Jochen Gläser[1] Wolfgang Glänzel[2,3] Andrea Scharnhorst[4]

Same data – different results?
Towards a comparative approach to the identification of thematic structures in science

Abstract: Science studies are persistently challenged by the elusive structures of their subject matter, be it scientific knowledge or the various collectivities of researchers engaged with its production. Bibliometrics has responded by developing a strong and growing structural bibliometrics, which is concerned with delineating fields and identifying thematic structures. In the course of these developments, a concern emerged and is steadily growing. Do the sets of publications, authors or institutions we identify and visualise with our methods indeed represent thematic structures? To what extent are results of topic identification exercises determined by properties of knowledge structures, and to what extent are they determined by the approaches we use? Do we produce more than artefacts? These questions triggered the collective process of comparative topic identification reported in this special issue. The introduction traces the history of bibliometric approaches to topic identification identifies the major challenges involved in these exercises and introduces the contributions to the special issue.

1. Introduction

Science studies are persistently challenged by the elusive structures of their subject matter, be it scientific knowledge or the various collectivities of researchers engaged with its production. The production of scientific knowledge is a self-organising process in which researchers recombine existing knowledge in their production of new knowledge claims. In this process, interpretations of knowledge and frames of reference in which interpretations occur are constantly changed, shared, or disassociated. The collectives (groups, networks, communities) involved in the production of knowledge evolve concurrently in a stream of temporary and partial memberships, thereby creating the fluid cognitive and social structures we commonly refer to as fields, specialties, or topics.

The identification and delineation of these socio-cognitive structures has been a central problem of at least three weakly related strands of scholarship: library and information science; sociology of science, and bibliometrics.[5] The oldest field dealing with these problems is librarianship, which has to fulfil its two central tasks of curation (keeping collections current and relevant) and information retrieval under conditions of a rapidly growing internal differentiation of science and hybridisation of fields. Books and bibliographies need to be brought under subject headings, which means that disciplines and fields

---

[1] Jochen.Glaser@ztg.tu-berlin.de; Center for Technology and Society, TU Berlin, Hardenbergstr. 16-18, 10623 Berlin, Germany

[2] wolfgang.glanzel@kuleuven.be; ECOOM and Department of MSI, KU Leuven, Naamsestraat 61, 3000 Louvain, Belgium

[3] glanzw@iif.hu; Department of Science Policy and Scientometrics, Library of the Hungarian Academy of Sciences, Arany Janos u. 1, Budapest 1051, Hungary

[4] andrea.scharnhorst@dans.knaw.nl; Data Archiving and Networked Services, Royal Netherlands Academy of Arts and Science (KNAW), Amsterdam, The Netherlands

[5] There are certainly more, e.g. topic modelling, which we will briefly discuss later (section 2).

must be reflected in science classifications (Kedrow 1975). The problem of changing socio-cognitive structures has also been addressed in the field of knowledge organization. Here, the problem appears under the label of "domain analysis" and is discussed in terms of the extraction of "concepts" rather than topics. The aim is to generate adequate knowledge organisation systems, i.e. taxonomies, controlled vocabularies, classifications, etc. (see Smiraglia 2015).

The sociology of science noticed early on that the processes of knowledge production it aims to explain vary considerably between fields, and introduced distinctions between 'hard' and 'soft' or 'basic' and 'applied' disciplines. The reception of Kuhn's (1962) model of paradigm and corresponding community development led to more fine-grained perspectives, which considered (but rarely compared) scientific specialties (Chubin 1976).

Bibliometrics has responded to these theoretical concerns and to political interests in selectively promoting societally relevant, interdisciplinary and lately emerging fields. It developed a strong and growing structural bibliometrics, which is concerned with delineating fields and identifying thematic structures. A similar and now often coinciding interest has been developed in applied bibliometrics. Bibliometric studies have observed that field-specific research practices are reflected in rhetorical, publication and citation practices. The resulting differences necessitate field-specific bibliometric indicators, measurement techniques, and time windows (Moed et al. 1985; Glänzel and Schoepflin 1995; Moed 2005). Field delineation has become an important tool for normalising citation counts in bibliometric evaluations (e.g. Leydesdorff et al. 2016) as well as mapping emerging, inter- and transdisciplinary fields (e.g. Leydesdorff and Bornmann 2011; Small et al. 2014).

