrusts. Recently, for example, Narin (1976) introduced the concept of evaluative bibliometrics, which he defined as the quantitative measurement of the properties of a literature in order to evaluate scholarly activity in a field. In addition, there is the term scientometrics, the scientific analysis of science and science policy. The latter focus was embodied in the formation in late 1978 of Scientometrics; An International Journal for all Quantitative Aspects of the Science of Science and Science Policy. This is the second of three recent, relevant journals. The first was Social Studies of Science; An International Review of Research in the Social Dimensions of Science and Technology (earlier entitled Science Studies, from its inception in 1971 until the end of 1974). The third journal, although of very recent origin, shows promising relevance. It is entitled

*A reference to an author during discussion of a unit has been footnoted only if the reference does not appear in the accompanying list of readings.
Teaching Bibliometrics

Knowledge: Creation, Diffusion, Utilization, and is aimed at bringing together researchers, policy-makers, research and development managers, and other practitioners engaged in the process of knowledge development. Of course, there are also a number of journals relevant to bibliometrics within the history and philosophy of science in terms of theoretical implications, notably the British Journal for the History of Science. Another important indicator of bibliometric advance was the inauguration in 1975 of the Society for Social Studies of Science, colloquially known as the "4S," which was reported to have attracted over 500 members by the end of its first year.

A comprehensive review of the literature of bibliometrics was published by Narin and Moll (1977), and a survey of developments to date by Hjerpe. In addition, more than thirty doctoral dissertations and several monographs on various aspects of bibliometrics have been published; among the notable monographs are those by Price (1963, 1975), Narin (1976), Elkana (1978), Garfield (1979), and Garvey (1979). Two other monographs have attempted to present an integrative overview of bibliometrics, Donohue2 and Nicholas and Ritchie,3 but neither has proven satisfactory.4 The definitive text awaits an author.

Narin (1976) has mapped out three research fronts in the literature of bibliometrics (see table 1). They are: (1) the size of the scholarly enterprise; (2) the properties (i.e., structure) of the literature of each enterprise; and (3) the productivity of scholarly authors.

Size of scholarly enterprise is generally expressed in terms of national or international comparisons among literatures. Recently, attempts have been made to correlate scientific productivity of a given country as indicated by its scientific literature with national economic vitality. Such an index may become particularly meaningful to the evaluation of progress in underdeveloped and middle-power nations.

The structure of a literature is generally expressed in terms of relationships among individual publications or among a set of publications such as journal literature, in terms of links between researchers, or in terms of maps of disciplinary phenomena. These relationships and links and maps can be used to identify key events, advances and patterns of scholarly research. Newer work such as cocitation analysis and multidimensional scaling can be used for evaluative functions as well as description, in comparing productivity among authors, journals or organizational entities such as funding agencies, university departments, professional associations, or countries. Suggested readings for this unit follow.

Terminology:

SUMMER 1981 161
# TABLE 1
**Chronology of Major Contributors to the Development of Bibliometric Analyses of Scientific Literatures**

<table>
<thead>
<tr>
<th>Year</th>
<th>Size of Literature</th>
<th>Structure of Literature</th>
<th>Productivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1910</td>
<td>Cole and Eales</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1920</td>
<td>Hulme</td>
<td>Gross and Gross</td>
<td>Lotka</td>
</tr>
<tr>
<td>1930</td>
<td>Bradford</td>
<td>Gross and Gross</td>
<td>Lotka</td>
</tr>
<tr>
<td>1940</td>
<td>Wilson and Fred</td>
<td>Cason and Lubotsky</td>
<td></td>
</tr>
<tr>
<td>1940</td>
<td>Gosnell (Bradford)</td>
<td>Fussler (Zipf)</td>
<td>Lehman</td>
</tr>
<tr>
<td>1950</td>
<td>Daniel and Louttit</td>
<td>Garfield</td>
<td>Schockley</td>
</tr>
<tr>
<td>1960</td>
<td>Kessler</td>
<td>Xhighnesse and Osgood</td>
<td>Price</td>
</tr>
<tr>
<td>1970</td>
<td>Narin and Carpenter</td>
<td>Garfield and Cole</td>
<td></td>
</tr>
</tbody>
</table>

Source: Narin (1976), adapted and slightly expanded.

**Reviews of the literature:**

-----

**Texts:**


Teaching Bibliometrics


Pritchard (1972), who attempted to relate bibliometrics to the information transfer process, conceptualizing the flow of information through channels as analogous to a chemical or industrial process. Another is Meincke and Atherton (1976), who have introduced the difficult but interesting concept of knowledge space or scientific space, in which concepts, fields of knowledge, and information items in a retrieval system are likened to physical objects (such as atoms) that occupy multidimensional vector space.

