Citation Analysis and Its Potential In Theological Libraries

Abstract

Citation analysis has played a critical role in the hard and social sciences for decades. Where has it been in the humanities? This essay provides a brief introduction, giving an overview of citation analysis, its critical role in scholarship, a look at how it can play a role in theology, religion, and philosophy, and some suggestions as to roles that citation analysis can play in the future of these areas of study.

Introduction

Citation analysis (CA) can be defined as the practice of calculating and analyzing the number of times a work of scholarship has been cited by others. Traditionally, CA has been performed on individual units of scholarly production — a journal title, journal article, or book. But more recently many other artifacts, including white papers, patents, and data sets, in a variety of formats, are also being counted. Librarians use CA to extrapolate the value of highly cited items in making collection development decisions, in helping scholars assess the impact of their scholarly work,[1] and in providing publication guidance, reasoning that the most important journals in a field are those that contain the most citations elsewhere (although, as we will see, CA faces ongoing practical challenges and theoretical critiques).

Originating as a form of bibliometrics in the context of the paper-based world of the mid-twentieth century, CA was given a conceptual and practical boost during the 1980s to mid-2000s by the development of huge online commercial indexers. These made it possible to leverage technology to track more expansively citations across publications in various scholarly fields and sub-fields, particularly in the sciences and social sciences.

For many years, the primary marker of scholarly metrics was the "Journal Impact Factor,"[2] a proprietary CA measurement of a journal's influence determined by citations in other scholarly journals, and available through subscription for-profit databases like (what is now known as) Thomson Reuters's product Web of Science.

Additionally, librarians regularly use forms of CA to undertake projects such as assessing local collections and communities of practice. For example, at Princeton Theological Seminary Library, one annual project is a CA of doctoral dissertations and faculty publications from the preceding year. Citations from these works are analyzed by number, resource type, and accessibility. Results are used to gauge a variety of considerations, including the relationships between faculty and student publication, and resource use patterns in research analysis and collection evaluation. The value of this snapshot for the library is that it provides a way to keep track of how the collection is being used in practice and how patrons are using or creating other kinds of resources.[3] Over several years, these patterns can grow in value.

---

[1] Ruth Lewis, Cathy C. Sarli, and Amy M. Suiter, *SPEC Kit 346: Scholarly Output Assessment Activities*, Association of Research Libraries (May 2015). http://publications.arl.org/Scholarly-Output-Assessment-SPEC-Kit-346/.

[2] Eugene Garfield, "The Evolution of the Science Citation Index," *International Microbiology* 10 (2007): 65-69.

[3] "Library Collection in Support of the Curriculum and Specialized Research" (2013 and 2014), *Measuring Our Performance*, Princeton Theological Seminary Library, accessed July 10, 2015, http://www.ptsem.edu/library/about/assessment/reports/. The two most recent years of the citation analysis study (2013 and 2014) are available in reports on the library website.

Jenifer Gundry is Director of Collections, Preservation and Assessment at Princeton Seminary Library, Princeton, NJ; Tim Senapatiratne is Reference Librarian at Bethel Seminary, St. Paul, MN; Garrett Trott is Reference/Instruction Librarian at Corban University, Salem, OR.

Changes in Scholarly Metrics Since 2006

Since the mid-2000s, CA has become more sophisticated, within a proliferation of scholarly metrics broadly. Changes in publishing workflows, debates about the objectivity and validity of traditional peer-reviewed networks, expanded acceptance of a variety of academic artifacts (beyond books and journals), and huge shifts in scholarly and social communication tools, patterns, and methods have contributed to the slow but steady growth of an expanded set of metrics, including article and scholar-level metrics, and a set of alternative metrics (a.k.a. "altmetrics") that focus on web and social media impact.[4] These changes coincide with new pressures in the academy as a whole for greater accountability, evidence of impact, and data-driven indicators.