The rise of bibliometric methods for field delineation and topic identification has been supported by the strong growth of the field of bibliometrics in general and by the increase in computing power. Over the last three decades, faster and more complex algorithms of language processing, network analysis, and visualization have been developed and applied to the problem of topic and field delineation. All three kinds of algorithms are also indicative of the import of methods from other fields such as linguistics, informatics, or physics.

In the course of these developments, a concern emerged and is steadily growing. Do the sets of publications, authors or institutions we identify and visualise with our methods indeed represent thematic structures? To what extent are results of topic identification exercises determined by properties of knowledge structures, and to what extent are they determined by the approaches we use? Do we produce more than artefacts? To our knowledge, these questions have first been raised by Leydesdorff (1987). Unfortunately, the methodological reflexivity in bibliometric research they should have triggered did not emerge at that time. The questions' persistent relevance is illustrated by a recent discussion of the tenuousness of the link between 'ground truths' about nodes and communities in networks. Between them, Hric et al. (2014) and their critics (Peel et al. 2016) came up with several reasons why topological community detection algorithms might not be able to reconstruct 'ground truths' about nodes. There might be more than one ground truth corresponding to the same network, these ground truths might be differently linked to the metadata describing network nodes, the network's community structure might not reflect the structure of the metadata representing a particular ground truths, or the community detection algorithm might perform poorly. These reasons are difficult to disambiguate, if at all.

Questions about the respective influences of real-world properties and methods on topic identification triggered the collective process of comparative topic identification that we report in this special issue. The process emerged from a project on measuring the epistemic diversity of research (Gläser et al. 2015). In order to measure the epistemic diversity of a research field, this field must be delineated, and topics in the field must be identified. In contrast to many other purposes of field delineation and topic

identification, the measurement of diversity is affected by variation in the delineation of fields or topics. This is why the discussion between the project and its advisory board soon focused on the questions formulated above. We invited further colleagues to a series of workshops on approaches to topic identification. At some point, the idea emerged to learn about our approaches by applying them to the same data set and to compare the outcomes.

The aim of this collaboration, whose results are presented in this special issue, is to use comparisons as a source of learning about one's own approach. Our premise is that there is not one 'best way' of identifying topics for two reasons. First, the structure of a body of knowledge is in the eye of the beholder, i.e. more than one valid thematic structure can be constructed depending on the perspective applied to the knowledge. Second, topical structures are reconstructed for specific purposes, so if at all, there might be a best method for a given purpose. This is why instead of finding the one best solution, we aim to develop a deeper understanding of the ways in which specific properties of the approaches we use create differences between our results.

The papers in this special issue address these questions in their presentation of approaches and of comparisons. Before we introduce the papers, we briefly sketch the history of topic identification (2), explain why we think that a comparative approach may help addressing key unresolved theoretical and methodological issues (3) and describe the data set used by all experiments published in this special issue (4).

2. A brief history of topic detection

The Mertonian sociology of science originated with structuralist concerns with the emergence and possibility of the social system of science in modern societies. However, questions about the social structure of and processes in the whole science system were soon enriched by questions about specific processes in specific fields. Empirical observations and models of scientific development such as Kuhn's (1962) made soon clear that scientists interact with each other quite selectively, and that a common concern with specific knowledge both directs and emerges from these interactions. The research interests of the sociology of science were oriented towards the emergence of scientific specialties (see the review of this literature by Edge and Mulkay 1976) and in the co-variation of cognitive differences between fields and their social structures (Whitley 1984). It soon became clear to sociologists of science that scientific specialties were the primary social units in which research was conducted, while disciplines remained the social forms of teaching (Chubin 1976). As for the general structure of science, several proposals of a hierarchical organization consisting of problem areas and research fields were made (e.g. Whitley 1974). Other notions such as 'invisible colleges' were also used but never particularly clearly defined in a sociological sense (Crane 1972).

The interest in field-specific processes was not accompanied by many efforts to properly delineate fields or other units. For many sociological investigations, the boundaries of fields were not important. For example, researchers studying the transmission of tacit knowledge (Collins 1974), scientific controversies (Collins 1975) or the standardization of research practices (Cambrosio and Keating 1988) only had to make sure that their empirical objects did belong to the field they were interested in. Such empirical investigations did not usually attempt to capture whole fields or scientific communities. At the same time, sociologists observed that the networks of researchers they were interested in could not be easily delineated. Neither properties of the literature nor questioning scientists led to unambiguous results (Crawford 1971; Crane 1972, pp. 41-47; Woolgar 1976). Sociologists came to the conclusion that the "research networks" they were interested in are "amorphous social groupings" (Mulkay 1977, p. 113), and that it is impossible to identify

unambiguously all members of such a grouping (Mulkay et al. 1975, pp. 188-190; Woolgar 1976, pp. 234-235; for a more recent confirmation of these findings see Verspagen and Werker 2003).