However, while theoretical advances in the sociology of science have been spectacular, little progress has occurred in our understanding of the nature of theoretical properties of the vast array of subject literatures. For example, Peritz has argued, convincingly, that citation analysis cannot properly be applied to historical research because citations representing the source documents for history cannot be sorted out from citations representing ordinary references. This may well have been the difficulty in the analysis by Brace of citation patterns in graduate library school doctoral dissertations, a large proportion of which have always been historical research. The same validity problem arises with respect to citation analysis of literary criticism studies.

Theoretical uncertainty goes deeper than this, however, for what we really need to understand better is under what conditions a literature structure may be said to be isomorphic to the referencing behavior and norms of its producers. Scientific literature is assumed to be isomorphic, or more nearly isomorphic, to the referencing behavior of scientific authors because scientists produce knowledge by building on previous knowledge, and so they acknowledge the antecedent work, the intellectual property, of their colleagues. Thus, both the scientific advances and the citing may be regarded as cumulative. Garfield, Malin and Small (1978) suggest that citation linkages in science reflect both the cognitive structure and the social structure of a specialty; this argument has not yet been adequately elaborated for empirical testing, however.

Like this theoretical hypothesis, there are many other challenges awaiting bibliometric inquiry. Some of these are to produce adequate explanations of the following problems and phenomena: how progress in scientific knowledge can be objectively identified, and how such progress is reflected in the literature; how the social systems of science and nonscientific scholarship differ, and how they reflect differing communication patterns, differing referencing practices and norms, and differing publication practices; how patterns of information...
exchange activity are related to the processes of scientific research, discovery, dissemination, and utilization by scientists, and how these processes vary from discipline to discipline or perhaps even from specialty to specialty; how the nature of a research front should be determined (is it in the formal or informal communication domain, and if in the formal, is it more accurately described as a citation front, as Garvey (1979) has perceptively argued?); how the hardness-softness metaphor describing a continuum of scientific rigor can be either operationalized and tested, or abandoned; how the identification of a susceptible in the epidemic theory of information diffusion proposed by Goffman and Newill (1964) can be determined; how the nature of a citation can be defined (is one citation to a paper equivalent to multiple citations to the same paper?); how the nature of a reference is to be agreed upon (is a reference to a scientific paper the same as a reference in historical inquiry and in literary criticism?); how information transfer or information flow are to be treated; what the relationship is between information, knowledge, ideas, and data; and finally, how the dissemination of knowledge differs between the paper disciplines and the product disciplines (that is, between scientific and technological research activities), and between them and the secret disciplines of military and industrial inquiry. These are only some of the exciting theoretical problems before us. Suggested readings for this unit follow.

Readings:


Teaching Bibliometrics


3. Research Traditions: Laws and Models (5 units)

This section is prefaced by an introduction to logarithmic distributions and nonparametric statistical procedures. This is necessary because bibliometric data have been found to exhibit geometric or exponential properties of growth and decline, rather than arithmetic properties.

From the bibliometrics literature, there is a strong impression that two research traditions have developed, more or less independently though concurrently. The one tradition is characterized by investigation into distributional properties, typically culminating in the formulation of a statistical law or a mathematical model of the logarithmic variety. This tradition derives from Lotka, Bradford and Zipf, and is represented by such researchers as Bookstein, Brookes, Coile, Fairthorne, Goffman, Kendall, Leimkuhler, O’Neill, Pratt, Vickery, Vlachy, and Wilkinson.

The other research tradition is more strictly empirical, focusing on counts of data and on first-order relationships among sets of data such as cocitation mapping describes. Notable contributors in this tradition are Fussler, Garfield, Griffith, Kessler, Line, Mullins, Narin, Price, Sandison, and Small. In passing, it should be noted that the creation of Science Citation Index, Social Sciences Citation Index and Arts & Humanities Citation Index by the Institute for Scientific Information in Philadelphia have vastly accelerated the potential advance of knowledge through the empirical tradition.

Bibliometric measures in general focus less on the central tendency of a distribution of data and much more on the extremes which characterize the distribution. Also, bibliometric measures are based on the frequency ranking of data, in most cases. However, if the essential information in the data is to be preserved and evaluated, nonparametric statistical tests for rank-ordered data cannot be utilized because such tests do not adequately preserve the magnitude of differences between rankings. Other nonparametric approaches must be
ALVIN SCHRADER

devised, so that the typical high concentration of data in a relatively small proportion of the population can be represented.

