Complete graphs and bibliographic coupling: A test of the applicability of bibliographic coupling for the identification of cognitive cores on the field level

Bo Jarneving

Swedish School of Library and Information Science, University College of Borås, Sweden

Abstract

The method of bibliographic coupling in combination with the complete link cluster method was applied for mapping of the field of organic chemistry with the purpose of testing the applicability of a proposed mapping method on the field level. The method put forward aimed at the generation of cognitive cores of documents, so-called 'bibliographic cliques' in the network of bibliographically coupled research articles. The defining feature of these cliques is that they can be considered complete graphs where each bibliographic coupling link ties an unordered pair of documents. In this way, it was presumed that coherent groups of documents in the research front would be found and that these groups would be intellectually coherent as well. Statistical analysis and subject specialist evaluations confirmed these presumptions. The study also elaborates on the choice of observation period and the application of thresholds in relation to the size of document populations.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Bibliographic coupling; Science mapping; Research fronts

1. Introduction

The sub-field of Information Science dealing with citation based mapping of research fields has widely been focused on the cocitation analytical method, which over time has been developed to an intelligence tool for science policy applications (Griffith, Small, Stonehill, & Dey, 1974; Small, 1973; Small & Griffith, 1974; Small & Sweeney, 1985). The "cocitation bibliometric modelling" has been claimed to provide with ". . .a detailed description of international research front" (Franklin & Johnston, 1988) and further development of the method, combining word analysis with cocitation clustering, has been accomplished (Braam, Moed, & van Raan, 1991a, 1991b). Less attention has been paid to another citation-based method, bibliographic coupling, which was introduces through a series of reports and articles in the early 1960s (Kessler, 1960, 1962, 1963a, 1963b, 1965). In these reports, with the focus on document retrieval, the applicability of bibliographic coupling as method for the coupling of similar documents was established. While the method of cocitation analysis prospered and was added with the author cocitation analytical method (White & Griffith, 1981), further important empirical findings did not show up until 1984 when Vladutz and Cook performed a successful large scale experiment testing the hypothesis that strong bibliographic coupling links imply strong subject relatedness between

E-mail address: bo.jarneving@hb.se.

articles. The question of cognitive resemblance between bibliographically coupled research papers was also elaborated in Peters, Braam, and van Raan (1995), where it was found that word profile similarity within document groups sharing a citation to a highly cited publication was significantly higher than between documents without such a relationship. One can conclude that the research on bibliographic coupling so far had not explicitly elaborated on the applicability of bibliographic coupling as a science mapping tool. This area of application for bibliographic coupling, however, was presented the same year in a proceeding by Glänzel and Czerwon, and subsequently in an extended version (1996), who claimed that bibliographic coupling could be used to identify "hot" research topics, meaning that documents connected by many and strong bibliographic coupling links can be applied for the mapping of research fronts. Their method was based on the identification of so-called "core documents", which were defined as documents coupled with at least 10 other documents with a set minimum of normalized coupling strength. The measure of normalized coupling strength applied had previously been suggested by Sen and Gan (1983) as the "coupling angle" (C.A.) between documents, analogous to the well known Salton cosine measure. Both empirical experience as well as theoretical consideration underlied Glänzel and Czerwon's definition, and they claimed that a too small number of coupling links may favour articles published in series and a too large number may exclude smaller research specialties. The authors recommended a threshold of normalized coupling strength of 0.25, which was deemed necessary for the filtering out of less significant links between documents. For their empirical experiment, Glänzel and Czerwon applied a whole volume of the *Science Citation Index* (CSI), comprising more than half a million research articles. Approximately 1% of these fulfilled the requirement of a core document. An important finding was that core documents tended to be cited above the average by subsequent research articles, which, from an additional angle, justified the label 'core document'. The authors concluded that a method of bibliographic coupling may well be applied for science mapping purposes. Most important, applying bibliographic coupling, a research article is available for analysis as soon as it is published, making the identification of emerging specialties more feasible in comparison with the cocitation method. In spite of its favourable features, the total research on bibliographic coupling as a method complementary to cocitation mapping has been meagre. The cause for this unobtrusive position of bibliographic coupling is not obvious and comparable and complementary results to the cocitation mapping approach have been reported (Jarneving, 2001; Persson, 1994; Sharabchiev, 1988).

This study had as its goal to elaborate further on this method's applicability in the context of science mapping. Considering the fact that only a small fraction of research articles would fulfil the requirements of a core document, the investigation of the applicability of bibliographic coupling as a method for science mapping should not be delimited to the analysis of core documents exclusively, and other models of document selection should be tried. On basis of theoretical considerations and previous findings, a method of bibliographic coupling is proposed and applied for the exclusive identification of cognitive cores in the research-front network, in accordance with requirements that stipulate both the shape of clusters as well as their internal strength of coherence.

1.1. Statement of purpose and research questions

The purpose of this study was to investigate the applicability of a compound method constructed for the purpose of the identification of coherent research themes within a defined science field. The two major components of the method put forward are (i) a method for the measuring of document–document similarity and (ii) a method for the partition of document populations into disjoint groups. With regard to (i), a normalized measure of bibliographic coupling strength was applied and with regard to (ii) a hierarchical agglomerative cluster method. The hypothesis was that this compound method would make it feasible to identify cognitive cores of documents, i.e. document groups such that the relationship between documents within groups is strong, and markedly stronger than the relationship to documents outside the group. The research questions stated were the following:

Q1 How well does the suggested method perform with regard to the generation of internally coherent and externally isolated clusters?
Q2 To what degree do generated clusters represent valid depictions of research themes?
Q3 How do results from the application of the suggested method comply with a subject specialist's apprehension of the cognitive structure of the research field?