



ELSEVIER

Contents lists available at [SciVerse ScienceDirect](http://SciVerse.ScienceDirect.com)

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Citation analysis: A social and dynamic approach to knowledge organization



Birger Hjørland*

Royal School of Library and Information Science, University of Copenhagen, 6 Birketinget, DK-2300 Copenhagen S, Denmark

ARTICLE INFO

Article history:

Received 7 February 2013
 Received in revised form 11 May 2013
 Accepted 1 July 2013
 Available online 25 July 2013

Keywords:

Approaches to knowledge organization
 Information organization
 Bibliometrics
 Citation analysis
 Epistemology

ABSTRACT

Knowledge organization (KO) and bibliometrics have traditionally been seen as separate subfields of library and information science, but bibliometric techniques make it possible to identify candidate terms for thesauri and to organize knowledge by relating scientific papers and authors to each other and thereby indicating kinds of relatedness and semantic distance. It is therefore important to view bibliometric techniques as a family of approaches to KO in order to illustrate their relative strengths and weaknesses. The sub-field of bibliometrics concerned with citation analysis forms a distinct approach to KO which is characterized by its social, historical and dynamic nature, its close dependence on scholarly literature and its explicit kind of literary warrant. The two main methods, co-citation analysis and bibliographic coupling represent different things and thus neither can be considered superior for all purposes. The main difference between traditional knowledge organization systems (KOSs) and maps based on citation analysis is that the first group represents intellectual KOSs, whereas the second represents social KOSs. For this reason bibliometric maps cannot be expected ever to be fully equivalent to scholarly taxonomies, but they are – along with other forms of KOSs – valuable tools for assisting users' to orient themselves to the information ecology. Like other KOSs, citation-based maps cannot be neutral but will always be based on researchers' decisions, which tend to favor certain interests and views at the expense of others.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

... the pattern is new in every moment.
 (Eliot, *East Coker*)

The present article is one of a series of papers considering the research traditions, “paradigms” and methodological approaches in knowledge organization (KO). Articles have previously been published on the facet analytic approach and the user-based and cognitive approach, respectively (Hjørland, 2013a, 2013b). Further papers are planned concerning approaches based on similarity measures (mainstream in information retrieval research, IR) and on domain analysis as an approach to KO. In Hjørland (2013c), my overall view of approaches to KO is presented. The goal is to identify distinct traditions in KO and to illustrate their respective strengths and weaknesses. Each approach is treated in a separate article, but overall understanding of my argumentation may be enhanced by the comparative perspective provided in Hjørland (2013c).

* Tel.: +45 67711970.

E-mail address: pnl617@iva.ku.dk

The present paper deals with bibliometric approaches, more specifically methods related to *citation analysis* (argued to be a distinct approach to KO). Bibliometrics (with informetrics and scientometrics) is both an interdisciplinary field and a sub-field of library and information science (LIS). It is predominantly considered a separate field from knowledge organization (KO). Bibliometrics and KO have, for example, separate textbooks without much overlap and the number of mutual citations between these two fields is low.

There are, however, important exceptions to the separation of KO and bibliometrics. Bibliometric techniques have been applied to construe knowledge organization systems such as automatic indexing (Salton, 1971), thesaurus construction (Rees-Potter, 1989; Rees-Potter, 1991; Schneider, 2004), KeyWords Plus indexing (Garfield & Sher, 1993), Research Front indexing (Dehart & Scott, 1991), the well-known Google “PageRank” algorithm from 1996, and mapping/visualizations of knowledge domains (Chen, 2003; Chen, Ibekwe-Sanjuán, & Hou, 2010; Vargas-Quesada & de Moya Anegón, 2007; White & McCain, 1998). Also studies such as Pao (1993) and Pao and Worthen (1989) considered citations as subject access points and examined their relative value in information retrieval. Bibliometrics, specifically citation analytic methods, should therefore be considered one among other families of approaches to KO, and as a family that competes with or supplements other approaches such as those already mentioned.

Using citation-based methods as a complement or alternative to conventional approaches to KO is thus not new in the bibliometric community, but rather tends to be neglected by the KO community. The paper should be understood as a comparative theoretical analysis of the assumptions in citation analysis compared with those in traditional forms of KO. Such an examination has not formerly been made.

2. Bibliometric maps as knowledge organization systems (KOSs)

In KO the concept of a knowledge organization system (KOS) is a generic term used for authority lists, classification systems, thesauri, topic maps, ontologies, etc. (Hodge, 2000). A KOS can be defined as a selected set of concepts together with an indication of (some of) their semantic relations. Currently, front-end research on KOSs is considering ontologies, KOSs with the most flexible variety of semantic relations and with formal structures, which allow automatic inference (i.e., the search for objects based on logical rules). Hodge (2000) presented a taxonomy of KOSs, but did not include bibliometric maps. However, it will be argued below that bibliometric maps should be considered a form of KOS in that they display a selection of concepts and an indication of (some of) their semantic relations.

KOSs organize *concepts*, for example, species and their relations to other species (in the form of, among others, genus–species relations and part–whole relations), as well as documents based on subject relations (e.g., books about the Vikings). KOSs therefore concern conceptual relations and subject relations. The unit in citation analysis in bibliometrics is, however, a single *document* and its bibliographical relations – in the form of references or citations – to other documents. If, for example, document A and document B are both cited by a third document (“co-cited”), the relation between A and B is a bibliographical relation, and only possibly (and secondarily) also a semantic relation or a subject relation.¹ Because this paper seeks to describe the principal differences between methods based on *citation analysis* (described below) and those based on traditional KOSs, it is important to emphasize that citation analytic methods are frequency relations based on relations between documents,² while conceptual relations and subject relations, which form the units of traditional KOS, are here (at best) *indirect* relations.³ This is probably one of the main reasons for the relative separation of KO and bibliometrics, and the reason that Hodge (2000) did not include KOSs based on bibliometric methods.⁴ Therefore an argument is needed for considering citation relations as a form of semantic relations and subject relations. This argumentation is made in the following section.

3. Citation relations, subject relations and semantic relations

The functionality of using bibliographical references in documents as well as citation indexes for subject retrieval is based on the assumption that there are (normally) subject relations and semantic relations between citing and cited documents.

¹ An anonymous reviewer wrote about a former version of the present paper: “[T]he units of analysis in bibliometrics are kinds of NAMES (nouns and noun phrases) as verbal types, and the key data are counts (i.e., frequencies) of occurrences or co-occurrences of verbal tokens of these types. There are occurrence counts not only for words or word-pairs, but also for names of papers (titles or author–title combinations), names of authors (oeuvres), names of collections of papers (titles of journals), and so on. We can also have co-occurrence counts for pairs of noun-types (e.g., author–author or paper–paper). In addition we can have co-occurrence counts for noun-types of different kinds (e.g., author–subject heading). The reason that all these kinds of noun-phrases can be mapped in the same semantic space (e.g., names of authors with subject headings or with keywords) is that their counts of their co-occurrences in bibliographic records (or full texts) can [be] correlated. It is the correlations, high to low, that determine semantic proximity in maps [...] analyses of various co-occurring noun-types (e.g., co-words, co-descriptors, co-cited papers, co-cited authors, and co-cited journals) have essentially the same purpose, which is to characterize subject space by showing empirical patterns of indexing data. Author-names and journal-titles are vaguer, more indirect indicators of subject than specific descriptors, classification terms, and keywords, but all are used toward the same end”.

² Yan and Ding (2012, p. 1314) wrote: “an article is usually a single research unit that can be aggregated into several higher levels, for instance, the author unit, the journal unit, the institution unit, and the field unit”.

³ In terms of citation-based analyses, this indirect relation is obvious, but when combining semantic and citation analyses, or when applying bibliometric methods to the semantic properties of documents, one could argue that the conceptual relations are addressed in a more direct way.

⁴ Traditionally, the two main functions of library classification have been (1) shelf arrangement and (2) information retrieval in catalogs. Bibliometric methods cannot be used for shelving. This may be another reason why the fields of KO and bibliometrics have not made better contact. In this paper only the IR function is considered.