There is still a great deal of investigation required into the underlying theoretical dimensions of the mathematical formulations expressed in Lotka's law, Bradford's law and Zipf's law. Various explanations to date have proposed a law of diminishing returns model, a cumulative or comparative advantage model issuing from the more generalized theory of stochastic processes, and an information theoretic model of the human mind. However, as Bookstein (1979) noted in a recent critique of the current views, these various models and laws all turn out to be mathematically identical, and this in itself is an interesting finding that invites investigation.

There is also a great deal of investigation required into methodological validity. Coile (1977) has documented several misuses of Lotka’s law, for example, and Wilkinson (1972) has pointed out that no two researchers have interpreted Bradford’s law in the same way. Some of the current questions are: whether these distributions are properly described as “laws” at all rather than simply probabilistic occurrences; whether Bradford’s law is reliable for small collections, what small means, and whether a collection can be one journal or whether a broad base of journals is required; whether Bradford’s law is biased toward journals that publish a large number of very short papers; whether sample size is a factor in making comparisons of scattering characteristics across fields; whether Bradford’s law can be explained as an artifact of journal editorial policy, as Fairthorne (1969) has speculated; and whether the performance of new journals, papers and authors can be predicted. Related issues are whether the investigation of one or two variables without a research hypothesis, as is the case with the empirical descriptions discovered by Bradford, Lotka and Zipf, constitutes an adequate basis for quantitative inquiry, and whether multivariate bibliometric analyses would be more fruitful. Suggested readings for this unit follow.

Logarithms:

Bradford and Zipf:

166

LIBRARY TRENDS
Teaching Bibliometrics


**Lotka:**


**Recent advances:**


"Cumulative Advantage Urn Games Explained: A Reply to Kantor." *Journal of the ASIS* 29(July 1978):204-06.


4. Research Traditions: Empirical Descriptions (5 units)

This section covers publication counting and citation analysis. Simple one-to-one citation links and the notion of bibliographic coupling were typical empirical approaches in the 1960s and before, but in the following decade the concept of cocitation clustering was invented and came to dominate the bibliometrics research front. The cocitation clustering technique has exciting potential for mapping the structure of scientific specialties and perhaps even entire fields of science, and for documenting changes and growth over time. Studies into the validity and limitations of citation analysis are also reviewed; contributions here are content analysis and typologies of citations, sometimes referred to as context analysis, and correlational analysis of citations with other quantitative and qualitative measures.
Scholarly norms of citing are complex and vary from field to field and from science to nonscience. Similarities in citing conventions between scientific literatures and humanities literatures are not adequately understood at all, but the social conventions determining citing behavior in a given field are crucial to theoretically valid characterizations of the structure of the field's literature.

The citing of antecedent research is a strong social norm among scientists and social scientists. Citation relationships are conceptualized as semantic relations between texts that constitute directed lines connecting later to earlier work. When these relations are graphed, they are said (borrowing from graph theory) to form a digraph. Such a digraph reflects semantic textual structures such that antecedent subject matter is linked to later subject matter. Citation analysis relies on the occurrence of the social norms of citing, but there are many other reasons for particular choices of prior authors and papers. As Lipetz (1965) and Weinstock (1974), among others, have noted, these choices could be motivated by any of the following: paying homage to pioneers; providing background reading; giving an example; modifying, correcting, criticizing, or refuting previous work; identifying the original publication of an eponymic concept or term such as Pareto's law; or window dressing. Refinements in citation analysis methodology are now being produced through contextual analysis of references. Also, studies have been undertaken in science to assess the correlation between citation data and peer judgments. Cole and Cole (1973) and Zuckerman (1977), among others, have demonstrated that straight citation counts are highly correlated with virtually every refined measure of research quality and other forms of scientific recognition, such as the Nobel prize and membership in a national academy of science.

Thus, although "errors" or deviations in citing behavior do occur, the accumulation of bibliographic links over hundreds or even thousands of acts of citing over time is seen to map out the cognitive domain of scientific knowledge in a given area; the self-correcting and cumulating nature of knowledge is a probabilistic process that sloughs off the errors or deviations and dead-end research programs. In effect, when an author cites he is classifying his own work with respect to the perceived domain of all prior scholarship.