"What is measured" has changed dramatically. Moving beyond books and journals, there are now metrics for assessing a wide array of scholarly outputs — articles, white papers, presentations, websites, digital objects, datasets, coding, and even individual scholars, projects, departments, or schools.

"What is counted" has changed as well. Though scholarly journal and book citations continue to be important measurements, there are now metrics that track references in blogs, news media, policy documents, Tweets, Facebook likes, LinkedIn, Wikipedia, Pinterest, online citation managers, social bookmarking sites (Zotero, Mendeley, CiteULike), and many others. In short, scholarly metrics are now much broader, more granular, and more immediate.[5] They also help to advance the concept of how to measure the value of scholarly work, as new metrics allow for emphasis on scholarly visibility not only among experts but within the wider public as well.

In some ways, "altmetrics" conceptually underscores the continued usefulness of CA, while it expands the foundation and practice of analysis. Though it originated in an earlier research and publication context, CA continues to be an important mechanism in assessing the quality and impact of a scholarly artifact.

Challenges to Citation Analysis in Religion, Theology, and Other Humanities Fields

Despite technological advances, the practice of CA in the Humanities broadly, and in the interrelated fields of theology, religion, and philosophy more narrowly, currently faces two significant challenges.

First, theology, religion, and philosophy are not adequately supported by the major tools used for CA at present. The major indexers and databases such Web of Science and Elsevier's Scopus that provide citation metrics have grown up around the sciences. Naturally, given their client base, these providers have pursued the indexing of publications in those fields. Both platforms do provide some content in theology, religion, and philosophy, generally folded into general Humanities categories, making searching difficult. For example, as of July 2015, Scopus[6] contains 3,198 peer-reviewed journal titles in its Arts and Humanities category amongst its 21,000+ total peer reviewed titles. Web of Science's[7] 12,000 total journal titles can be searched by category and title. Its Arts & Humanities Citation Index is slightly more browser friendly than Scopus, subdividing its Humanities offerings in categories such as Religion (138 titles), Philosophy (179 titles), and others of potential interest, including Archeology, Classics, Folklore, Medieval & Renaissance Studies, etc.

In either case, it is clear that titles in fields of primary interest to theological libraries are not significantly represented in these products. There is some gradual improvement in more general databases such as EBSCO, Google Scholar

---

[4] Jason Priem, Dario Taraborelli, Paul Groth, and Cameron Neylon, "Altmetrics: A Manifesto" (October 26, 2010), http://altmetrics.org/manifesto.

[5] For example, PlumX, Altmetric, Impact Story, Google Scholar, and others. For an excellent overview of the status of metrics currently, see Robin Chin Roemer and Rachel Borchardt, *Meaningful Metrics: A 21st Century Librarian's Guide to Bibliometrics, Altmetrics, and Research Impact* (Chicago: Association of College and Research Libraries), 2015.

[6] "Content," Elsevier, accessed July 10, 2015, http://www.elsevier.com/solutions/scopus/content. Scopus content spreadsheets can be downloaded and searched through their Journal Title Lists and Book Title List here.

[7] "Master Journal List," Intellectual Property & Science, Thomson Reuters, accessed July 10, 2015, http://ip-science.thomsonreuters.com/mjl/. Web of Science content can be searched via the Master Journal List here.

Citations, and others to begin incorporating citation metrics across disciplines including the Humanities, but growth has been slow.

The comparative absence of theology, religion, and philosophy titles in the major indexes has several implications. It greatly hampers wide-scale CA; without major databases to do the heavy lifting of indexing, CA is very time consuming, as librarians and scholars must undertake the painstaking work of counting and cross-referencing works that may not be included in the major indexes. Also, the absence of this scholarship from the major indexers artificially limits the visibility and impact of these related disciplines in the work of other fields, and makes cross-discipline analysis difficult.