However, what is meant by this is far from a trivial question.⁵ A human being may intuitively perceive whether a citing and the cited document are “about the same subject” or not. Such a judgment cannot, however, be verified unless we are able to operationalize the concept of “subject”, which involves deep philosophical questions. Such relations cannot be determined by comparing titles or by measuring the similarity of citing and cited papers by means of word co-occurrences because co-word analysis is itself a measure that needs to be validated by other methods. Recent studies have used statistical measures of textual coherence (e.g., Boyack & Klavans, 2010), but also this approach needs further motivation.⁶ We end up with some deep theoretical problems: are subject relations a priori “given”? Are they established empirically? Are they dependent on context (and thus socially and historically relative)? Are they partly determined by pragmatic and political factors?

Partly inspired by bibliometrics, Hjørland (1992) proposed an understanding of “subject” as the informative or epistemological potentials of documents. According to this view, documents do not “have” subjects but are assigned subjects by somebody in order to facilitate the implicit or explicit goal underlying this assignment. “Subjects” are thus relative in terms of different human goals and interests (and citation analysis is an important tool for mapping subject relations relative to such different interests and paradigms). A given bibliographical reference may intend to refer the reader to another document on what the author considers to be the same subject. However, it may also serve other functions (for example, to show separation between subjects), and its understanding of subject-relatedness may be considered more or less adequate by the readers. The relation between citing and cited documents may therefore be considered subject relations *relative* to how references are used by authors (and the seeming strength of citation relations for information retrieval is based on the condition that authors use references to a high degree as subject designations in an adequate manner).

Semantic relations are meaning relations, i.e., relations between concepts. Typical semantic relations in thesauri and classification systems are generic relations, part–whole relations, synonym and homonym relations, among others (see Hjørland, 2007, pp. 404–405 for a long list of semantic relations). My claim is that bibliometrics may provide KO with highly relevant and much needed philosophical implications. First of all, whereas many approaches to KO tend to consider concepts and their semantic relations as stable⁷ (if not as a priori relations, cf. Svenonius, 2000, p. 131),⁸ bibliometrics provides a dynamic view of concepts and semantics, which seems to be much more in accordance with the contemporary philosophy of science and with the derived views of concepts and language, for example, the views developed by Thomas Kuhn (see, e.g., Andersen, Barker, & Chen, 2006; Hjørland, 2009; Thagard, 1992).⁹ A static view of semantic relations could state that “A is a kind of X”, whereas a dynamic view would state that “A is considered a kind of X by some documents (at a given time), but is considered a kind of Y by other documents (at another time)”. In other words, changes take place in terminological structures over time and such changes are determined by developments in subject theories and what Kuhn (1962) called scientific “paradigms”.¹⁰ Such a dynamic view is opposed to the traditional view in library science emphasizing the *standardization* of KOSs, which tends to provide rather static systems.¹¹

Both subject relations and semantic relations are thus shown in a new light by being considered in relation to citation analysis. This philosophical understanding has not so far influenced attempts to examine semantic relations and subject relations between citing and cited papers, although Small (2011), for example, seems close to doing so. Harter, Nisonger, and Weng (1993) examined the semantic relations between citing and cited documents by comparing how the papers were classified or indexed by librarians or information specialists. Again, however, such a classification also needs to be validated. We are dealing with two competing views, namely the library and information specialists’ tradition and view of how to clas-

⁵ Bibliometric coupling and co-citation analysis are described in Wikipedia (<http://en.wikipedia.org/wiki/Co-citation>, January 2, 2013) as *semantic similarity measures for documents*. However, the point here is that they may often be used as such, but that it needs to be established theoretically that they are so, or rather it needs to be established when and to what extent they can be considered measures of semantic similarity.

⁶ Statistical methods such as cluster analysis, vector space models, latent semantic indexing, etc., are used in both IR approaches and bibliometric approaches (e.g., Janssens, Glanzel, & De Moor, 2007) and will not be discussed in the present paper. Suffice it to say that Cooper (2005) concluded that one cannot select empirical variables for numerical techniques for classification without a basis in domain-specific theory. This also corresponds to the following quotation: “The quality of a SOM map [self-organizing map] or an MDS [multidimensional scaling] map should be evaluated by experts in the area studied, as no objective means exist for assessing unknown domains. This opinion is shared by Tijssen [1993], [...] he [Tijssen, 1993] offers empirical data to show that the cognitive perception of a group of experts in one subject area with respect to the same map can be very diverse” (Moya-Aregon, Herrero-Solana, & Jimenez-Contreras, 2006, p. 72). Additionally, the authors stated: “we would agree with those authors who consider MDS, SOM and clustering as complementary methods that provide representations of the same reality from different analytical points of view” (p. 73).

⁷ Francis Miksa, for example, wrote: “In the end, there is strong indication that Ranganathan’s use of faceted structure of subjects may well have represented his need to find more order and regularity, in the realm of subjects, than actually exist” (Miksa, 1998, p. 73).

⁸ “Thesauri and classifications build on these [genus–species type relations], but often (despite guidelines proscribing it) go beyond them to include relationships that are syntagmatic or extralexical. Unlike lexical or definitional relationships, which are wholly paradigmatic or a priori, syntagmatic relationships are contingent or empirical. The former express tautological relationships among ideas; the latter express relational knowledge about the real world” (Svenonius, 2000, p. 131). See also Svenonius (2000, pp. 168–169).

⁹ Thagard (1992, p. 7): “From theses 1 and 2 follows the conjecture that all scientific revolutions involve transformations in kind-relations and/or part-relations”.

¹⁰ Small (2011) also recognized Kuhn’s idea of a lexical structure to represent a scientific specialty or paradigm and its importance for bibliometrics. Small’s interest in that paper was however to provide a basis for distinguishing kinds of citation motivations by utilizing terms from the text surrounding references in scientific papers.

¹¹ Such standardized systems often make internal *conventions* (e.g. to classify social psychology with sociology). Such conventions make the system more stable (and reduce the need to update the system and to reclassify documents), but this comes at a cost: the more internal conventions and standardization are used, the less the system is able to reflect developments in the domains being classified. It becomes an isolated island without contact with the surrounding world and an alienating element for users.

Co-occurrence map of *ERIC* keywords and cited authors from nine LIS journals 1998–2000. Included are the 53 most frequent occurring keywords and authors.

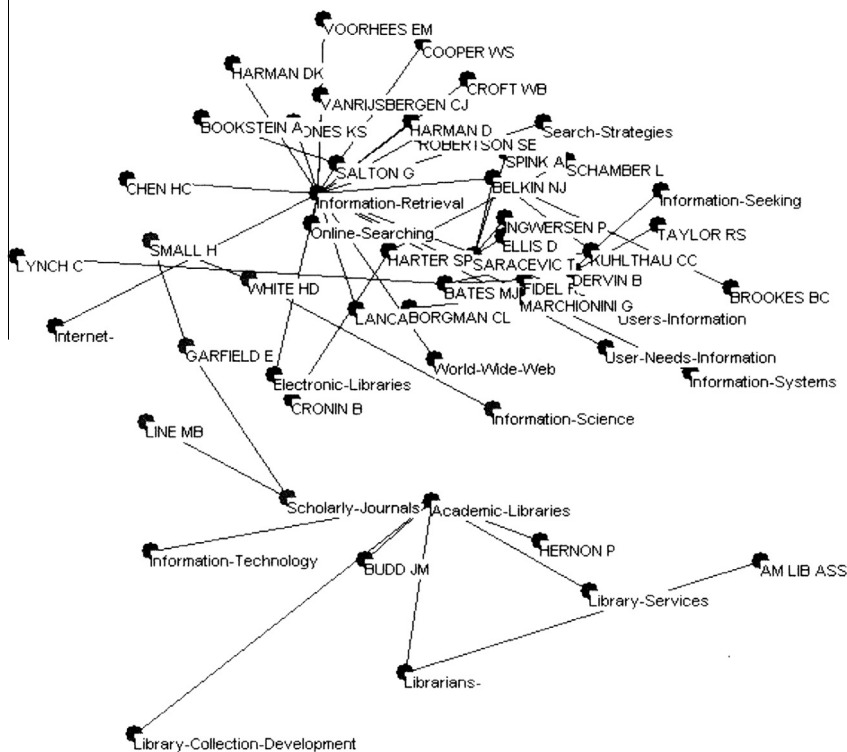


Fig. 1. A bibliometric map of LIS combining author co-citation analysis with co-word analysis (from Åström, 2002, p. 193; reprinted with permission from the author).

sify documents versus the authors' of scientific papers choice of which documents they consider it relevant to cite.¹² When we are considering bibliometrics as one among other approaches to KO, we cannot a priori assume that one of these approaches is the correct one, one that can be used as a gold standard for testing other approaches: There are no neutral points of view from which the different approaches can be compared.