The field of bibliometrics rapidly grew at this time. The availability of the Science Citation Index (SCI) and improving access to computing power enabled efficient explorations of the intellectual structure of the sciences. During the last four decades, these explorations have grown into an important topic of bibliometrics. This development has partly been driven by three changes to bibliometrics' resources, namely the availability of electronic versions of the large abstract and citation databases (Web of Science and Scopus), the subsequently introduced online access to these databases, and the local availability of the full citation databases in major bibliometric centres, which improved these centres' opportunities to consolidate the data and develop dedicated software.

The basic strategy of all explorations of intellectual structure has remained the same ever since the first experiments by Henry Small and his colleagues (Small 1973; Griffith et al. 1974; Small and Griffith 1974; Small 1977). A set of publications is delineated, and its intellectual structure analysed with algorithms that utilise properties of these publications. The first step sometimes uses an entire database, which in itself covers only a sub-section of scientific publications. Alternatively, a sub-section of the database is created by downloads based on dedicated search strings for titles, keywords, or abstracts, lists of specified journals, or subject categories provided by the database. In some cases, this download is expanded by additionally applying citation-based methods.

The second step, the actual exploration of intellectual structure by identifying and delineating topics, was dominated for some time by co-citation clustering, which was first suggested independently by Small (1973) and Marshakova (1973) and soon used for the identification of specialties and research fronts (Small 1973; Griffith et al. 1974; Small and Griffith 1974). The 1980s saw a significant increase in the variety of properties of publications utilised and approaches utilising them. White and Griffiths (1981) introduced author co-citation, which was used for the analysis of intellectual structures of disciplines (White and McCain 1998). Callon et al. (1983) suggested co-word analysis, an approach that treated the co-occurrence of words in a document analogously to the co-occurrence of publications in reference lists. Similar to author co-citation, co-word analysis has been used widely in the exploration of thematic structures of paper sets that were otherwise retrieved (Rip and Courtial 1984; Tijssen 1992; van Raan and Tijssen 1993).

Leydesdorff suggested using journal to journal citations, which were applied by him and his colleagues both on the whole Web of Science and for the analysis of subfields (Leydesdorff 1986; Leydesdorff and Cozzens 1993; Leydesdorff 2004; Leydesdorff and Rafols 2009).

Quite surprisingly, the pendant to co-citation, bibliographic coupling, has been used for topic identification with significant delay. Although introduced early on by Kessler (1963), it was first used for topic detection by Schiminovich (1971) but his suggestion did not lead to a development similar to that of co-citation analysis. Instead, bibliographic coupling was rediscovered much later. Glänzel and Czerwon (1996) used bibliographic coupling to identify core documents. Jarneving (2001) appears to have been the second to use bibliographic coupling for topic detection. Bibliographic coupling is, however, an important part of hybrid methods that combine bibliographic coupling and text-analytic methods (Glenisson et al. 2005; Janssens et al. 2007).

Some research groups were able to exploit their access to local versions of Thomson Reuters' or Elsevier's citation databases and the opportunity to develop programmes that operate directly on these databases. Combined with increased computing power and the publication of ever more efficient algorithms, this access enabled the clustering of all publications indexed in the Web of Science (respectively Scopus) based on co-citations or direct citations (Boyack et al. 2005; Klavans and Boyack 2011; Waltman and van Eck 2012).

The most recent development in the context of topic identification is stochastic topic modelling. This approach assumes that words in a document arise from a number of latent topics, which are defined as distributions over a fixed vocabulary (Blei 2012). Topic modelling uses full texts of documents in order to identify topics, and assumes that all documents share all topics in varying proportions. The most widespread method of topic modelling, latent Dirichlet analysis, assigns words to topics according to the prevalence of words across topics and of topics in a document (Yau et al. 2014). Stochastic topic modelling is used on a wide range of document sets. While the potential of topic modelling remains to be explored, it has one feature that is somewhat at odds with the science studies approach to topic identification, namely the necessity to decide ex ante how many topics there are in a given set of papers.

