Research evaluation and citation analysis: key issues and implications

Nancy K. Herther
University of Minnesota Libraries, Minneapolis, Minnesota, USA

Abstract

Purpose – Citation and ranking information are becoming key aspects of knowledge management in academic and research institutions. By examining changing user needs and products, this paper aims to encourage information professionals to better understand and manage these resources and better respond to user needs.

Design/methodology/approach – Literature reviews, work with faculty clients and initial product testing are used to present coherent information on the current climate and practice of competitive analysis of researchers and their institutions.

Findings – As more sources for citation information have become available – even many scholarly databases today offering cited reference data – the need to identify, access and manage these resources is becoming acute. Information professionals need to become more proactive in their strategies to support these applications and users.

Originality/value – This article builds on previous analyses of the roles and nature of citation analysis in research institutions and examines potential new roles and contributions that information professionals can take on to better serve their users.

Keywords Research, Higher education, Information management, Statistics

Paper type Conceptual paper

Introduction

In the 2006 report from the US Commission on the Future of Higher Education, the role of statistics was clear in terms of both evaluating where academe is today and planning for where it needs to be in the future (Commission on the Future of Higher Education, 2006). This and other recent reports discuss many critical, data-intensive issues facing academe and are changing societal assumptions about the role and nature of education:

- **Accountability**: to students, the institution, funding sources and to society.
- **Performance**: specific measures or expectations for student outcomes, value for the dollar, etc.
- **Mission**: scholarship of engagement, and other movements, seek to redefine the role of higher education in light of broader social or cultural needs bringing added dimensions to assessments.
- **Image**: the need to distinguish one institution’s ‘product’ from others.
- **Competition for research dollars**: from foundations as well as from international and non-‘academic’ candidates – requires that institutions include information on their competitive standings as a part of the application process.
These needs have led to the creation and use of a variety of statistical resources and databases on campuses. The collection of admissions information, for example, has grown substantially: “Until about five years ago, only a handful of colleges used sophisticated statistical formulas. But now there are few four-year institutions that do not.” (Farrell, 2006).

Sources of aggregated data, ranking US colleges, include the Center for Measuring University Performance at the University of Arizona (mup.asu.edu/), Academic Analytics (www.academicanalytics.com/), the National Research Council (www7.nationalacademies.org/resdoc/index.html) and the University of Florida’s Center for Studies in the Humanities and Social Sciences (mup.asu.edu/). In the UK, the Research Assessment Exercise (www.rae.ac.uk) and the Australian government-sponsored RQF (Research Quality Framework) are examples of national efforts to develop performance measurements for higher education. (Cheek et al., 2006) In the Australian framework, a series of measures are suggested including the number of articles published in high-quality journals, the number of highly cited articles published and the number of citations contained in articles from “high-quality journals” (Allen consulting group, 2005).

The US Commission on the Future of Higher Education has also entered the picture, proposing the establishment of a:

[...] database that would allow consumers to rank colleges based on variables of their choosing and intends to endorse a controversial plan to create another database to track the educational progress of every college student in the USA (Field, 2006).

Ranking of American institutions by popular media – US News and World Report (which began its ratings reports in 1983), Maclean’s, Business Week, Gorman Report, Washington Monthly and others – are widely read, but also widely criticized by the academic community. Commenting on the inclusion of “patriotism” in the Washington Monthly rankings, Rebecca Goldin commented:

No longer are academic excellence, the advancement of human knowledge and the preservation and proliferation of ideas sufficient reasons for a university to exist – or sufficient demonstration of benefit to the country” (Goldin, 2006).

The “audit culture” & tenure
Pressures on academe to provide explicit quality measures have led to what has been called an “audit culture.” (Cheek et al., 2006) Inherent in this accountability process is the evaluation of contributions by productivity, quality and impact. Productivity defined by some type of quantitative measure of works attributed to a given scholar or institution over some specified time period. Quality – more difficult to define – is most often linked to expert judgment by peers or independent experts. Impact, defined as the frequency that someone’s work is cited by other authors, is a reflection of the importance of some specific research to the overall growth of knowledge in a field.

As issues of performance and accountability have become particularly acute, the focus has intensified on the productivity and output of faculty. Historically, academic tenure has been used to promote and protect the academic freedom of faculty; helping institutions to attract and retain high quality faculty, providing certain specified employment conditions and guarantees, while outlining conditions and processes for potential termination (due to poor performance, etc). However, even tenure, itself, is undergoing change.
Academic institutions are using increasing numbers of non-tenure track appointments today to fill vacant positions, with the assumption that this allows them more flexibility in filling positions in a time of shifting economic conditions. A November 2005 survey in the Chronicle of Higher Education found, in fact, that “53 percent of college presidents agreed that tenure for faculty members should be replaced by a system of long-term contracts” (Fogg, 2005).

