

Appendix 3

Citation and Content Analyses

Claire Donovan

Citation Analysis

The professors' survey (Chapter 8 and Appendix 1) shows who, for them, were the most influential and celebrated sociologists in the twentieth century. Of course, the preferences of the professors do not necessarily reflect the views of the profession at large today or throughout various stages of the twentieth century. A citation analysis of British sociology journals may appear to provide an alternative 'objective' measure of influence but (and this may at first seem counter-intuitive) citation analysis is not as straightforward a measure of influence as the professors' survey. While citation analysis quantifies citation patterns, the methods and the interpretations of such analyses have been hotly contested. It is therefore necessary to explain our particular choices of method, and how this affects the outcomes of our study and what realistically can be measured.

Citation analysis is familiar through the efforts of the Philadelphia-based Institute of Scientific Information (ISI)¹ established by Eugene Garfield in 1958, and through such resources as the *Web of Science*.² The ISI compiles the Science Citation Index (SCI), the Social Science Citation Index (SSCI) and the Arts and Humanities Citation Index (AHCI) from references to journal articles and some edited collections of papers published as books. The analysis operates by recording the references made in the footnotes, endnotes, or bibliographies of these papers and presents the total frequencies (and source details) of citations of authors, books, papers, and journals. For the social sciences, ISI (in the form of SSCI) covers 1,725 journals which span 50 disciplines and add to over 30.15 million entries. ISI data were originally employed in the late 1960s and early 1970s by historians of science and sociologists of science. But what precisely does citation analysis measure?

Citation analysis is used to indicate relative levels of impact by particular researchers or particular works in certain subjects. How often an author or a work is cited, is treated as a gauge of influence and of research quality. Citation

¹ <http://www.isinet.com/isi/>

² <http://wos.mimas.ac.uk/>

patterns are taken as an objective measure of research quality free from any subjective biases, providing a simple empirical record of who influences the work of others in their field (Bavelas, 1978). So far so good. However, the production and interpretation of citation analyses have not only been the object of sociological use but also the subject of sociological investigation, so we should expect to find that these processes are not as straightforward as they may at first appear. The traditional view takes for granted the question of what citation analysis really is, something that has become by default 'embodied in the procedures of ISI' (Hicks and Potter, 1991: 481). Critics have seized upon two areas of weakness, with citation analysis generally (Phelan, 1999) and ISI procedures in particular (Chapman, 1989).

In the late 1970s, citation analysis was criticized by 'new' sociologists of science who linked attitudes towards citation to competing philosophies of science. The traditional and social constructivist views are 'two entirely different accounts of the validity and function of citation where on the one hand citations are unproblematic and citation analysis provides valid data for analysis, and on the other hand citations are seen as a deeply problematic basis for data evaluation' (MacRoberts and MacRoberts, 1996: 438). While some critics believe that despite some obvious flaws, if handled correctly, citation analysis remains a fairly reliable measure of research impact, constructivists largely dismiss citation analysis as an indicator of anything concrete. The social constructivist approach thus entails a sceptical view of the worth of citation analysis as its basic assumptions 'are clearly false' (1996: 422) and we must question citation data as a true measure of the value of a work or as a means of bestowing credit on colleagues.

The constructivist approach mentioned above differs from the traditional perspective because it is concerned with social aspects of citation; 'a scientific contribution does not become legitimised until it has been endorsed by other scientists' (Baldi, 1998: 829–30). Thus, scientific progress is part of a socially constructed pattern of agreement and endorsement. Attention is given to the context of citation or 'how authors use their colleagues work' (Hargens, 2000: 857), and social science is thought to differ from natural science (the model of disciplinary organization upon which the traditional approach depends) because social scientists tend to cite foundational rather than recent publications. In this light, citations are taken to have a rhetorical use and social scientists in particular use citations to justify and contextualize their work by aligning their efforts with influential texts and authors. This has become part of the expected 'packaging' for journal articles because social science is written for a 'heterogeneous' audience of allegiances and fields. If there is little assumed shared knowledge the author is obliged to go back to first principles to explain why his or hers is a significant contribution (Hargens, 1999, 2000).

Several authors note that when social scientists refer to foundational or 'classic' works, in contrast to natural scientists they rarely cite empirical evidence to elaborate research methods and they instead tend to cite general themes, sometimes mistakenly (Platt, 1984; Chapman, 1989; Delamont, 1989). Citations

AQ: In Hargens, 1999, 2000, Is it 1991 as listed.

are often used as 'totemic representations' of general approaches initiated by foundational works and become 'shorthand markers of general perspectives' (Hargens, 2000: 859–60), while past empirical results are, it seems, 'delivered to an empty house' (Cozzens, 1985: 147). It is felt that there is often citation without knowledge, 'No doubt some authors sometimes cite in a perfunctory fashion, without themselves having more than a superficial knowledge of particular papers in their reference lists' and this may add to the citation counts of 'established, cited works and eminent persons' (Chapman, 1989: 341).