There are many different ways to explore citation relations and semantic relations today. Avram, Caragea, and Dumitrache (2012) suggested an improvement to bibliometrics by introducing *citation value weighting* (by using a semantic similarity degree). Such an approach assumes that statistical similarity among documents can be taken as a measure of subject relatedness (without a discussion of this assumption). That two documents may be conceptually related although they are not statistically similar is easy to demonstrate by considering two documents about the same subject written in different languages.¹³ It is important to emphasize that from an epistemological point of view things are not just similar: Documents (and anything else) are similar in certain respects and dissimilar in other respects. What kind of similarity is relevant and how it can be measured must be qualified and for this a domain-specific theory is required (cf. note 6).

We have so far observed that citation relations, semantic relations and subject relations are three different kinds of relations and concluded that citation relations are indirect indications of subject relatedness and semantic relatedness. How well citation patterns represent KOS is an empirical question. However, the maps discussed in Section 4 (see Fig. 1) seem to be a very strong indication that maps constructed by citation analytic techniques should indeed be considered forms of KOS because they are able to map concepts and some of their relations.

¹² The KO conducted by information specialists has to serve the people using the information, including writers of papers. In a way, their bibliographical references are signs of what was needed at the time of writing. Patterns in authors' use of references are thus something that KO has to consider. The problem with doing so is mainly that the patterns are very complex and dynamic: "the pattern is new in every moment" (Eliot, 1944).

¹³ Although Braam, Moed, and van Raan (1991, p. 234) pointed out: "If different researchers work on the same set of subject-related research problems and concepts, one would expect that they use, to a relatively large extent, the same words for important concepts and problems in their speciality". Besides the problem that researchers publish in different natural languages, there is also the problem of different "paradigms" developing different terminologies.

4. Bibliometric maps

A map by Åström (2002) is shown in Fig. 1 in order to illustrate the relation between bibliometric maps and KOS. This map is the third in Åström's paper. His first map showed how the 52 most cited authors in nine LIS journals are related (their relative distances from each other as measured by co-citations). The second figure added descriptors from the ERIC database to the SCI records and clustered the 47 most frequently occurring descriptors. Åström's third map (Fig. 1) combined author co-citations and word co-occurrences in one map.

Åström (2002, pp. 191–192) remarked that Fig. 1 represents “the third part of the analysis [in which] the keywords and citations were merged and ranked, and the 53 most frequently occurring keywords and authors were selected, coupled, mapped and clustered [...] The structure of this map is basically the same as in the former two analyses [...] In this map, the separation between areas is not as clearly distinguishable as with the cited authors. But the same structures and areas can still be found, with the same location on the map”.

In bibliometrics, there are several methods for mapping documents. We have seen that Åström (2002) used co-citation analysis and word co-occurrences. These methods and a third are here considered core bibliometric methods for mapping the similarities of properties in documents (or among authors, journals, or other aggregated units). Two of these are based on citation relations:

- (1) Documents are said to be *bibliographically coupled* if they have one or more bibliographical reference in common. If document A and document B both cite document C, then A and B are bibliographically coupled (sometimes termed *retrospective coupling*).¹⁴ Bibliographic coupling strengths are counts of the number of references a set of documents have in common and a high coupling strength may be hypothesized to indicate a high degree of similarity of subject matter.
- (2) Documents are said to be co-cited¹⁵ if they appear together in the reference lists of other documents. If document C contains a reference to both document A and document B, then A and B are co-cited (sometimes termed *prospective coupling*). The co-citation frequency is defined as the frequency with which two documents are cited together: If papers A and B are both cited by many other papers, they have a stronger co-citation relationship. The more papers they are cited by, the stronger their relationship is. A strong co-citation relation may again be hypothesized to indicate a high degree of similarity of subject matter.

Another kind of relation often seen in bibliometric research and used by Åström (2002) is the relation between words, namely co-word occurrences (studied by co-word analysis).¹⁶ Regarding words in the titles of documents, in abstracts, in descriptors, in references, or in full texts:

- (1) Two words co-occur if they are used in the same records (or in the same field in the record) in a database. The number of times two words both appear in the same records (field) in the database is an indication of the co-occurrence of that set of words.

Do such different techniques provide identical maps or different maps? If they are different, how can such differences be understood and explained? We have seen that in maps based on co-occurrence compared with co-citation relations “the separation between areas is not as clearly distinguishable as with the cited authors. But the same structures and areas can still be found, with the same location on the map”. In other words, there is a degree of similarity, but the methods do not provide exactly the same results.

In general, research has so far been inconclusive in relation to measuring the relative strength or validity of various bibliometric methods (see the next section). In the rest of this paper, only methods based on citation relations will be considered further because they are here viewed as a special kind of approach to KO (whereas, for example, co-word analysis is considered more related to the methods used by the IR tradition, which is reserved for another paper).

In Fig. 1 it is most obvious that the pattern based on co-word analysis represents a kind of KOS: We have terms representing concepts and we have indications of the relative distances between these terms: The closer the terms are the closer are their meanings (i.e., a kind of semantic relation). But how can it be that a map of authors also represents a

¹⁴ The concept of bibliographical coupling was introduced by Kessler (1963), who argued for the subject relatedness of bibliographically coupled documents. See also Kessler (1965), who concluded: “This report does not pass judgment on the utility of either method to any specific application”, i.e. when bibliographical coupling should be preferred for “analytic subject indexing”.

¹⁵ The co-citation concept was constructed independently by Marshakova (1973) and Small (1973), document co-citation analysis was introduced by Small (1973), and author co-citation analysis was first used by White and Griffith (1981). “Co-citation analysis was adopted as the de facto standard in the 1970s, and has enjoyed that position of preference ever since [but] there has been a recent resurgence in the use of bibliographic coupling that is challenging the historical preference for co-citation analysis” (Boyack & Klavans, 2010, p. 2390). Co-citation analysis may be performed on different types of units: documents, authors, journals, countries (as represented by authors' addresses), and so forth. The most used type of co-citation analysis is author co-citation analysis (ACA) and it has often been employed to display what has been termed “the intellectual structure” of a specific scientific field. McCain's (1990) work is an often used standard for conducting an author co-citation analysis (ACA).

¹⁶ Co-word analysis was proposed by Callon, Courtial, Turner, and Bauin (1983) as a content analysis technique that is effective in mapping the strength of association between information items in textual data.

KOS? The first thing to observe is that there is relative agreement between the two methods, indicating that maps based on co-citations seem to provide a fair match to maps based on co-word occurrences. A second argument is provided by Small (1978), who found that a scientific paper may be cited frequently over time because it is used by many authors to stand for a particular idea, such as a method or a finding. The paper thus comes to symbolize that particular method or finding as a concept; evidence that this is so can be gleaned from the co-text surrounding the citation itself in the body of the paper:

[A]s a document is repeatedly cited, the citers engage in a dialogue on the document's significance. The verdict or consensus which emerges (if one does) from this dialogue is manifested as a uniform terminology in the contexts of citation. Meaning has been conferred through usage and what is regarded and accepted as currently valid theory or procedure has been socially selected and defined (Small, 1978, p. 338).¹⁷

In citation-based maps, authors may thus be understood as concept symbols and author names can be considered equivalent to concepts. Therefore maps based on citation analysis may be considered forms of KOS. We still have to explore, however, the relative merits of co-citation analysis and bibliometric coupling, as well as the relative merits of citation-based KOS relative to other kinds of KOS.

