

Citation Analysis and Theory

K

James Tsung Juang Wang

National Taipei University of Technology, Taiwan

INTRODUCTION

Citation describes a relationship between two scholarly authors or among several scholarly authors. The relationship may be studied by citation analysis, now usually called citationology, for the purpose of determining comparisons, contrasts, or, most importantly, influences between and among authors. The nature of the relationship has two aspects: intellectual and social. The intellectual dimension is concerned with the access and retrieval of information and the development of research and ideas. The social dimension is mainly concerned with evaluating the scholarly performance of an author, and as such may be instrumental toward that author achieving tenure, promotion, and funding within the academic community. Actually, the two conceptual aspects of citation might be said to be so intertwined that it would be impossible to separate the intellectual aspect from the social aspect. As Merton (1942/1973) argues, the ethos of science assumes that knowledge is communal and freely shared, so it is, in effect, not only by discovering knowledge but also by sharing that discovery that a scientist builds a reputation within the scientific community. Citations are both markers of intellectual accomplishment and the extent to which that knowledge has been shared with others and has been influential. Thus the intellectual and social components of the concept of citation are virtually inseparable. The purpose of this article is, first, to illuminate the important features and controversies relating to both the social and the intellectual components of citation, and, then, to investigate the possibility of the establishment of a somewhat belated generally acceptable theory of citation.

BACKGROUND

The Development of Citation Practice

The traditional metaphor of citation is Newton's famous aphorism that if he had seen more, it was because he was standing on the shoulders of giants. Since the advent of modern science in the 17th century citation has been regarded as a moral responsibility among scholars, a matter of paying intellectual debts through formal acknowledgement and making the sources of all arguments openly accessible to everyone, allowing other scholars to add to the conversation about what Bacon called the advancement of learning. By the 19th century science had become thoroughly codified and structured and conceptualized as how we know it today. In the 20th century citation flourished in the form of extensive footnotes and bibliographies. Although citation is common to both science and the humanities, the interest of scholars of citationology is especially focused on the physical sciences, so much so that the standard unit of analysis is the scientific paper.

Two outstanding scholars in the tradition of citation analysis are Robert Merton and Eugene Garfield. In the 1940s Merton published a sociological argument outlining the ethos or culture of science. Then in the 1950s Merton advanced the claim that scientific knowledge is a public good, a fact that facilitates priority being recognized as the main feature of the reward structure of science. Priority – being the first to discover something in science – is a concept that resonates with both the intellectual and the social facets of citation analysis, for it is central to both information retrieval and social recognition in the scientific community. Meanwhile Garfield, dissatisfied with the existing methods of cataloguing and classifying published information, and fortuitously inspired by the American system of legal documentation, developed the *Science Cita-*

DOI: 10.4018/978-1-4666-5888-2.ch442

tion Index in 1961. This was soon followed by *the Social Sciences Citation Index (SSCI)*, and *the Arts & Humanities Citation Index (A&HCI)*. Ultimately these Reuters Thomson citation indexes became the *Web of Science*. In recent years two more important citation databases, *Scopus* and *Google Scholar*, have appeared. Not only are these databases comprehensive, but they are also multidisciplinary.

The immediate future promises many new developments in the discipline of citation analysis. Cybermetrics and webometrics reconstruct the relationship of cited and citer and journal and reader from a slowly extending chain to a rapid and constantly changing network, while scientific work becomes increasingly more collaborative, making citations ever more complex. Moreover, the epistemological paradigm of constructivism is challenging the traditional notions of the objectivity of both nature and scientists, thereby disrupting both the cognitive assumptions and the social ethos of science. In the not-too-distant future scientific communication and citation analysis might both look different than they do today.