Critical concerns with citation analysis may be divided into two areas: 'conceptual concerns'—what citation analysis truly measures; and 'methods'—the technicalities of how citation analyses should be conducted.

Limitations of citation analysis: conceptual concerns

An issue of fundamental importance is what citation analysis really measures. The traditional ideal is that citation levels point to the inherent quality of the research cited but quantity does not simply mean quality. For example, there is the idea of negative citation, that referring to a work 'does not necessarily denote approval' as some highly cited research may be 'poorly regarded, perhaps defective' if 'there has been a failure to replicate or verify' or 'even because the research is known to be fraudulent'. Yet negative citations do indicate the impact of research as academics tend to ignore unimportant work and negative citations may indicate that 'a piece of work has substance' as a valuable step in the development of knowledge (Chapman, 1989: 341). An alternative approach is to accept that citations measure 'impact' so that uncertainties surrounding any unrecognized or inherent qualities of a piece of work may be avoided (Martin, 1996). Judgements of research 'quality' are beyond the scope of what citation analyses can measure while assessments of 'impact' are not.

Yet this approach too has its shortcomings. Although we may accept that citation analysis is a reasonable indicator of research impact, there are many influences that it cannot measure. There is the phenomenon of 'obliteration' where original sources are no longer cited because sociological concepts and methods have become assumed knowledge (Hicks and Potter, 1991; Hargens, 2000). This is an important consideration because the traditional view assumes that scientists cite work that has influenced them, yet MacRoberts and MacRoberts (1996) found that bibliographies typically cover only 30 per cent of the influence evident in the body of a paper, a figure they maintain applies equally to natural science and to sociology. Indeed, their analysis finds one-third of references credited to someone other than the original researcher. In this sense, citation is a far more complex process than assumed by the traditional approach, particularly when we add that informal influences, such as the views of colleagues, go uncited (Cronin *et al.*, 1993) and influence credited in the acknowledgements section of a paper may be more significant than a particular reference made or the efforts of a co-author. Citation patterns may also be distorted by the activities of invisible colleges or networks, although it

is fair to say that where these exist they are 'little more than a manifestation of the power relations existing within that field. That citation counts reflect this reality is not a methodological shortcoming' (Phelan, 1999: 124).

When considering 'impact' a further factor is the assumption that 'all papers have the same probability of being cited, the same potential citing audience'. However, while sociologists working in a small specialist area may gain few citations this does not mean that their work has relatively low impact, but reflects the fact that in a narrow field of research there are few people to cite and to be cited by. It follows that, if we are to 'determine if a paper... has received the proper number of citations, [we] would first have to know its potential audience. This is never known' (MacRoberts and MacRoberts, 1996).

Limitations of citation analysis: methods

Alternative methods will yield different results. For example, the range of sources from which bibliographies are drawn will affect the outcome of any study. Moreover the question arises whether all the works cited in chosen bibliographies should be included for analysis. As potential sources, should all journal entries be treated equally irrespective of length and content? Should research papers, review articles and short book reviews be equally weighted? Are reviews original research? The ISI includes all types of journal entries, and while incongruities will occur because it is 'unavoidable that very large-scale, computerised enumeration lacks sophistication' we may remain uncomfortable with the idea that all journal contributions are treated equally (Chapman, 1989: 341). As a crude measure of the significance of a contribution, smaller-scale citation analyses often exclude bibliographies that, for example, belong to papers less than ten pages long although it is unusual to exclude any work cited by an accepted bibliography. The ISI does not include all the fringe journals of various disciplines or fields, and the efforts of editors of journals and books may go without credit, factors that all smaller-scale analyses must grapple with.

One of the most pressing issues, particularly when dealing with the social sciences, is that bibliographies of monographs are not included as sources for ISI information. Citation analysis based upon journal publications evolved from the study of publishing patterns in the natural sciences where the dynamic of research compels scientists to report their findings quickly before they are 'scooped': and so they publish in journals. Social scientists, according to some, are not so concerned about this speed of publication and often prefer 'lengthy exposition and extensive reference to past work' in book form (Hargens, 1991). In this light the SSCI is 'asymmetric' as the impact of books is greatly underestimated (Clemens *et al.*, 1995). It may, however, be noted that while the SSCI does not include the bibliographies of books it does represent the impact of books that are cited by papers.