5. Bibliographical coupling versus co-citation analysis

A number of scholars have addressed the problem of whether bibliographic coupling and/or co-citation are good indicators of subject relatedness. Small (1973) found that bibliographical coupling and co-citation analysis provided significantly different patterns, and suggested that bibliographic coupling is a less reliable indicator of subject similarity than co-citation. Small mentioned different kinds of relations that co-citations may reflect. Co-citations may

- (1) be analogous to a measure of descriptor or word association (p. 265);
- (2) reveal relationships that are strongly recognized by people in the specialty (which may be recognized explicitly in the papers);
- (3) measure subject similarity (p. 267);
- (4) reflect the "semantic" relations among cited papers;
- (5) identify the "core" literature in a specialty.

These relations are not, however, all clearly defined by Small (1973): No data or speculations are provided concerning the validity and reliability of subject relatedness, the conditions under which bibliographic coupling or co-citation may be a good indicator of subject relatedness. Concepts such as "subject relatedness" and "semantic relations" are used very vaguely, without any hints concerning their empirical operationalization.

The relation between bibliographic coupling, co-citation analysis, as well as other kinds of network relations has since been reconsidered. Among the studies to do so are those by Boyack and Klavans (2010), Jarneving (2005), and Yan and Ding (2012).

Boyack and Klavans (2010) found that bibliographic coupling slightly outperforms co-citation analysis but that a hybrid approach that couples both references and words from titles/abstracts improves upon the bibliographic coupling. The levels of accuracy were compared by using two metrics – within-cluster textual coherence as defined by the Jensen–Shannon divergence and a concentration measure based on the grant-to-article linkages indexed in MEDLINE. The textual coherence measure is based on clusters of documents with similar sets of words in which a less diverse set of words will have a lower divergence. The authors wrote: "Given that a textual coherence is likely to favor text-based solutions over citation-based solutions, we needed a second accuracy measure, and one that was less biased toward either text or citation" (p. 2399). The grant-to-article measure was chosen because it was considered unbiased.

While Boyack and Klavans's (2010) work represented an original solution to overcome a difficult methodological problem (and introduced an important additional criterion for the measurement of citations), no measures are unbiased. If, for example, North American Grant Nos. dominate in the text corpus, then the citation of North American articles might perhaps indirectly be favored by applying this measure.

Jarneving (2005) compared bibliographically coupled documents with co-cited papers and found that the research front was portrayed in two considerably different ways depending on the methods applied. It was concluded that the results in this study would support a further comparative study of these methods at a detailed level and on a more qualitative ground.

Yan and Ding (2012) found that topical networks and coauthorship networks have the lowest level of similarity; co-citation networks and citation networks have a high level of similarity; bibliographic coupling networks and co-citation networks have a high level of similarity; and co-word networks and topical networks have a high level of similarity.

¹⁷ "It should also be mentioned that "books tend to have lower degrees of uniform usage than research papers, probably due to their greater diversity of content" (Small, 1978, p. 337).

However, no measure was applied to establish the relations between forms of citation measures and subject relatedness, only a measure of the statistical similarity of different kinds of networks. By applying network theories to citation analysis, the study was, however, able to capture the complexity of research communication and scholarly interaction more precisely than traditional bibliometric mappings.

The literature thus displays divergent findings concerning the relative “validity”¹⁸ of bibliographic coupling and co-citation. This lack of a concrete conclusion may be caused by the lack of philosophical perspective formerly introduced.¹⁹

Understanding of bibliographical coupling could probably benefit from taking as its point of departure White's (2001) concept of “an author's citation identity”, that is the researchers' individual profiles in selecting references for their publications over time. To understand bibliographical coupling is thus to understand the degree of overlap in different authors' citation identity (including their degree of individuality or ego-centeredness). Such an overlap may partly be determined by differences in domains (as further discussed below in relation to Whitley, 2000): In some fields, authors have high degrees of freedom in selecting research problems, research methods, and, by implication, relevant literature. In other fields, they are much more restricted by collectively developed norms and conventions. Citation identities should therefore display greater variability in some domains than in others; as such, they are not just a psychological tendency by individuals and should thus not primarily be studied through psychological approaches, but by studies of scholarly fields. Citation identities are expected to be less “ego-centered” in mature disciplines and to display greater variability in disciplines labeled “fragmented adhocracies” by Whitley. Consequently, the study of citation identities and bibliographical coupling might benefit from a kind of sociological study in the manner of Whitley (2000).

To understand co-citation patterns is by contrast to understand the reception history and scholarly impact of documents. Each document among all the documents ever produced may potentially be relevant to existing and future researchers and may therefore potentially be cited by them. What determines whether or not a given paper is found relevant and cited is first of all determined by current research interests and theory.²⁰ Developments in scholarly theories determine what is cited, but also why papers are co-cited or not. If, for example, Thomas Kuhn is considered important in order to understand co-citation patterns, then it should be expected that Kuhn is co-cited with bibliometric authors, for example Howard D. White.²¹ If Kuhn's view is later abandoned, this co-citation relation should decrease.

This understanding of bibliographic coupling and co-citedness may explain the divergent findings concerning the relative validity of these methods in the literature: These methods measure different things and their interpretation has to be undertaken in relation to a specific analysis of the kind of conceptual developments over time. Such interpretation presupposes subject knowledge and the need for bibliometric patterns to be verified by experts is frequently mentioned in the literature, e.g., by Yan and Ding (2012, p. 1325).

6. The intellectual and the social organization of the sciences

In order to understand a major difference between traditional KOS and KOS based on citation analytic methods, the distinction between the intellectual and the social organization of the sciences seems to be important.²² An academic discipline is both a body of intellectual knowledge and a social unit²³:

- The intellectual aspects of knowledge are organized in concepts, propositions, models, theories, and laws. *Such intellectual organizations are primarily structured via relations of explanatory coherence* (Thagard, 1992, p. 9), which are again primarily related to questions concerning truth.

¹⁸ The concept “validity” presupposes that there is a correct representation, which is an understanding that will be considered problematic in this article.

¹⁹ It should be acknowledged, however, that some of the classic bibliometric researchers in particular, e.g., Henry Small, did explore bibliometrics by considering the dynamics in the fields they mapped. Small, for example, sometimes read the physical literature he mapped in order to interpret his findings in greater depth.

²⁰ Taking myself as an example, my co-citations seem to be determined primarily by which topics interest most researchers in information science (e.g., the concept of information) although I belong to a group of researchers arguing for the concept of documents. My bibliographic coupling, on the other hand, is determined more by my individual “citation identity” (e.g., favoring references to document theory and epistemology) and thus relating to authors with similar citation identities.

²¹ Such co-citation should be expected at least for a period of time, after which it may decrease due to the phenomenon known as “obliteration by incorporation” (McCain, 2012).

²² This distinction is inspired by the title of Whitley's book (1984, 2000), which did not, however, define the terms “intellectual organization” and “social organization” of the sciences. In an email, Whitley wrote on January 2, 2013: “The short answer is that I did not bother to specify these terms because at the time, the early 1980s, there seemed little need to do so. Broadly speaking, intellectual organization refers to the structure of ideas, concepts, everyday research practices, intellectual strategies etc. that constitute scientific fields, while the social organization refers to the socio-economic environment in which research is conducted, including employment relations, formal organizational structures, resource allocation procedures and control, careers and reputational systems. Empirically, of course, this distinction is difficult to maintain, but it served to clarify the analytical distinctions I was concerned to make and the nature of the causal processes involved”. Recently Guns (2013) also proposed the social dimension (people and groupings of people) and the epistemic or cognitive dimension (topics and ideas) in addition to the documentary dimension (documents) as the entities and relations studied by informetrics.

²³ These two aspects might also be termed the “content knowledge”/“cognitive aspect of knowledge” versus the “institutional aspects” of scholarship, i.e., the professional forums.