The second challenge that complicates bringing this material into existing CA technological platforms is rooted in the inherent nature of scholarship in religion, theology, and philosophy. Not only are the publishing patterns of the disciplines very different from the sciences (particularly, much longer publication cycles and a wide proliferation of important but small journals), but the texts of our disciplines in theological librarianship are very different. For example, they share a wide set of ancient primary texts, and millennia worth of translations across both dead and living languages that would inevitably have an impact on how citations are referenced and counted. Though the sciences, too, deal with some ancient and historical texts, this has not been a focus of the indexers to date; with religion, theology, and philosophy, these texts are core to everyday contemporary practice in these fields.

Response to Citation Analysis Challenges

As noted previously, CA still bears weight in multiple disciplines.[8] However, there is also much questioning of its value,[9] principally regarding whether or not citations are a useful means of assessing the scholarly nature of an article.[10]

In response to this, Zhang, Ding, and Milojević have integrated content analysis into CA, creating a hybrid approach referred to as *citation content analysis* (CCA).[11] CA, historically, has only considered the number of times a resource or author has been cited. This approach has generated some skepticism because it tends to carry the built-in assumption that every citation bears the same weight. "Content Analysis" is a broad term referring to the study of the text. While it does carry differing weights in distinct fields of study, in this context it is looking at the contextual value of a citation. However, CCA *combines* these two means of analysis by looking not only at the number of citations but assessing the *contextual value of each*.

Zhang, Ding, and Milojević created a matrix with twelve measurements to measure CCA: eight syntactic (i.e., looking at the particular elements of that citation) and four semantic (i.e., looking at the value suggested by the context of that

---

8 Johan Bollen et al., "Toward Alternative Metrics of Journal Impact: A Comparison of Download and Citation Data," *Information Processing & Management* 41, no. 6 (2005): 1419–40; Amy E. Dunbar and David P. Weber, "What Influences Accounting Research? A Citations-Based Analysis," *Issues in Accounting Education* 29, no. 1 (2014): 1–60; Dominick L. Frosch et al., "Assessing the Scholarly Impact of Health Psychology: A Citation Analysis of Articles Published from 1993 to 2003," *Health Psychology* 29, no. 5 (2010): 555–62; Dangzhi Zhao and Andreas Strotmann, "Counting First, Last, or All Authors in Citation Analysis: A Comprehensive Comparison in the Highly Collaborative Stem Cell Research Field," *Journal of the American Society for Information Science and Technology* 62, no. 4 (April 1, 2011): 654–76.

9 Michael H. MacRoberts and Barbara R. MacRoberts, "Problems of Citation Analysis: A Critical Review," *Journal of the American Society for Information Science* 40, no. 5 (September 1, 1989): 342–49; Per O. Seglen, "Citation Rates and Journal Impact Factors Are Not Suitable for Evaluation of Research," *Acta Orthopaedica* 69, no. 3 (1998): 224–29; Per O. Seglen, "Why the Impact Factor of Journals Should Not Be Used for Evaluating Research," *BMJ* 314, no. 7079 (1997): 497; P. O. Seglen, "Citations and Journal Impact Factors: Questionable Indicators of Research Quality," *Allergy* 52, no. 11 (1997): 1050–56; Per O. Seglen, "Causal Relationship between Article Citedness and Journal Impact," *Journal of the American Society for Information Science* 45, no. 1 (1994): 1–11.

10 Dirk Schoonbaert and Gilbert Roelants, "Citation Analysis for Measuring the Value of Scientific Publications: Quality Assessment Tool or Comedy of Errors?" *Tropical Medicine & International Health* 1, no. 6 (1996): 739–52.