Various approaches to recording authors' details will affect the outcome of citation analyses. ISI lists only the first author of a work rather than all named contributors. So, for example, in the *Affluent Worker* series, Goldthorpe gains

the full ISI citation and Lockwood, Bechhofer, and Platt go uncredited, while for a later joint publication with Erikson, Goldthorpe is unrecognized. This is clearly a problem when attempting to assess the influence of particular persons. Too great or too small a weighting may be given to an individual's contribution. A moot point is whether reference to one's own work should be allowed. At an aggregate level this may not be a problem although 'removing self-citations is an important prerequisite when comparing the performance of specific researchers' (Phelan, 1999).

A crucial consideration is that ISI does in some respects, and not in others, take account of the *distribution* of references. For example, ISI will allow only one citation per work no matter how many times it is cited by one source, but will allow the same source to provide numerous citations to an author or journal if different works are mentioned. Chapter 9 demonstrates that this has dramatic consequences in frequency counts.

While there is suspicion of the value of international data sets, an emerging counter-trend within the recent literature is to develop alternative bibliometric approaches designed for specific purposes, and small-scale nationally oriented studies are proving to be more revealing than ISI data alone (Hicks, 1999; Phelan, 1999). In spite of various criticisms of citation analysis, this self-aware smaller-scale approach is a simple empirical measure of the most cited individuals and works in particular contexts, and whatever motivations may or may not lie behind citation behaviour, the outcome is representative of sociologists' citation choices at given times.

Citation analysis: our approach

So how did we grapple with the various conceptual and methodological concerns about citation analysis? And how did we model our research design? We followed the lead of Hicks and Phelan and set out to construct a small-scale nationally oriented study designed to overcome the perceived shortcomings of the ISI approach.

Our sources were the three mainstream British sociology journals: *Sociological Review*, the *British Journal of Sociology*, and *Sociology*, taken decennially from 1910 to 2000. We included all types of journal entries that were at least ten pages in length (including notes and bibliography). The result was a sample of 371 papers.

The credit for jointly authored papers was divided equally between all authors rather than purely on a first author only basis. Self-citation was allowed and noted so that the number of self-citations could be subtracted if required. The distribution of references in each paper was limited to one per journal, author, and specific collaboration, although the unweighted totals were recorded for comparison. Our main concern was with the number of citations to authors and journals, and once the most highly cited authors were listed their most cited texts were identified. This revealed that the impact of books is not lost in a journal-based analysis.

Our citation study provides a snapshot of the *influence* of particular sociologists and specific texts on the citation habits of sociologists in Britain in different decades of the twentieth century. Our methods modify and improve traditional approaches and because we measure the direction of sociology through the sum total of citing authors' actions, constructivist arguments about their underlying motivations become redundant.

Content Analysis

In contrast to citation analysis, content analysis of journal papers is a less common activity and perhaps as a consequence has historically been less controversial. Content analysis does not derive its data from bibliographies—the object of study is the construction and content of articles. Past examples of the content analyses of sociology papers tend to be allied to the history of sociology and trace the popularity of various areas or fields of sociology (Carter, 1968; Simon, 1969; Collison and Webber, 1971) and study the research methods employed by sociologists (Bechhofer 1981, 1996; Platt, 1981; Bulmer, 1989; Gartrell and Gartrell, 2002). Our content analysis expands and modifies these approaches, largely by extending coverage from 1910 to 2000. We also include an original study of the ideologies which underpin the sample papers, mapping the rise and fall of various sociological perspectives (or 'isms') throughout the twentieth century. Thus our content analysis provides a comprehensive empirical study of the development of research methods, the changing fashions in popular areas of sociological enquiry and shifts in ideology, and how these interact, as represented by the mainstream journals for the century that sociology has existed in Britain.

AQ: Platt 1981 ref.
not listed

Our starting point was to analyse the papers originally used in our citation analysis. In order to further analyse trends we extended the years covered to include quinquennial years from 1975 onwards, yielding a sample of 649 papers. The detail of our analysis meant that each paper had to be individually scanned and while this process was intense and lengthy it produced a rich and valuable resource.

Our analysis was constructed on a 'text unit' basis (Collison and Webber, 1971) so that each paper is a potentially divisible unit. So, for example, when we calculate the number of women who employ highly quantitative methods in a particular year, if a paper is a female/male joint collaboration this will equal 0.5 of a text unit. All variables were treated in this manner, where appropriate. In addition to the variables listed below data recorded for each sample paper included: year of publication, journal, journal number and volume, name(s) of author(s), gender, institutional affiliation, discipline, and nationality.