- The social aspects of knowledge are organized into academic departments, disciplines, cooperative networks, administrative bodies, etc. *Such social organizations are primarily structured by the social division of labor in societies*, which are again mainly related to questions concerning social relevance, authority, and power.²⁴

We thus have two kinds of KO driven by criteria that may support or oppose each other in complex mutual interactions. Toulmin (1972), for example, suggested that science is generally continuous because either the content or the institution will remain stable while the other changes. In response, then, the first will adapt, in an iterative process of constant change and constant stability.

A given intellectual organization of knowledge is as stable as the knowledge and theory on which it is based: When theories change, KO should be updated accordingly. We can see such changes in the history of scholarly taxonomies, such as the biological taxonomy, the periodical system, and other classifications. A given social organization of knowledge, on the other hand, is as stable as the power relations and interests that support it. Such changes can be seen, for example, in the organization of academic units, cooperative patterns among researchers, and in bibliometric maps based on citation relations.²⁵

Traditional KOSs are to a high degree based on intellectual organization: Many classes and semantic relations in such systems are representations of, for example, biological taxonomy, the medical classification of illnesses, the periodical system of chemistry and physics, or geographical structures or other kinds of intellectual organization. These are based on models of reality and represent ontological structures, which organize (parts of) the world according to our scholarly and public knowledge.

Citation-based methods, on the other hand, are models of patterns in scientific communication and organization: They are social models, displaying the social structures among scientists and scholars (cf. Rousseau, 2008). In Fig. 1 we can see clusters of researchers (e.g., a bibliometric cluster with Small and Garfield, an IR cluster with Salton and van Rijsbergen, and a library research cluster with Hernon and Budd). These clusters represent social organizations of researchers working in the same specialties *and the concepts displayed in the same figure also reflect this social organization*. Bibliometric methods are important for showing developments in research fields. Zhao and Strotmann (2008), for example, updated the White and McCain (1998) study on information science for the years 1996–2005. This time period was considered particularly significant in that it was the first decade of the rise to prominence of the World Wide Web and allows us to glimpse its effects on the IS field.

This example demonstrates how the dynamics of scholarly fields can be modeled by methods based on citation analysis. It is, however, different from an intellectual KO:

The fact is that traditional classification involves structures that cannot be produced by any empirical analysis of the documents (or of the users for that matter). A geographical structure, for example, places different regions in a structure that is autonomous in relation to the documents that are written about those regions. You cannot produce a geographical map of Spain by making, for example, bibliometric maps of the literature about Spain [yet such autonomous structures as maps of Spain are often very useful for information retrieval about Spain] (Hjørland, 2002, p. 452).

Intellectual KO seems thus not to be superseded by bibliometric maps. But how should we understand the relative importance of intellectual versus social approaches to KOSs, and when – and to what degree – are citation-based methods able to reflect ontological models? To understand when and to what degree approaches to KO based on citation analyses overlap with KO based on intellectual methods is important in order to understand the limitations and potentialities of each approach.

Although research may improve our understanding of the relation between KO based on bibliometrics versus KO based on ontological models, we cannot expect a bibliometric map ever to correspond fully to an ontological model: There are always more factors determining social organization than pure theoretical models display. In general, bibliometrics is supposed to be the strongest in displaying trends in specific fields as well as in scholarship in general, whereas KO based on ontological models may provide more explicit semantic relations between terms.

7. Epistemological issues

This section presents two theses: (a) that citation analysis provides KO with a historical perspective which is fundamentally distinct from “similarity” perspectives, and (b) that no KOS can be considered neutral. Therefore KOS based on citation analysis should also be considered tools that support some views, goals and interests at the expense of other views, goals and interests.

²⁴ Journals (and publishers) form parts of the social structure, although the opposite has been claimed: Leydesdorff (2007, p. 25) wrote: “In science studies, this operationalization of the intellectual organization of knowledge in terms of texts (journals) as different from the social organization of the sciences in terms of institutions and people would enable us to explain the scientific enterprise as a result of these two interacting and potentially co-evolving dimensions”. Ni, Sugimoto, and Jiang (2013), on the other hand, confirmed my social understanding: “These author communities comprise all the authors who have submitted to the journal. These authors, and their conceptual markers, facilitate in creating the intellectual and social identity of this journal. Therefore, grouping journals by their shared author profiles may provide evidence of an underlying social and intellectual community”. Concepts such as scholarly terminology, special language, and genres seem to a higher degree to bridge this cognitive–social dichotomy.

²⁵ Examples of studies of social KO are those by Oleson and Voss (1979) and Wallerstein et al. (1996).

7.1. The historical perspective

Citation analysis can be compared to the paradigm shift in biological taxonomy over recent decades. The classical approach to biological classification (exemplified by the Linnaean taxonomy) is based on classifying organisms *on the basis of shared properties* (e.g., number of stamens), that is to classify according to similarity of certain properties. Cladism represents a paradigm shift in biology in which organisms are *classified solely on the basis of a common ancestor* (what Ereshefsky, 2001, called “the historical approach”). This new approach has made fundamental changes in the classification of plants and animals and this revolution is not yet complete. In the same way as cladism represents a revolution in biological taxonomy, citation analysis may be considered a revolution in KO and information retrieval. Both are based on a historical rather than a structural approach to classification. The implication for KO is that the domains and scholarly traditions to which documents belong are considered their most important criteria of classification (rather than, for example, their statistical word patterns). Scholarly theories determine what is to be considered related and different theories imply different criteria of relatedness. Thus:

Co-citation patterns change as the interests and intellectual patterns of the field change (Small, 1973, p. 265).

One way to implement this historical understanding is to bring historical studies of science and conceptual changes into play. In order to interpret co-citation patterns, it is necessary to study the history of intellectual changes in the field (for example, two papers in the bibliometric tradition are seen as more related than two papers in different traditions, say the facet analytic tradition and the bibliometric tradition). *The relations between papers in a certain tradition are used as criteria of subject relatedness rather than just classifying documents on the basis of shared properties.* It should be said, however, that citation-based approaches are sometimes used in an ahistorical way in which sets of documents are classified according to statistical similarity based on shared references or co-citations. In such cases, citation-based techniques are used as similarity measures as in mainstream IR. The historical perspective is not yet mainstream: it represents potential, but still has to win ground. Epistemologically, bibliometrics may still be driven by traditional empiricist/positivist ideals, but bibliometrics also introduces historicism as an epistemology to the field of KO.

7.2. KOSs cannot be neutral

Is it possible to construe a neutral, objective KOS? Or are KOSs necessarily tools created in order to support some goals and values at the expense of others? The first view corresponds to the traditional positivist/empiricist epistemological positions, whereas the latter view corresponds to pragmatic and critical epistemologies and philosophies of science (see also Hardeman, 2012). A precondition for using bibliometrics in accordance with the pragmatic/critical philosophy is first and foremost to realize that the literature on which bibliometrics is performed is not one neutral body of findings, but a merging of different points of view, traditions, “paradigms”, etc. The next thing to realize is that bibliometric mapping cannot be neutral in relation to such underlying views, but that any specific map tends to support some views at the expense of others.

It has often been implied that bibliometric maps are objective (Börner, Chen, & Boyack, 2003, p. 217; Silva & Teixeira, 2012). When Silva and Teixeira (2012, p. 616) claimed that bibliometric techniques “arguably rely less on the judgments and perceptions of researchers, and have a higher degree of certainty”, this is correct in the sense that the same maps can be constructed by different researchers using exactly the same techniques and data sets but still reveal the ideal of objectivity. In some contrast, Small (1999, p. 799) wrote in relation to maps of science: “Rather, it is a structure we impose on a collection of objects” (i.e., not an objective structure we discover).

To the extent that this view is maintained by Small and other bibliometric researchers, we could say that it lives up to the ideals of pragmatism and critical theory. The overall impression is that most bibliometrics, including most of Small’s writings, do not correspond to this view (Small’s point was only about whether maps should be two-dimensional or N-dimensional, not about whether two documents are related or not). The overall tendency in bibliometrics seems to be to discover structures rather than to construe structures that are in accordance with specific goals and values. Small (1999) again came close to the pragmatic/critical understanding when he said:

The choice of what coupling measure to use, of course, depends on the goals of the analysis. For a mapping of current papers the analyst might elect to use BC [bibliographic coupling] only. If the goal is to map older key papers from a current perspective, the best choice might be a co-citation (p. 802).