A good fundamental source about the complex ways that rapidly developing Information and Communication Technologies (ICTs) are changing citation analysis is given by Thelwall's (2004) book *Link Analysis*. In webometrics link analysis is analogous to citation analysis, and as such it serves to focus study in this emergent discipline. Ingwersen (1998) introduced the concept of Web Impact Factor, and, according to Björneborn and Ingwersen (2004, p. 1217), the four main areas for webometrics research are "(1) web page count analysis; (2) web link structure analysis; (3) web usage analysis (including log files of users' searching and browsing behavior); Web technology analysis (including search engine performance)." Related to link analysis is the increasing capabilities for computerized mapping and visualization that are outlined and explained in some detail by Börner, Chen, and Boyack (2003). One obvious and important difference between citation analysis and link analysis is that links are infinitely more detailed and complex than citations in the amount of information they yield, and one of the great challenges of the future appears to be the development of sound methodologies for coping with that profusion of possibilities.

Kinds of Citation

Citation begins with the classic footnote and extends all the way to the multiple and various contemporary systems of referring to previously published or unpublished documents. The original purpose of citation was assumed to be the civilized acknowledgement of intellectual debts and the invitation to readers to peruse the same documents that the author of the book or article at hand had read and considered in the process of writing. All that changed with the introduction of the *Science Citation Index* by Eugene Garfield in 1961. Suddenly the "lowly footnote," as Norman Kaplan ironically called it in 1965, became the unit of analysis for the new science of citationology. Citation counts, made readily available through the *SCI* and the growing technology of computerization, soon revealed patterns and clusters that information scholars could interpret in rapidly expanding ways. Sher and Garfield (1966), for instance, revealed a significant relationship between the frequency of citation of scientific papers and the winning of the Nobel Prize.

Co-citation may be defined as the frequency that two documents are cited together. Small (1973) first proposed this concept as a means for studying the specialties of science through the observation of networks and clusters revealed by data collected from the *SCI*. Whereas the older methodology of bibliographic coupling links source documents, co-citation links cited documents. Moreover, co-citation patterns do not tend to agree with bibliographic coupling patterns, but they do tend to agree with direct citation patterns. The patterns themselves are a measure of semantic similarity. Gipp and Beel (2009) proposed a refinement of co-citation analysis with the introduction of Citation Proximity Analysis (CPA). This is an approach that identifies how close citations are to each other in a cited text, the assumption being that the closer citations occur, the more likely they are to be semantically related. CPA promises greater precision in the identification of both related materials and related sections.

7 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/citation-analysis-and-theory/112893

Related Content

Classification of Polarity of Opinions Using Unsupervised Approach in Tourism Domain

Mahima Goyal and Vishal Bhatnagar (2016). *International Journal of Rough Sets and Data Analysis* (pp. 68-78).

www.irma-international.org/article/classification-of-polarity-of-opinions-using-unsupervised-approach-in-tourism-domain/163104

Risk Regulation Regimes of Radio Frequency Information Technology

Joshua M. Steinfeld (2015). *Encyclopedia of Information Science and Technology, Third Edition* (pp. 6282-6294).

www.irma-international.org/chapter/risk-regulation-regimes-of-radio-frequency-information-technology/113084

Strategy for Performing Critical Projects in a Data Center Using DevSecOps Approach and Risk Management

Edgar Oswaldo Diaz and Mirna Muñoz (2020). *International Journal of Information Technologies and Systems Approach* (pp. 61-73).

www.irma-international.org/article/strategy-for-performing-critical-projects-in-a-data-center-using-devsecops-approach-and-risk-management/240765

Methodologies for Investment Evaluation: A Review and Assessment

Egon Berghout and Theo-Jan Renkema (2001). *Information Technology Evaluation Methods and Management* (pp. 78-97).

www.irma-international.org/chapter/methodologies-investment-evaluation/23669

FLANN + BHO: A Novel Approach for Handling Nonlinearity in System Identification

Bighnaraj Naik, Janmenjoy Nayak and H.S. Behera (2018). *International Journal of Rough Sets and Data Analysis* (pp. 13-33).

www.irma-international.org/article/flann--bho/190888