11 Guo Zhang, Ying Ding, and Staša Milojević, "Citation Content Analysis (cca): A Framework for Syntactic and Semantic Analysis of Citation Content," *Journal of the American Society for Information Science and Technology* 64, no. 7 (2013): 1490–1503.

citation). One of these measurements stands out as having potential value to CA in many disciplines, but particularly theology, religion, and philosophy: disposition of citation.[12]

The disposition of a citation asks the question "what value does the citation carry?" Zhang, Ding, and Milojević suggest four options: positive, negative, mixed, and neutral.[13] A positive disposition would suggest that in this context, the author is favoring the work being cited, whereas a negative disposition suggests that the author is critiquing or disagreeing with the work that is being cited. A mixed disposition is a context where the author expresses mixed feelings about a work, the context does not strongly support the work, nor is the author adamantly opposed to it. In a context where an author simply makes reference to a work without necessarily inferring his or her opinion on it, the citation would likely be classified as having a neutral disposition.

This approach has the potential to enhance greatly the value and trustworthiness of CA. If a vast majority of cited articles are being used in a manner that reflects a positive disposition, it can be said that CA accurately reflects the nature of research and scholarship for that discipline. Alternately, in the event that the results are a combination of positive, negative, mixed, and neutral, one could look at journals cited with a positive disposition and arrive at CA data based on those indicators. Unless the citations overall bear a positive semantic value, using citation data without discretion would be of very limited benefit. Analysis based upon CCA could provide a more accurate list of journals that are valued in the discipline.

As a test case in disciplines familiar to *Theological Librarianship* readers, utilizing a revision of Zhang, Ding, and Milojević's matrix, we have analyzed citations in four issues of the *Journal of the Evangelical Theological Society* (*JETS*). While it provides intriguing data, it is a fairly small sample and therefore it is difficult to offer anything but insights and suggestions.

As noted above, a major question regarding the value of CA is the fact that authors use citations for a variety of purposes: not every work cited by an author necessarily means that the authors endorses that work. As Zhang, Ding, and Milojević's matrix assesses this, how does one year of *JETS* citations look when analyzed through this matrix?

There were 2,772 items cited in four issues (54:1-4) of *JETS*. This project discovered that while not all citations are ranked "positive," a vast majority of them (75 percent) are. The remaining citations are broken up into the categories of neutral (11 percent), mixed (6 percent) and negative (8 percent). With a vast majority of citations trending strongly in the direction of favorable, this would suggest that utilizing traditional CA (i.e., simply tracking cited journals without their semantic value) has the potential to bear some value in that discipline, specifically showing what resources scholars are basing their own research upon and thus bear value for that discipline.

However, do the data bear the same findings when only journal articles are assessed? Of the 2,772 citations, only 436 of them (16 percent) were journal articles. When looking at journal articles, the data indicates the following: 82 percent of journal articles cited were ranked as "positive." The remaining 18 percent of citations were broken up into the three categories of neutral (7 percent), mixed (6 percent), and negative (5 percent).

While the data sample is small and thus inconclusive, it does suggest some answers and a few hints as to the role of CA in theology, religion, and philosophy. First, while there has been some work on CA in relation to theology, religion, and philosophy[14] their scarcity is notable. Perhaps the fact that such a small percentage of resources cited in *JETS* were journal articles (16 percent) is evidence of the fact that journal articles do not play as critical of a role in the disciplines of theology, religion, and philosophy as they do in others. Thus, the relatively infrequent use of journal articles for citation

---

[12] Ibid.

[13] Ibid.

[14] Jenifer Gundry, "Library Collection in Support of the Curriculum and Specialized Research," *Library Reports and Outcomes* (Princeton, NJ: Princeton Theological Seminary, 2015); Timothy Senapatiratne, "A Citation Analysis of Ecclesiastes Scholarship: A Test Case Using Citation Analysis in Biblical Studies," *Advances in the Study of Information and Religion* 4, no. 1 (2014), http://digitalcommons.kent.edu/asir/vol4/iss1/2; Garrett Trott, "Blind Men and Elephants: A Multi-Method Analysis of Periodical Literature," *The Christian Librarian* 54, no. 2 (2011): 40–48.

in this sample may suggest why CA has thus far not played as important of a role in theology, religion, and philosophy as it has in other disciplines, such as the hard sciences.