The insight that co-citations may “map [...] papers from a [...] perspective” is extremely important and it represents an expression of the pragmatic/critical view that papers are always written as well as read and cited from some specific perspectives. He also wrote: “Co-citation patterns change as the interests and intellectual patterns of the field change” (Small, 1973, p. 265). However, more than old papers and current perspectives are at play: There are competing views in the old papers and there are competing contemporary perspectives from which older papers can be viewed. By implication, any set of publications used for bibliometric purposes will, more or less, be a merging of different theoretical positions/points of view (cf. Hjørland, 1998). Many technical choices are made during the construction of bibliometric maps, and the claim here is that these choices have important implications in relation to which goals are best served and which goals are suppressed.

One issue that is particularly important in this respect is the selection of the documents on which the bibliometric maps are based. Imagine that we are going to create a map of LIS. As Åström (2002) showed, former maps, such as that of White and McCain (1998), seem to have a bias towards information science. In order to provide a better alternative, Åström also included more library-oriented journals in his study. However, there is no objective criterion for judging which documents best represent LIS, and *any selected set of journals can always be shown to have a bias in some direction or another*.

Both White and McCain (1998) and Åström (2002) were explicit about which journals²⁶ they used in their studies. However, the claim put forward here is that they did not make explicit arguments for how the journals were selected *in relation to their conception of the field*. It is as if the authors' view of what information science is and should be is considered "obvious" or of no consequence. As a result, their selection of journals is not based on arguments about which aspects of information science are being favored and which are being suppressed. This is not an insignificant complaint because such maps are extremely vulnerable in relation to such choices. The whole idea of a non-biased set of journals belongs to positivist ideals of science, which are simply untenable from the pragmatic point of view. We have here a kind of hermeneutic circle: How can we identify a field by a set of journals, a set of departments, a set of scholars, etc., unless we already know the field? And how can we know the field unless we know its journals, its research institutions, and its leading scholars? The answer is not that it is hopeless, but that it requires an iterative process.²⁷ Information science consists of a number of research traditions, metatheories, paradigms, etc.

It might be argued, however, that this problem can be avoided by considering another kind of map. Klavans and Boyack (2011) argued that "local" maps (e.g., White & McCain, 1998; Åström, 2002) are less accurate than "global" maps in which a single domain is mapped in the context of all scholarly disciplines. They demonstrated convincingly that even the clustering within a given domain may change if the domain's relations to other domains are considered. Based on limited data sets, they claimed that they are able to improve the patterns revealed by traditional local maps.²⁸

Klavans and Boyack's (2011) argument is based on the assumption that we are not dealing with different kinds of representations, just with more or less *accurate* representations of the same thing. They rest on the claim that maps based on all the available information are more accurate than maps based on only a fraction of that information. A counterargument can be based on the thesis that one needs to take into consideration *the nature* of the citations. LIS, for example, is a field struggling between computer-related and cultural-related views. From the cultural perspective, many citations from computer science may influence global maps of IS in a way that represents a biased view. Cultural people may say that in local maps cultural studies are better represented, and may not find that Klavans and Boyack (2011) provide a more accurate map of LIS.²⁹

Schneider (2004) provided a further development of Rees-Potter's research (1989, 1991) and demonstrated that candidates for thesaurus terms may be produced by means of bibliometric methods. One of the innovations in this research is the application of an advanced parser to identify noun phrases in small windows by citations in the text. This method is clearly an example of the principle of literary warrant (first formulated by Hulme, 1911), and as such is a very explicit application of that principle.

... the case study of periodontology clearly demonstrates that the applied bibliometric methods of co-citation analysis and citation context analysis are able to select important candidate thesaurus terms. ... We believe that the special

²⁶ The choice is not just a matter of selecting journals, but also choosing other kinds of documents or works. Often journals are chosen simply because they are indexed, but this produces serious "bias" in relation to fields such as computer science in which conference papers dominate, or in relation to history in which monographs dominate. Also, within a given field, the choice of journals at the expense of monographs may favor some paradigms (such as cognitivism, which is journal-centered), at the expense of others (such as psychoanalysis, which is to a greater extent monograph-centered). Finally, to consider a given journal, such as JASIST, a representation of a field is also problematic, because, as demonstrated by Chua and Yang (2008), "Top authors [in JASIST] have grown in diversity from those being affiliated predominantly with library/information-related departments to include those from information systems management, information technology, business, and the humanities". Therefore, bibliometric maps based on JASIST cannot simply be taken to represent the library/information field without further examination.

²⁷ A person working in the facet analytic tradition of KO would miss, among others, S.R. Ranganathan on White and McCain's (1998) map. Exploring the journals used by White and McCain (1998) shows that Ranganathan's absence is *partly* due to the elimination of journals in which Ranganathan is highly cited, such as Libri (14.8% of the references to Ranganathan), Aslib Proceedings (14.3%), and International Classification (09.7 + 07.6% = 17.3%) in the period up to and including 1995 (more journals citing Ranganathan are not included here). That is, more than 46.4% of the citations to Ranganathan were excluded by White and McCain, although these references were in the database (i.e., in addition to the bias implied by the database coverage). The omission of these journals reflects a view of information science that downgrades the tradition of facet classification. The main argument here is that it is not done explicitly.

²⁸ They mentioned, however, that today's science does not have at its disposal computer algorithms powerful enough to process all the data in the Thomson Reuter or Scopus databases (today's limit is 2,500,000 concepts compared with the necessary 10^8 (1,000,000,000) concepts). Moreover, we could add that even when this is achieved, these databases still do not represent the total world literature. The idea that these databases reflect in a non-biased way the most important literature may also be a problematic assumption (as the whole idea of an objective hierarchy of journals in each discipline within which each scholar competes to publish is problematic, cf. Andersen, 2000).

²⁹ As another example, Karl Marx was the most cited author in the Arts and Humanities Citation Index in 1977–1978. Garfield (1980, p. 53) wrote: "The appearance of Marx, Lenin, and Engels on the list may be surprising. It reflects our definition of the humanities and the resulting composition of our database. Half of the citations to Marx come from philosophy journals, with nearly two-thirds of these from one journal – *Deutsche Zeitschrift für Philosophie*. . .". Of course it is an objective fact that Marx was the most cited author in this database at that time. However, it is also the case that the distribution of citations to his works is extremely skewed. Unless one is told that nearly two-thirds of the philosophical references came from one journal from the former Soviet Union, one would gain the wrong impression of Marx's influence in the humanities internationally. As Garfield (1980, p. 53) wrote, "It reflects our definition of the humanities and the resulting composition of our database". My point here is that more information does not necessarily make for a more accurate map and that the idea of an accurate map seems problematic (but to produce such maps and accompany them with adequate interpretations is highly relevant, of course). An argument could perhaps also be that the global map tends to introduce a Matthew effect (i.e., the theory of cumulative advantage) whereby minority views are misrepresented.

selection procedures inherent in the methodical steps of the two components ensure that a significant number of the selected primary candidate thesaurus terms turn out to be important index terms. Hence, the conclusion is that the applied bibliometric methods are very suitable for selection of candidate thesaurus terms in the specialty area of periodontology (Schneider, 2004, p. 323).

What Schneider demonstrated was that it is possible to identify by bibliometric means terms in the literature that also exist in thesauri such as MeSH[®].³⁰ He did not demonstrate that a sample of terms from MeSH[®] could *all* be identified in the literature. In other words, he demonstrated that MeSH[®] is at least partially based on concepts in the scientific literature and that some of these concepts may be retrieved by bibliometric methods. He used MeSH[®] (and the *Glossary of Periodontal Terms*) as “the gold standard” by which he evaluated the bibliometric methods. It should be considered, however, that tools such as MeSH[®] are also based on certain assumptions and should be evaluated. A knowledge organization tool such as a thesaurus or a bibliometric map is never a neutral or objective representation, but different underlying views and interests in domains demand different representations. This issue was not considered by Schneider.