What lends further credence to the validity of citation analysis, at least in science, is the consensus factor; that is, the journal-refereeing system requires a consensus among selected scholars on the worth of the work being submitted for publication, and one of the criteria for judging such worth is coherence with past research, presumably as represented by the researcher's choice of citations to antecedent work. However, it should also be noted that citation anomalies having a small effect on the average might have serious distorting effects in a particular instance, for example, anomalies such as obliteration, eponyms and highly unpopular claims like those of Arthur Jensen.

Thus, citing theory is in its infancy. Among the factors influencing the nature and frequency of citation are the following: the size of the field and number of authors in a field; the nature of the field, especially its degree of theoretical integration or codification; whether a field is a paper- or product-producer, and especially what proportion of a field may be said to be engaged in secret research, such as for military and industrial organizations; the age of a field; differing growth rates of fields; journal editorial policies, such as rates of publication, language of publication, length of articles; journal function (e.g.,
Teaching Bibliometrics

reporting research or current awareness; journal quality and prestige; author eminence; average number of references per journal article; the degree of anomalous citation behavior in a field; perceived social utility of the field and funding for research; rates of multiple versus single citation to a paper; rates of multiple versus single authorship; variability in quality and importance of papers; relationships between obsolescence and changes in journal size; and above all, differential reference functions and norms among the sciences, social sciences, technological fields, and the nonsciences. Suggested readings for this unit follow.

Citation analysis:


Context analysis:
Bertram, Shelia J.K. "The Relationship Between Intra-Document Citation Location and Citation Level." Ph.D. diss., University of Illinois at Urbana-Champaign, 1970.
Chubin, Daryl E., and Moitra, Soumyo D. "Content Analysis of References: Adjunct or Alternative to Citation Counting?" Social Studies of Science 5(1975):423-41.
Lipetz, Ben-Ami. "Improvement of the Selectivity of Citation Indexes to Science Literature through Inclusion of Citation Relationship Indicators." American Documentation 16(1965):81-90.
Murugesan, P., and Moravcsik, Michael J. "Variation of the Nature of Citation Measures with Journals and Scientific Specialties." Journal of the ASIS 29(May 1978):141-47.

5. Applications for Professional Practice (2 units)

There is a great deal of controversy about the appropriateness of bibliometric applications to practical problems. Some authors have argued that underlying theoretical explanations of the bibliometric distributions are too weak to guide information facility policy decisions, that bibliometric theory is not ready for practical application. Others have urged even greater application, particularly to library collection management. Several reviews have been published, notably those of Broadus (1977), Buckland (1978), Fitzgibbons (1980), and Lancaster (1977). Moll edited a special issue in 1978 of Collection Management devoted to bibliometrics in library collection management.

However, a number of major application problems have not been adequately addressed in the bibliometrics literature. First, most of the mathematical models which have been proposed are static models, i.e., they assume fixed economic conditions, for example, with respect to journal acquisitions costs versus interlibrary loan costs, fixed subject areas, fixed user interests and homogeneous information demands, and fixed information facility objectives and policies. Second, the models are simplistic and do not adequately reflect reality in that they assume—but are unable to demonstrate operationally—that user satisfaction can be defined and measured, and that individual user dissatisfaction is unimportant to the advance of scholarship. Third, the mathematical
models have weak explanatory power. They are unable, for example, to predict the performance of new journals, new researchers and new papers. Fourth, the variables in the models are only vaguely linked to sociological concepts. For example, citation analysis treats the formal communication process, while use and user studies concern demands on an information facility. Are identical or highly dissimilar processes and modes of social communication behavior thus being measured? How valid is the assumption that citations reflect information facility use patterns? Fifth, almost all information facility objectives and, in particular, collection policies are so unclearly expressed that they boil down to assertions that cannot be operationalized and tested. Fundamental concepts such as information need, user satisfaction, and even information facility use, are inadequately articulated. Until information facilities begin to support development inquiry on a grand scale, with funds for researchers rather than for computers and computer applications, progress in applying bibliometric theory will be very slow. Finally, almost all the models and bibliometric explanations to date have been focused on scientific journal literatures, scientific information facilities, and scientific researchers. More work is needed to determine what form practical applications should take in public and academic libraries as they are presently constituted, with amorphous, heterogeneous user populations exhibiting highly diversified demand patterns.

These are some of the difficult but challenging problems ahead. Suggested readings for this unit follow.

**Reviews of the literature:**


**Readings:**


ALVIN SCHRADER

"No-Growth Libraries and Citation Analysis; or, Pulling Weeds with ISI's Journal Citation Reports." Current Contents: Life Sciences 18(30 June 1975):5-8.


Notes