In the course of this evolution of topic detection, the validity of attempts to identify topics and delineate fields has always been a concern. Considerations of validity (or "accuracy", Boyack and Klavans 2010) led to numerous partial comparisons of approaches to field delineation and topic identification, most of which either focused on similarity measures for papers (see e.g. the review in Boyack and Klavans 2010; Klavans and Boyack 2015) or on clustering methods (Šubelj et al. 2016). These papers also highlight the central problem of these exercises: Since there is no 'ground truth' available, comparisons have to use arbitrary yardsticks. In most cases, several similarity measures are compared against one that is set as universal yardstick. Klavans and Boyack (2015) use documents with more than 100 references, while Šubelj et al. (2016) analyse statistical properties of solutions and use their own expertise in scientometrics. None of these approaches indicates convergence on a shared, theoretically justified standard for establishing validity. Interestingly enough, the attempts to bring consistency to research on the identification of thematic structures appear to suffer from the disease they attempt to cure: Comparison exercises reference each other without serious engagement, and the growing body of literature does not produce a coherent body of knowledge.

Attempts to let experts from the analysed fields validate the outcome of bibliometric approaches to the identification of thematic structures illustrate the problem. If maps are presented to experts, they tend to agree that most of the structure presented to them makes sense, while some aspects of the maps do not (Healey et al. 1986; Bauin et al. 1991, p. 133; Noyons 2001; Noyons 2004). This is not surprising given that experts have to match a map given to them to one of a set of equally valid scientific perspectives. As Law et al. (1988, p. 262) put it:

> … though scientists recognize the themes of research that we have identified, and generally think that they represent a reasonable tally of acidification-relevant work, the adequacy of our assessment of the *relationship* between these themes is more difficult to test. We are currently in the process of analysing our interviews in order to derive an interview-based depiction of the structure of acidification research. This task is complicated, not only by the fact that judgements have to be made about what count as synonyms, but also because each scientist has her own special view of science. Our problem, then, is analogous to that of putting together the views of Londoners, Mancunians, Glaswegians and Aberdonians in order to create a map of the United Kingdom. It is possible, but it is not straightforward.

There are also findings of bibliometric research that point to principal limitations of attempts to delineate thematic structures. Katz (1999) and van Raan (1990; 2000) described the fractal structure of scientific knowledge as represented in publications. These findings are supported by failed attempts to identify 'natural' boundaries of fields, i.e. significant changes in bibliometric parameters that may indicate the existence of a topic's or field's boundary (Noyons and van Raan 1998; Zitt and Bassecoulard 2006).

At the current stage of this development, the challenge to the field of bibliometrics posed by the task of topic identification can be described as follows. Thematic structures are doubtlessly very important

to scientists and shape their work more than many other conditions of research. However, in spite of numerous attempts at topic delineation with an ever-growing number of methodologies of increasing sophistication, bibliometrics still cannot be sure about the relationship between the clusters of papers produced by its methods and the 'ground truths' about scientific topics. We know that we construct something relevant and important, and that active researchers can make sense of many structures we construct. However, we still have an insufficient understanding of the way in which our constructions relate to the thematic structures that shape research.

One of the reasons for this state is that the methodological developments we mentioned in our brief account have largely occurred in isolation from each other. As a rule, bibliometricians do not reproduce each other's results because they are either unable or not interested to do so (see Leydesdorff et al. 2015 for a recent case of irreplicability). As a consequence, there is no methodology of topic identification as a consistent body of knowledge that systematically relates types of data and their properties, research or application problems, and approaches to the identification of thematic structures in science. In this respect, topic identification appears to be art rather than science, with bibliometricians separately feeling their way through the exercise until results make sense.

3. A comparative approach to unresolved issues

So far, approaches to and results of topic identification exercises appear to be incommensurable. This is highly unsatisfying because there is no cumulative growth of knowledge, which means that the local progress made by many bibliometricians does not translate into progress of topic identification as a research area. Progress of a research area is possible only when findings can be related to each other and be placed in a larger consistent framework, which evolves with each contribution that is placed in it.

This kind of progress is difficult for various reasons, some of which are beyond the control of bibliometricians. The problems begin already with the nature of our empirical object. Although topics are doubtless 'out there' and influence researchers' actions, they are ambiguous because they look different from several equally valid scientific perspectives and because they are incompletely represented in any set of papers we might analyse. There are multiple partial ground truths, which cannot easily serve as a yardstick for topic identification.