The point defended in this section is that bibliometric researchers, whether they realize it or not, make subjective decisions that are important for “bias” in the maps they produce. Bibliometric studies have to be accompanied by studies of traditions and paradigms in the domains they map. For each decision, the bibliometric researcher should make clear which theoretical positions are supported and which are relatively suppressed (cf. Hjørland, 2009, p. 1527). Such subjectivity may seem uncomfortable for positivist-minded researchers but it is better to have explicit subjectivity than to have subjectivity disguised as objectivity.

8. Conclusion

Knowledge is a cultural entity and keeps shifting its pattern like a kaleidoscope.

An emergence of the new knowledge modifies the structure of the whole.

Contrary to H.E. Bliss (1870–1955) there is no permanent order in knowledge. “Pattern is new every moment” said T.S. Eliot (1888–1965), with a poetic vision.

Satija (1992, pp. 40–41), paraphrasing McGarry (1991, p. 148).

Bibliometrics is important because scientific knowledge claims are to a very great extent based on contributions published in the scholarly literature. If researchers want to provide arguments for their views, then their published arguments have a privileged status because they can be examined by other researchers who can be traced by bibliographical references. It is an important part of scholarly work to consider claims in the literature and to discuss them in relation to one’s own work. The importance of findings in the literature is not just about the truth or falsity of a claim, but also about the organization of its subject matter, which is what is represented by KOSs. For example, if nocturnal enuresis (bed-wetting) is shown to depend on psychological factors, the concept “enuresis” belongs to psychology and is thus part of the terminological structure of psychology. If, on the other hand, it is shown to belong to genetics, then it belongs to the terminological structure of genetics or physiology. Of course, it may belong to both fields, but the relative strengths of the associations are determined by current research activity, which again is related to current theory.

It is also important to realize that the scholarly literature (and not, for example, dictionaries or thesauri) are the primary sources regarding the meanings of words and other symbols used in scholarly fields (from there often spread to languages in general). Concepts are dynamically negotiated in the scientific literature (Hjørland, 2009). In order to identify scientific concepts and terms and their relation to other terms, in the end one needs to inspect the primary literature and bibliometrics is an important tool for such an inspection. Changes in conceptual structures have become an important issue in cognitive science,³¹ and it is exactly such shifts in conceptual structures that bibliometrics is well suited to map and that make it a dynamic approach to KO.

It has also been argued above that bibliometrics should be understood as a social approach to KO based on cooperation patterns among researchers (which, of course, are partly theoretically motivated). As such, it stands in contrast to KOSs based on ontological models of reality. However, the relation between social and intellectual KO is complex. There is no reason to believe that a bibliometric map may ever be able to produce intellectual structures as known, for example, from the periodical system of chemistry and physics, from biological taxonomy, from geographical maps, etc. Generally, therefore, maps based on citation analyses should be seen as supplements rather than replacements. There is, however, a further need to study the interaction of these two kinds of KOS.

We also need to consider an important distinction in the literature of KO: assigned versus derived indexing. Derived indexing is the use of words from the texts that are indexed, whereas assigned indexing is the indexer’s assignment of labels to a document. We saw above different techniques for using derived indexing in bibliometrics, above all Schneider’s (2004) use of an advanced parser to identify noun phrases in small windows around citations in the text. An important theoretical question is “Can all relevant concepts always be supposed to be in the texts which are indexed (or mapped)?” Could it be

³⁰ Whether what Schneider was performing systematically was in the past performed impressionistically by committees of subject experts is an open question. Although it is ordinary practice to use subject specialists for the development of classification schemes (in addition to the classification of each document), this activity is not reflected in the research literature.

³¹ See, for example, Andersen et al. (2006) and Thagard (1992).

that using indexing checklists (as in MEDLINE) could improve retrievability by adding conceptual distinctions that are not available in the documents? Could it be that KO is a creative act of creating new labels? We are dealing here with the epistemological question of whether to describe things passively (the ideal of objectivity) or whether to construct conceptions and labels/keywords actively (the ideal of subjectivity).

The main conclusions concerning citation analysis as an approach to KO are summarized in the following points:

(A) Advantages of bibliographic references and citations as subject access points:

- References represent a form of “literary warrant” and are thus empirically based in the scholarly literature.
- Citations are provided by researchers (highly qualified subject specialists).
- The number of references reflects the indexing depth and specificity (the average of scientific papers is about 10 references per article).
- Citation indexing is a highly dynamic form of subject representation (each new document published and indexed updates the pattern).
- References are distributed through papers, allowing the utilization of the paper structure in the contextual interpretation of citations.
- Scientific papers form a kind of self-organization system.
- Citation based maps identify groups of researchers working in the same specialties.

(B) Disadvantages of bibliographic references and citations as subject access points:

- The relation between citations and subject relatedness is indirect and somewhat unclear (related to the difference between the social and the intellectual organization of knowledge).
- Bibliometric maps do not provide a clear logical structure with mutually exclusive and collectively exhaustive classes.
- Explicit semantic relations are *not* provided (e.g. genus–species relations and part–whole relations) (but future systems may distinguish between different kinds of citation links/motivations).
- Only derived indexing is provided: Concepts not represented in the literary sample is not assigned.
- There is a tendency to mix different theoretical structures due to the merging of literatures in the samples (rather than providing a system based on a pure theoretical basis).
- Namedropping and other forms of imprecise citation may cause noise.

References

- Andersen, H. (2000). Influence and reputation in the social sciences – How much do researchers agree? *Journal of Documentation*, 56(6), 674–692.
- Andersen, H., Barker, P., & Chen, X. (2006). *The cognitive structure of scientific revolutions*. New York: Cambridge University Press.
- Åström, F. (2002). Visualizing library and information science concept spaces through keyword and citation based maps and clusters. In H. Bruce, R. Fidel, P. Ingwersen & P. Vakkari (Eds.), *Emerging frameworks and methods: Proceedings of the fourth international conference on conceptions of library and information science (COLIS4)* (pp. 185–197). Greenwood Village: Libraries Unlimited.
- Avram, S., Caragea, D., & Dumitrache, I. (2012). A new approach to bibliometrics based on semantic similarity of scientific papers. *Control Engineering and Applied Informatics*, 14(3), 35–42.
- Börner, K., Chen, C. M., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37, 179–255.
- Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389–2404.
- Braam, R. R., Moed, H. F., & van Raan, A. F. J. (1991). Mapping of science by combined co-citation and word analysis. I. Structural aspects. *Journal of the American Society for Information Science*, 42(4), 233–251.
- Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22, 191–235.
- Chen, C. (2003). *Mapping scientific frontiers: The quest for knowledge visualization*. New York: Springer-Verlag.
- Chen, C., Ibeke-Sanjuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal of the American Society for Information Science and Technology*, 61(7), 1386–1409.
- Chua, A. Y. K., & Yang, C. C. (2008). The shift towards multi-disciplinarity in information science. *Journal of the American Society for Information Science & Technology*, 59(13), 2156–2170.
- Cooper, R. (2005). *Classifying madness: A philosophical examination of the diagnostic and statistical manual of mental disorders*. Berlin: Springer.
- Dehart, F. E., & Scott, L. (1991). ISI research fronts and online subject access. *Journal of the American Society for Information Science*, 42(5), 386–388.
- Eliot, T. S. (1944). *The four quartets*. London: Faber.
- Ereshefsky, M. (2001). *The poverty of the Linnaean hierarchy: A philosophical study of biological taxonomy*. Cambridge: Cambridge University Press.
- Garfield, E. (1980). Is information retrieval in the arts and humanities inherently different from that in science? The effect that ISI's citation index for the arts and humanities is expected to have on future scholarship. *Library Quarterly*, 50(1), 40–57.
- Garfield, E., & Sher, I. H. (1993). KeyWords Plus – Algorithmic derivative indexing. *Journal of the American Society for Information Science*, 44(5), 298–299.
- Guns, R. (2013). The three dimensions of informetrics: A conceptual view. *Journal of Documentation*, 69(2), 295–308.
- Hardeman, S. (2012). Organization level research in scientometrics: A plea for an explicit pragmatic approach. *Scientometrics*, Online First™, July 20, 2012. ><http://www.springerlink.com/content/uhw3660427525277/fulltext.pdf>< Retrieved 12.01.13
- Harter, S. P., Nisonger, T. E., & Weng, A. W. (1993). Semantic relations between cited and citing articles in library and information science journals. *Journal of the American Society for Information Science*, 44(9), 543–552.
- Hjørland, B. (1992). The concept of “subject” in information science. *Journal of Documentation*, 48(2), 172–200.
- Hjørland, B. (1998). Information retrieval, text composition, and semantics. *Knowledge Organization*, 25(1/2), 16–31.
- Hjørland, B. (2002). The methodology of constructing classification schemes. A discussion of the state-of-the-art. *Advances in Knowledge Organization*, 8, 450–456.
- Hjørland, Birger (2007). Semantics and knowledge organization. *Annual Review of Information Science and Technology*, 41, 367–405.
- Hjørland, B. (2009). Concept theory. *Journal of the American Society for Information Science and Technology*, 60(8), 1519–1536.
- Hjørland, B. (2013a). Facet analysis: The logical approach to knowledge organization. *Information Processing & Management*, 49(2), 545–557.