A recent task force from the Modern Language Association of America found that “fewer than 40 percent of the PhD recipients who make up the pool of applicants for tenure-track positions obtain such positions and go through the tenure process at the institutions where they are initially hired, and a somewhat larger number of modern language doctorate recipients – more than 40 percent – never obtain tenure-track appointments. In the aggregate, then, PhDs in the fields represented by the MLA appear to have about a 35 percent chance of getting tenure.” (Modern Language Association of America, 2006).

Guidelines and principles for tenure exist from accrediting agencies, scholarly organizations (such as AAUP) and professional associations. The 2007 National Research Council Assessment of Doctoral Research Programs (http://www.nationalacademies.org/nrc), for example, specifies that faculty complete a questionnaire, which asks for information on honors, publications, and other types of scholarly activity going back five years (for sciences and social sciences) or ten years (for the humanities). Although it is not required that faculty provide citation information as a part of their assessments, many faculty prefer to include this data in their reports.

In practice, tenure and review policies vary by college and department. (Colquhoun, 2003) Many in competitive or highly interdisciplinary areas, especially, cite a stronger need for corroborating data on the “quality” of citations to their work. (Wallin, 2005) In an area that appears to be shifting quickly, the available guidance from peers or institutions varies widely (Pfirman et al., 2005).

Citations & recognition: an academic tradition
Scholarly publication is a key component of the social system of science, having evolved over the centuries as a key aspect of scholarly research:

Only by publishing their work can scientists make their contribution (as the telling word has it) and only when it thus becomes part of the public domain of science can they truly lay claim to it as theirs. For that claim resides only in the recognition of the source of the contribution by peers (Merton, 1979).

Communication in academe has been called “the essence of science.” (Garvey, 1979) A. J. Meadows referred to communication as being as “vital for research as the actual investigation itself” (Meadows, 1974). Jonathan R. Cole and Stephen Cole referred to communication as “the nervous system of science – the system that receives and transmits stimuli to its various parts.” (Cole and Cole, 1973) A key component of formal, academic communication, through the publication process, is the use of references and citations to prior works, to both frame new contributions and to give fair credit for the role of other researchers in the evolution of theory and research.

Despite the enormous changes in technology, and the increasing rate of scientific knowledge in the past hundred years, these principles have remained constant. All of this is reinforced by the reward system of science. Since recognition by qualified peers
is the basic form of extrinsic reward (all other extrinsic rewards deriving from it)” which includes tenure, promotion, merit pay, etc (Merton, 1995).

**Assessment tools**
Critics often cite the “bottom line” nature of faculty assessments – metrics such as the number of hits to some particular article, etc. – instead of focusing on the quality or future research value of their efforts. (Schuster and Finkelstein, 2006) Today, there are also many instances of outside, political influences on the process of tenure (Gravos, 2006; Fogg, 2006; Smallwood, 2005). Although tenure processes vary, there are strong commonalities. In some fields, vita entries are checked against listings in core scholarly databases as a method to indicate importance. Scholarly, databases are sometimes viewed as quality filters since the articles are taken from carefully chosen refereed journals, recognized by the appropriate core associations as relevant to the discipline. This “gatekeeper” role of scholarly databases, however, has received little study in either the scholarly or information/library literature (Horan and Erickson, 1991; Wallin, 2005).

Schemes to deal with issues such as multiple authorship, usage of articles, establishing links to cohorts, etc. have also been discussed as potential methods of measurement. (Gelman and Gibelman, 1999) Outside experts may be sought to provide independent judgments and opinions on candidates as well. One of the most popular assessment methods, however, remains the use of bibliometric data from citation databases. In Finland, for example, “government funding for university hospitals is partly based on publication points, with a sliding scale corresponding to the impact factor of the journals in which researchers publish their work” (Adam, 2002). The fact that this has happened is somewhat logical due to the nature of the bibliometrics which makes it easy to glean statistics from citation databases.

**The citation indexes & quality measurement**
Thomson Scientific’s (formerly ISI) citation databases were developed by Eugene Garfield as a tool to measure trends in science, as reflected in formal publication, and for tracking these changes, connections and developments over time. As Garfield described it, “the system would provide a complete listing, for the publications covered, of all the original articles that had referred to the article in question.” (Garfield, 1955) The significant phrase here is for the publications covered. Garfield’s purpose was not to create a system that would record all potential sources for citations – books, for one example, were never included despite their value in many areas of the sciences (and even more so in the social sciences, arts and humanities).