Methods used by sociologists

The previous studies mentioned above are confined to particular periods of time and tend to focus on 'positivism' or empirical methods, and quantification in particular. Our aim was to extend the coverage for the period 1910–2000 and to

study the development of both quantitative and qualitative research methods and levels of technical sophistication or otherwise. Sample papers were categorized under the following headings: theory only, empirical, quantitative, qualitative, and both quantitative and qualitative. Quantitative papers were further divided into: descriptive, low, high, and high and theory only. Qualitative papers were classified as belonging to either the survey or anthropological tradition. (See Chapter 10 for further elaboration.)

Areas studied by sociologists

Past studies of the areas that sociologists choose to study are based upon categories developed by the Sociological Abstracts Classification Scheme (SACS),³ which divides sociology into a number of areas. We modified this starting point so that, for example, we could distinguish between social theory and sociological theory, and we both expanded and compressed several SACS classes to better handle and interpret our data. Our aim was to provide a map of the ebb and flow of interest in the various sociological fields for the whole as the last century as represented by the core journals. The areas covered include:

- Ageing—includes generations, gerontology, the elderly
- Biology and behaviour
- Sociology of the body
- Comparative sociology
- Consumption
- Culture—includes museums, music
- Demography—includes migration
- Economic organization—includes industrial relations, political economy, development theory
- Education—includes universities
- Ethnicity and race
- Family—includes marriage, divorce
- Gender—also includes masculinity
- Health and medicine—includes deviance and labelling in terms of mental health, addiction, accidents, death
- History of sociology—includes the contemporary state of sociology, the future of sociology
- Sociology of knowledge
- Law and crime—includes deviance and labelling connected with law breaking behaviour or perceptions of potential criminality or recidivism, prisons, the police, punishment, crime prevention, surveillance
- Leisure—includes tourism
- Methods—includes practical issues concerning research methods, research ethics, problems of method, discussion of perceived methodological flaws in sociological research
- Media and communication—includes television, censorship

³ <http://www.csal.co.uk/csa/factsheets/saclass/shtml>

Occupations—specific to particular occupations and professions (not occupational scales, which are classified under stratification/social differentiation)
 Organizational theory—includes state level (bureaucracy, corporatism) and macro level (individual company structures and management schemes)
 Political sociology—includes political theory, nationalism
 Religion—includes paranormal behaviour/beliefs
 Rural sociology
 Science and technology
 Social anthropology
 Social geography—includes ecological sociology and environmentalism, regional sociology and planning (rural planning is classified as ‘social geography’ and ‘rural’), community studies
 Social history
 Social movements
 Social policy
 Social psychology
 Social theory—theoretical papers that do not use or discuss the use of sociological data
 Sociological theory—theoretical papers that use or discuss the use of sociological data
 Social welfare—includes social work
 Sociology of sport
 Stratification/social differentiation—includes class, occupational scales, mobility
 Urban sociology—includes town planning
 Youth

The sample papers could belong to more than one of these areas. The list was reduced to cover any area that was represented by 5 per cent or more papers in any given sample year.

Sociologists and ideology

As far as we are aware, this aspect of our citation analysis is a unique attempt to chart the rise and fall of ideological interests (sociological perspectives or ‘isms’) for the duration of the twentieth century. We took our starting point as the commitment to ideologies expressed in the professors’ survey and we added new categories as we encountered them in our sample. The resulting list is as follows:

Action/agency/structuration theory—includes Parsonian/positivist action theory, agency and structure (again in the positivist sense), holism v. individualism debate, structuration
 Critical theory—includes critical realism
 Empiricism—this does not include all articles that use quantitative empirical evidence, but refers to work that is explicitly empiricist or that may be placed in this paradigm

Exchange theory—includes the Mauss/Levi-Strauss collectivist approach, whereas the individualist approach is classified under ‘rational action/choice’

Feminism

Foucauldianism

Freudianism

Functionalism

Interpretivism—includes interactionism, ethnomethodology, hermeneutics

Le Playism

Marxism—includes Neo-Marxism

Positivism—includes papers that discuss positivism or that make statements that support or are commensurate with the view that social science should aspire to equivalence with the natural and physical sciences

Post-modernism—includes post-industrialization and post-Fordism

Post-structuralism—intertextuality, anti-humanism

Rational action/choice—includes rational action theory, rational choice theory, methodological individualism, individualist exchange theory

Rationalism/realism—excludes critical realism which is classed as ‘critical theory’

Relativism

Social constructionism

Structuralism (a)—the view that society is prior to individuals

Structuralism (b)—unobservable social structures that generate observable social phenomena (e.g. Levi-Strauss)

Weberianism—includes neo-Weberianism

Papers could subscribe to one, several or none of these perspectives. Again, this list was reduced to ideologies covered by at least 5 per cent of sample papers in any year.