- Hjørland, B. (2013b). User-based and cognitive approaches to knowledge organization: A theoretical analysis of the research literature. *Knowledge Organization*, 40(1), 11–27.
- Hjørland, B. (2013c). Theories of knowledge organization – theories of knowledge. In Keynote presentation at the 13th Meeting of the German ISKO (International Society for Knowledge Organization), Potsdam, March 19–20, 2013. *Knowledge Organization*, 40(3), 169–181.
- Hodge, G. (2000). *Systems of knowledge organization for digital libraries. Beyond traditional authority files*. Washington, DC: The Council on Library and Information Resources. <<http://www.clir.org/pubs/reports/pub91/contents.html>> Retrieved 12.01.13.
- Hulme, E. W. (1911). Principles of book classification. *Library Association Record*, 13, 354–358, October 1911; 389–394, November 1911; & 444–449, December 1911.
- Janssens, F., Glanzel, W., & De Moor, B. (2007). Dynamic hybrid clustering of bioinformatics by incorporating text mining and citation analysis. In *KDD-2007. Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 360–369). San Jose, California, USA, August 12–15, 2007. New York: ACM.
- Jarneving, B. (2005). A comparison of two bibliometric methods for mapping of the research front. *Scientometrics*, 65(2), 245–263.
- Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14, 10–25.
- Kessler, M. M. (1965). Comparison of the results of bibliographic coupling and analytic subject indexing. *American Documentation*, 16(3), 223–233.
- Klavans, R., & Boyack, K. W. (2011). Using global mapping to create more accurate document-level maps of research fields. *Journal of the American Society for Information Science and Technology*, 62(1), 1–18.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Leydesdorff, L. (2007). Visualization of the citation impact environments of scientific journals: An online mapping exercise. *Journal of the American Society for Information Science and Technology*, 58(1), 25–38.
- Marshakova, I. V. (1973). A system of document connection based on references. *Scientific and Technical Information Serial of VINITI*, 6(2), 3–8.
- McCain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science*, 41(6), 433–443.
- McGarry, K. (1991). Epilogue: Differing views of knowledge. In A. J. Meadows (Ed.), *Knowledge and communication. Essays on the information chain* (pp. 132–152). London: Library Association.
- Miksa, F. L. (1998). *The DDC, the universe of knowledge, and the post-modern library*. Albany, NY: Forest Press.
- Moya-Anegón, F., Herrero-Solana, V., & Jimenez-Contreras, E. (2006). A connectionist and multivariate approach to science maps: The SOM, clustering and MDS applied to library science research and information. *Journal of Information Science*, 32(1), 63–77.
- Ni, C., Sugimoto, C. R., & Jiang, J. (2013). Venue-author-coupling: A measure for identifying disciplines through author communities. *Journal of the American Society for Information Science and Technology*, 64(2), 265–279.
- Oleson, A., & Voss, J. (Eds.). (1979). *The organization of knowledge in modern America, 1860–1920*. Baltimore: Johns Hopkins University Press.
- Pao, M. L. (1993). Term and citation retrieval: A field study. *Information Processing & Management*, 29(1), 95–112.
- Pao, M. L., & Worthen, D. B. (1989). Retrieval effectiveness by semantic and pragmatic relevance. *Journal of the American Society for Information Science*, 40(4), 226–235.
- Rees-Potter, L. K. (1989). Dynamic thesaural systems: A bibliometric study of terminological and conceptual change in sociology and economics with application to the design of dynamic thesaural systems. *Information Processing & Management*, 25(6), 677–691.
- Rees-Potter, L. K. (1991). Dynamic thesauri: The cognitive function. Tools for knowledge organization and the human interface. In *Proceedings of the 1st International ISKO Conference*, Darmstadt, August 14–17, 1990 (Part 2, 1991, pp. 145–150).
- Rousseau, R. (2008). Publication and citation analysis as a tool for information retrieval. In D. Hoh & S. Foo (Eds.), *Social information retrieval systems: Emerging technologies and applications for searching the web effectively* (pp. 252–268). London: Information Science Reference.
- Salton, G. (1971). Automatic indexing using bibliographic citations. *Journal of Documentation*, 27(2), 98–110.
- Satija, M. P. (1992). Book review of Meadows (1991): Knowledge and communication: Essays on the information chain. *International Classification*, 19(1), 39–41.
- Schneider, J. W. (2004). *Verification of bibliometric methods' applicability for thesaurus construction*. Aalborg: Royal School of Library and Information Science [PhD dissertation]. <http://pure.iva.dk/files/31034882/jesper_schneider_phd.pdf> Retrieved 12.01.13.
- Silva, M. C., & Teixeira, A. A. C. (2012). Methods of assessing the evolution of science: A review. *European Journal of Scientific Research*, 68(4), 616–635. <http://www.europeanjournalofscientificresearch.com/ISSUES/EJSR_68_4_15.pdf> Retrieved 12.01.13.
- Small, H. G. (1973). Co-citation in the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 256–269.
- Small, H. G. (1978). Cited documents as concept symbols. *Social Studies of Science*, 8(3), 327–340.
- Small, H. G. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, 50(9), 799–813.
- Small, H. G. (2011). Interpreting maps of science using citation context sentiments: A preliminary investigation. *Scientometrics*, 87(2), 373–388.
- Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge, MA: MIT Press.
- Thagard, P. (1992). *Conceptual revolutions*. Princeton: Princeton University Press.
- Tijssen, R. J. W. (1993). A scientometric cognitive study of neural network research: Expert mental maps versus bibliometric maps. *Scientometrics*, 28(1), 111–136.
- Toulmin, S. (1972). *Human understanding. The collective use and evolution of human concepts*. Princeton, New Jersey: Princeton University Press.
- Vargas-Quesada, B., & de Moya Anegón, F. (2007). *Visualizing the structure of science*. Berlin: Springer.
- Wallerstein, I., Juma, C., Keller, E. F., Kocka, J., Lecourt, D., Mudimbe, V. Y., et al (1996). *Open the social sciences: Report of the Gulbenkian Commission on the restructuring of the social sciences*. Stanford, CA: Stanford University Press.
- White, H. D. (2001). Authors as citers over time. *Journal of the American Society for Information Science and Technology*, 52(2), 87–108.
- White, H. D., & Griffith, B. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32(3), 163–171.
- White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science*, 49(4), 327–355.
- Whitley, R. R. (2000). *The intellectual and social organization of the sciences* (2nd ed.). Oxford: Oxford University Press (First edition published 1984).
- Yan, E., & Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and cword networks relate to each other. *Journal of the American Society for Information Science and Technology*, 63(7), 1313–1326.
- Zhao, D., & Strotmann, A. (2008). Information science during the first decade of the web: An enriched author co-citation analysis. *Journal of the American Society for Information Science & Technology*, 59(6), 916–937.