Garfield has often warned against the use of citation analysis and Journal Impact Factors for individual assessments; however, it may be fair to consider the PR for both the Journal Citation Reports (JCR) and Essential Science Indicators (ESI) as promoting these applications. Thomson Scientific promotes their ESI product as an: “in-depth analytical tool (that) offers data for ranking scientists, institutions, countries, and journals.” (Thomson Scientific, 2008a). JCR is described as “quantifiable statistical data that provides a systematic, objective way to evaluate the world’s leading journals and their impact and influence in the global research community.” (Thomson Scientific, 2008b) Although articles critical of these tools and their applications abound in academic presses, they continue to be widely used, probably due to the established reputation of the Thomson Scientific citation indexes in academe.
Using either JCR or ESI is restrictive for most researchers due to the nature and structure of each database. The rather artificial science and social science divides and the lack of any treatment of the arts and humanities are just two of the most critical problems faced by researchers, often leaving them to find their own methods to download information and create datasets to perform advanced citation research.

**A formidable task for the individual**

Citation counting remains the most common method of individual assessment today. In 1998, Garfield himself estimated that citation data and analysis “have been used in the USA to evaluate 5,000 departments at the leading universities” (Garfield, 1998). Lowell Hargens and Howard Schuman surveyed a sampling of departments of sociology and biochemistry in 1990 finding that 35 percent of biochemists and 60 percent of sociologists reported that their departments had “ever used citation counts in decisions about hiring, promotions or salaries” (Hargens and Schuman, 1990).

The degree of manipulation or analysis required in these assessments can vary from a mere listing of the number of citations for each published work to the manipulation of this data, adding other elements or analysis, to determine the relative value or importance of research endeavors. (Kostoff, 2002). Many institutions have chosen to use Thomson Scientific’s Journal Impact Factor as a surrogate for a more indepth, complex analysis, leading to what has been called “the impact factor industry” (Bensman, 2007). Anthony van Raan, director of Leiden University’s Centre for Science and Technology Studies has called the Journal Impact Factors “the poor man’s citation analysis” with good cause (Adam, 2002).

These controversial numbers are certainly easy to pull out of the JCR database, but the numbers themselves are highly controversial and widely disputed for this purpose:

Important papers, the argument goes, will be cited more frequently. As a general rule, that is a reasonable assumption. But apply it blindly, without regard to the quality and limitations of the raw data, and the conclusions you draw may be far from reasonable (Bensman, 2007).

Given the lack of comprehensive data, and the difficulty of defining qualitative objectives at a level of specificity needed to perform this type of research, the search for unbiased information to judge or value research continues.

**Alternative sources of citation data**

Today, an increasingly competitive marketplace is developing for citation data:

- **CiteSeer**: (http://citeseer.ist.psu.edu/) which currently focuses on computer/information science.
- **Scopus**: (www.scopus.com/scopus/home.url) the Elsevier product which covers approximately 15,000 journals from 1996.
- **Google Scholar**: (http://scholar.google.com/) which attempts to cover anything available on the Web. Noted for its innovation, the results are widely scattered, lacking authority control for basic data elements and including no explanations on how citation rates are calculated, etc.
- **SciFinder Scholar**: (www.cas.org/SCIFINDER/SCHOLAR/) which covers over 9,500 currently published journals (1999 to the present) and patent information
from more than 50 active patent issuing authorities, focuses on chemistry &
medicine.

- **Faculty of 1,000:** (www.facultyof1000.com) for Biology and Medicine which
  include about 5,000 “leading researchers and clinicians share their expert
  opinions by highlighting and evaluating the most important articles in biology
  and medicine.”

- **Scholarly databases:** including those from Cambridge Scientific, EBSCO and
  other publishers, are increasingly providing citation information from their core
  journals and adding these to the reference information for particular articles.

None of these products, by themselves, are necessarily replacements to the Thomson
Scientific databases, but clearly more data sources for citation information are reaching
the market.

**Quantity & quality: the case of Elsevier’s Scopus**
The rise of more sources of citations has largely taken place without much oversight
from the library/information science community. Has our need for numbers caused
institutions to rush into acquiring and offering resources of unstudied value? Google
Scholar’s methods are treated as proprietary information, making a good evaluation
nearly impossible. With the noble goal “to be the most complete and comprehensive
resource for all research literature in science, social science, technology and medicine,”
Elsevier released Scopus in 2004. Currently the database includes 15,000 peer-reviewed
journals, mostly from the sciences. We have yet to see detailed reviews of the
database’s content coverage.

