Chapter XIII
Publication and Citation Analysis as a Tool for Information Retrieval

Ronald Rousseau
KHBO (Association K.U. Leuven), Belgium

ABSTRACT

In this chapter an overview of citation analysis is presented, emphasizing its formal aspects as applied social network theory. As such citation linking can be considered a tool for information retrieval based on social interaction. It is indeed well known that following citation links is an efficient method of information retrieval. Relations with Web linking are highlighted. Yet, also social aspects related to the act of citing, such as the occurrence of invisible colleges, are discussed. I present some recent developments and my opinion on some future developments. In this way I hope the reader will realize how the fields of citation analysis and Webometrics can be helpful in building social information retrieval systems.

INTRODUCTION: DEFINITION OF CITATION ANALYSIS

We define citation analysis as that subfield of bibliometrics where patterns and frequencies of citations, given as well as received, are analyzed. Such an analysis is performed on the level of authors, journals, scientific disciplines, and any other useful unit or level. Citation analysis further studies relations between cited and citing units (documents, authors, countries, etc.). From an application point of view, citation analysis may be considered as a collaborative peer effort to analyze and promote the quality of scholarly publication and research. For a review of citation analysis as a subfield of informetrics, we refer to Wilson (1999) and Borgman and Furner (2002).

Science is a social and accumulative endeavor, even if occasionally whole areas are overturned by new evidence. No scientific discovery or activity is conducted in splendid isolation, and new work is always based in some way on the work
of predecessors. Citations reflect this social and accumulative aspect by linking the past with the present, and possibly the future. They represent an intrinsic part of the progress and development of science.

Contributions to scientific knowledge are often crystallized in the form of a scientific article. Such contributions may take the form of new facts, new hypotheses, new theories or theorems, new explanations, or a new synthesis of existing facts (Russell & Rousseau, 2002). In each case a transition has taken place from an existing, say ‘old’, situation to a ‘new’ one. The transition itself takes place in the head of the investigators with the help of scientific equipment and is usually invisible for outsiders, but scientific tradition requires that an author refers to earlier articles that relate to the theme of his or her paper. The author must clarify his or her starting point. ‘The old’ is revealed by identifying those predecessors whose concepts, methods, and discoveries have inspired or were used in developing ‘the new’. Stated otherwise, the author acknowledges a group of inspirational articles written by earlier researchers by referring to them. The term ‘referring’ means here: mentioning in the proper context and giving an explicit bibliographical statement in the reference list. The older articles are then cited by—receive a citation from—the new one.

A similar process takes place when a new Web site is created and links are attached to existing Web sites. Yet, while the act of citing is accompanied with a time stamp, a time stamp on the Web (if it exists at all) usually gets lost very soon. Of course there is another important difference between the two types of linking. When article A cites article B, then because of this time factor, B will usually not cite A. Note though that recently Rousseau and Small (2005) published a strange case where a giant Escher staircase (Rousseau & Thelwall, 2004) was discovered between 13 articles. Between Web sites or Web pages, reciprocal linking is quite feasible. This is, however, also true for citation links between journals or authors.

Articles connected to other articles through citation linking, journals connected to other journals and to itself, also by citation linking, scientists connected to colleagues through co-authorship—all these relations can be considered as special cases of networks or expressed in mathematical terminology as graphs. Consequently, graph theory and network analysis will constitute an important thread in this contribution. Information retrieval using citation analysis techniques takes advantages of the existence of these links. Garner (1967) was among the first to show the relation between citation indexing and general graph theory. For a deeper insight in all aspects of citation analysis, the reader is referred to Citation Analysis in Research Evaluation by Henk F. Moed (2005).

Although this contribution focuses on citation analysis and Web linking, we cannot ignore the enormous influence of the ‘free online availability’ movement. Preprint servers, open access (gold road), self-archiving (green road), institutional repositories, and other new modes of publication are changing the landscape of scholarly publishing (Bose & Harnad, 2005; Brody, Kampa, Harnad, Carr, & Hitchcock, 2003). It has been shown that open access increases later citation counts and that earlier Web usage can be used as a predictor for later citation counts (Brody, Harnad, & Carr, 2006). Web access has given rise to download counts as a new indicator.

**What is the Relation Between Publication and Citation Analysis on the One Hand and Information Retrieval on the Other?**

Citing is not a purely logical or objective act. It can be compared to decision making under partially unknown circumstances. As all other decisions it involves a socially determined choice. Using publication and citation data for evaluation pur-
Related Content

Extensive Use of Facebook Marketing by Entrepreneurs in Gaining a Wider Reach
[www.igi-global.com/article/extensive-use-facebook-marketing-entrepreneurs/122012?camid=4v1a](www.igi-global.com/article/extensive-use-facebook-marketing-entrepreneurs/122012?camid=4v1a)

Situated Evaluation of Socio-Technical Systems
[www.igi-global.com/chapter/situated-evaluation-socio-technical-systems/21442?camid=4v1a](www.igi-global.com/chapter/situated-evaluation-socio-technical-systems/21442?camid=4v1a)

Importance Analysis of a Blog Quality Model for Criteria and Families in Different Blog Categories
[www.igi-global.com/article/importance-analysis-blog-quality-model/122010?camid=4v1a](www.igi-global.com/article/importance-analysis-blog-quality-model/122010?camid=4v1a)

Some Research Challenges for Studies of Virtual Communities Using On-Line Tells
[www.igi-global.com/article/some-research-challenges-studies-virtual/52987?camid=4v1a](www.igi-global.com/article/some-research-challenges-studies-virtual/52987?camid=4v1a)