As noted above, citations are used for multiple purposes and are not always used to provide a citation to support a scholar's work. However, the data shows that a vast majority of citations (75 percent) were used for exactly that. When only journal articles are measured, the percentage increases so that 82 percent of citations were used in a positive manner. This data suggest that scholars tend to use citations in a positive manner, and even more so when using journal articles as sources for citation.

There are distinctions between CCA and CA. If an author refers to two items in his article and one of these is cited once while the second is referred to ten times, CA tends to give them identical weighting, whereas CCA does not. This raises the question of whether or not CA is a valid means of assessment. However, the strong correlation (0.80) of the two data sets (one utilizing CCA and the other utilizing CA) from issues 54:1-4 of *JETS* is notable and suggests that CA not only bears weight because of the very strong tendency of citations coming from journal articles to be positive in disposition, but also the strong correlation between the data sets generated from CCA and CA.

Further exploration, based on a much broader body of literature, will be required to verify whether the trends suggested here are consistent. But again, this small sample at least suggests that CA has continued value, specifically for following scholarly research trends in the areas of theology, religion, and philosophy.

Future Considerations for Citation Analysis in Biblical and Theological Studies

What might this tell us regarding the future of CA in the fields of theology, religion, and philosophy? One major issue that needs to be considered is the question of the categorization of CA studies. As any religious studies scholar knows, the histories of theology, religion, and philosophy as disciplines are long and complicated. In light of this, CA studies face the problematic challenge that the multitude of sub-disciplines within the field make the topic extraordinarily complex. These categorizations are especially problematic within the humanities in general, and especially within religious studies where interdisciplinary work is commonplace. This can, at best, make conclusive findings from CA studies elusive, since a specific CA study will normally tend to be very narrow. At worst, this could create situations where a single CA results in data that are badly skewed by the narrowness of the study being extrapolated to wider generalizations. Consequently, those embarking on various studies would do well to carefully consider the categories for doing the study and the specific purpose of the study. For example, the ATLA (or other catalog) subject headings may offer a suitable categorization of resources for creating viable data sets.[15]

This leads to another question for theological librarians who might wonder what role the *ATLA Religion Database*® could have in CA of theological, religious, or philosophical works. Currently one of the biggest challenges for CA of theological, religious, or philosophical topics is the amount of time that CA studies demand. For example, the previously cited example of a CA of Ecclesiastes scholarship examined about 30 journal articles. The gathering and sorting of the data required over 60 hours of Senapatiratne's time, whereas Trott's previously cited work[16] required over 120 hours of data gathering, analysis, and synthesis. With this type of time commitment required, large-scale studies that would be the most helpful for theological librarians are unlikely to be done. Thus, the use of technology to develop cross-referencing within databases is essential. Databases that link cited material in one article to its source article are invaluable for CA studies since much of the time-consuming work has already been done. In this regard, the *ATLA Religion Database*® could be a major part of the solution for sustainable and significant work. Developing this sort of cross-referencing system could move the CA efforts from a narrow and limited scope to one that could eventually even allow for developing impact factors for theological, religious, and philosophical journals.

At this point there seem to be two areas for immediate growth. The first is for librarians to encourage journals and databases to create better avenues for capturing the data necessary for doing citation analyses. This could be as simple, for example, as *Theological Librarianship* hyperlinking footnotes to those materials available through the ATLA databases.

---

[15] See pages 4 and 5 of Senapatiratne (2014) for justification of this idea.

[16] Trott, 2011.

The second area of growth is for librarians to undertake the difficult task of doing CA work. No amount of theoretical knowledge of CA will replace the praxis of doing the work first hand. Doing the work of CA will likely produce thoughtful and innovative ways of capturing these data and, more importantly, developing better librarianship practices as a result of their conclusions.