If you compare the number of titles, for example, in the social sciences and arts and
humanities between the Web of Science (2,840 titles) and Scopus (2,850 titles), you find
comparable numbers. In Web of Science, however, all titles are given cover-to-cover
indexing, while Scopus excludes certain types of materials (largely obituaries and
media reviews) from their coverage. The social sciences category, as defined in Scopus,
includes the fields of: Arts & humanities, psychology, decision sciences, econometrics,
economics, finance, business, management, accounting, and “social sciences.”
Since Web of Science includes psychology and operations/management sciences in
the Science category, it is fair to assume that Scopus provides far less coverage to the
social sciences than does Web of Science.

In a cursory sort through their title lists done in 2007, the present author found only 191
titles under Classification Primary (Level 1) for arts and humanities. Of the 191 titles, there
was an 80.1 percent overlap with Web of Science. Many of the titles, exclusively in Scopus,
are ones that may not be considered core to many people in these fields, such as *Eotovas
Lorand Tudomanyengyetem*, and *Anatolia*. Furthermore, the dates of coverage in Scopus
vary widely, with many gaps in coverage. *American Antiquity*, for example is listed as

Clearly more, detailed research on Scopus and other emerging sources of citation
data is needed. However, will more data solve users’ needs or address the issue of
quality assessment?
Role of the library
Since research databases are generally provided by libraries for their organizations, issues of research design and data sources are critical for academic librarians and information professionals working to provide the best available resources to serve user needs. Citation databases allow for a level of traditional bibliographic retrieval and citation information for known papers or researchers. However, increasingly assessment research requires comprehensive information on citedness and contribution – something that no individual tool can provide today. Given the nature of such a global search – for citations or mentions in published works, acknowledgements and other potential indicators of quality – finding information has become very complex. This, however, provides a unique opportunity for librarians in these evaluation processes.

Even if librarians may understand the existing tools, do we adequately understand the complicated context for these research efforts? For tenure review, in particular, these users are engaging in a very personal research effort, that is often neither well-defined nor clearly designed. In effect, these efforts are often more forms of marketing oneself or one’s institution rather than a traditional, impartial research experiment. The questions that are often the hardest to answer for these clients are: “How am I doing?” or “Is this enough?” Whether for self-evaluation or meeting the expectations of someone’s department, college, university or funding agency, these subjective questions are complex and any effort to approach them is tentative, at best: Are 50 citations “good enough?” What about 100? 1,000? 10,000? Working on evaluation issues involves far more than database searching skills or access to available resources. Understanding and interpreting existing collegial or professional/association guidelines is critical to the task as well.

Long range issues
The increasingly multidisciplinary nature of research today complicates efforts to establish value in the traditional frameworks of most databases. In 2005, the University of Minnesota Libraries received a grant from the Andrew W. Mellon Foundation to develop a multi-dimensional model for assessing support for scholarship and research in the context of a large research campus. As a part of the study, faculty in 16 departments were surveyed about their research. One compelling finding was that over 90 percent of the surveyed faculty considered their work to be multidisciplinary or interdisciplinary (McCready, 2006). For many, “cutting edge” research may also be more difficult to defend by traditional definitions or standards of quality.

In the past 20 years there have been other significant changes: In database access and use, due to the rise of the Web; the growth in both non-traditional publishing and diversified scholarly publishing platforms; and increasing collaboration within and outside the institution. Research is needed not only to reassess the needs of researchers and the support that is required, but also to gauge the impact of these changes on academic work life.

Today academe seems poised to address some of these critical issues:

Part of what we are seeing is a tension between structural and interpretive approaches and between quantitative and qualitative methods [...] We view these developments as
complementary, for each challenges the other. Models must be explained, and theories must be validated (Borgman and Furner, 2002).

Librarians also need to address metrics such as Journal Impact Factors – their nature, development and critics – with our users:

How can a score count for so much when it is understood by so few and its value is so uncertain? In defence, worshippers of impact factors say we have no better alternative. Isn’t it time for the scientific community to find one?” (Abbasi, 2004).

In 2005, Jorge Hirsch, University of California at San Diego physicist, suggested a new type of measure for individual contribution which has been called the h-index:

A scientist has index h if h of his/her $N_p$ papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each. (Hirsch, 2005).

Hirsch believes this to be “a useful yardstick to compare different individuals competing for the same resource when an important evaluation criterion is scientific achievement, in an unbiased way” (Hirsch, 2005). Early research is finding this to be a worthwhile, potentially more effective “index over an index based on total citations and total publications” (Vanclay, 2007).

In 2006, an international group of educators and administrators met to develop the “Berlin Principles on Ranking of Higher Education Institutions,” a set of voluntary guidelines developed as “the beginnings of a self-regulatory process” for academe. (Bollag, 2006) Efforts such as these may provide a better framework for understanding the factors that can and should be examined in future academic reviews. This information would be critical for librarians in the evaluation and selection of appropriate citation databases and services.

**Redefining the role of the library**

Although the Thomson Scientific databases are not only sources of citations for secondary literature, but also they remain the standard source for most academic fields. Thomson Scientific hasn’t helped the situation in the face of competitive challenges. Their help pages offer no in-depth coverage of these types of applications. Scopus’ coverage is equally problematic. As commercial entities, these companies are perhaps less interested in covering aspects of the structure and contents of their databases that may result in less reliance on their products. However, librarians have a fiduciary responsibility to their institutions and users to provide this type of information for all types of citation products that they offer users.

The traditional role that librarians have played with citation information has been similar to the support provided for other types of databases or reference needs: We have focused on hands-on searching instruction, through workshops or other means, as well as offering assistance at reference desks or on request to users. Many library web sites provide information on citation workshops; however, most appear to emphasize either basic searching techniques or alert features. Although useful, this level of information is not adequate for the types of evaluative applications that currently perplex, confound and frustrate the academic community.

If seen as a spectrum of potential roles from reactive to highly engaged, librarians have generally settled for general, introductory levels of service in this area, rather than exploring how we might fulfill some of our users’ more complex needs and interests.
Change would require that we move to a position of advocacy and consultation – and potentially even working as innovators as new systems and theories are tested and developed which might better meet these deeper information needs of our organizations:

- Participating in studies of the reliability of Journal Impact Factors, h-index and other indicators of quality, especially in areas of emerging or interdisciplinary research.
- Examining the journal contents of the citation databases, scholarly databases and other products, in various subject areas – and especially in newer, growing, multidisciplinary areas that we face – for comprehensiveness and breadth of coverage.
- Comparing and evaluating the various potential sources for citation and other recognition data (such as acknowledgements) that might be useful in these evaluations.

Academic librarians need, also, to reach out to our users – especially tenure committees and administrators – to provide our unique perspectives in such areas as:

- Researching the meaning and process of determining Journal Impact Factors, the h-index and other measures; encouraging the better understanding of various methods and definitions of “impact.”
- Exploring the limitations of existing data and the nascent growth of newer tools, educating our users, not only on the available databases, but how they can be best used.
- Linking available resources to existing models and research efforts.

Working with our vendors, librarians are in a unique position to participate in the design of future products as well as to advocate for the interests of our communities – participating, with unique skills and perspectives, as agents for change – on such issues as:

- Enhancing data integration features – such as the number of records downloadable, and easier porting of data to spreadsheets, statistical packages, relational databases and other products.
- Improving search and retrieval features that more easily allow for identifying related literatures, cohorts and other key relationships.
- Providing clearer information on the contents and methods used by each vendor in creating their databases and assigning values.
- Developing tools to aid in the manipulation of citation data, especially in integrating data from multiple sources, by establishing common field structures, etc.

By participating in these ways, we would provide not only a better service for our communities, but also add significantly to the perception and role of the library in this critical aspect of knowledge management for our institutions.
Conclusion
The past 20 years has seen many, significant changes in the ways in which research is perceived, funded and rewarded. The use of citation metrics has grown to the point of being an industry of its own today; dominated by such powerhouses as Google, Elsevier and Thomson Scientific. These newer applications of citation information require new and different efforts on the part of research libraries and their staffs to provide the best and most comprehensive information available. As arbiters of information, librarians and other search professionals need to re-evaluate their role in this process working with vendors on product development as well as with research organizations as they work to better define and articulate research value for the future.

References


Further reading


About the author
Nancy K. Herther currently serves as the Anthropology/Sociology Librarian at the University of Minnesota Libraries. She received her MLS from the University of Minnesota with a minor in Adult and Continuing Education. She is a full, elected member of Sigma Xi, the Scientific Research Society, and is a former columnist and editor. In 2006 she was awarded a quarter research leave from the University to study the citation indexes from Thomson Scientific/ISI and other digital sources. She has served on the Editorial Advisory Board for The Electronic Library since 2003. Nancy K. Herther can be contacted at: n-hert@umn.edu

To purchase reprints of this article please e-mail: reprints@emeraldinsight.com
Or visit our web site for further details: www.emeraldinsight.com/reprints