Under these conditions, the activity we commonly call 'topic identification' is much more an exercise in topic *construction*. We make decisions on how to model the data based on the information contained in a set of publications, choose algorithms with specific properties, and set parameters for these algorithms. The decisions we make affect the outcome of topic identification procedures in ways that are not yet fully understood. We do not simply 'discover' the topics that 'are in the data' but actively construct them. This is not to say that our results are arbitrary and have nothing to do with the reality of topics in sets of publications as constructed and perceived by researchers.[6] Instead, being aware of the constructive nature of our approaches should help us to identify the relative influence of 'ground truths', data models, cluster algorithms, and other elements of our approaches.

The constructive nature of our approaches is also expressed by the definition of 'topic' we use. Three different kinds of definitions appear to coexist. *Operational definitions*, which define 'topic' as the outcome of a specific procedure, appear to be most common. We can also find *pragmatic* definitions, which define a topic as the thematic structure an external client wants to be investigated. *Theoretical* definitions, which define 'topic' in the context of the state of the art of science studies, have become

---

[6] The fact that researchers actively construct scientific topics, and do so differently depending on their own research, is one of the main causes for the difficulties involved in bibliometrically reconstructing topics.

exceedingly rare. Although we often proceed without explicit definition, the definition of our object we use implicitly affects our approach to identifying it.

Naturally, the definition of 'topic' influences the assessment of the solutions we produce. On what grounds do we consider a cluster solution to be 'satisfactory'? As we argued above, this assessment cannot be based on explicit arguments about the validity of approaches for two reasons. First, we have no way to determine whether our approaches truly reconstruct what we intend them to reconstruct. Second, we know that several different but equally valid solutions are likely to coexist. Our feeling that a solution is 'good' often includes an implicit notion of validity, which is based on experience. Apart from that, technical criteria such as stability or pragmatic criteria such as 'fits the purpose' appear to dominate. If, however, the suitability of an outcome for the purposes at hand plays an important role in the assessment of solutions, we need to make explicit purposes of topic identification and the assessment criteria derived from them.

To make things even more complicated, any clustering exercise of document spaces is a combination of bibliometric approaches and generic algorithms that have been developed in statistics, network analysis, or computer science. When engaging in topic identification, bibliometricians usually apply algorithms developed for non-bibliometric purposes to their bibliometric data models.

Thus, a further important decision that is sometimes made without full awareness of the consequences is the choice of clustering algorithm that is applied to the data. Algorithms are based on assumptions about their input, i.e. they 'presume' that the data they are applied to have certain properties. They also create an output with specific properties, properties that are not necessarily in accordance with the intentions of topic reconstruction. The best-known example for such properties is probably the 'chaining effect' of single-linkage clustering, i.e. the tendency of the algorithm to create 'long' clusters whose opposite ends may be thematically very different. Modularity-optimising algorithms "may fail to identify modules smaller than a scale which depends on the total size of the network and on the degree of interconnectedness of the modules, even in cases where modules are unambiguously defined" (Fortunato and Barthélemy 2007, p. 36). Other algorithms tend to produce clusters of uniform size or create a specific size distribution of clusters. If the real-world properties of document spaces we want to cluster deviate from those 'preferred' by algorithms, the algorithms create artefacts. In other words, algorithms have 'bad habits' that may severely distort the reconstruction of topics. We do not know enough about 'bad habits' of our algorithms. Systematic knowledge about these effects could contribute much to our understanding of the cluster solutions we produce.

How does the comparison presented in this special issue help us to address these problems? When we began the comparative cluster analysis, we were well aware of the limits. We do not have a yardstick for assessing cluster solutions, either. Our lack of expert knowledge of the subject matter – astronomy and astrophysics – prevents us from associating solutions or part of them with accepted scientific perspectives of the field. Any interpretation of cluster solutions is limited by our general knowledge of the field. Any comparison we conduct ourselves is limited to finding commonalities and differences between cluster solutions produced by different approaches.

But even with these limitations, we feel that the exercise in comparing approaches to topic reconstruction was useful. First, in the microcosm of interaction, in the lab-like sphere and in particular in the last phase of working together on one data set, we learned how to better present our own and inquire about other approaches. The learning process is documented in the contributions to the special issue, especially in the comparison paper. As a consequence, all contributions to this special issue are results of close collaborations.

Second, the comparison triggered the demand for more comparison. We began to 'exchange' data models. Velden applied the Infomap algorithm (as described by Velden et al. 2017a) to the 2010

network of papers and their cited sources, and provided the results for use by Havemann et al. (2017). Havemann et al. (2017) applied their algorithm to the 2003-2010 direct citation network that was used by Velden et al. (2017a) and van Eck and Waltman (2017). Wang et al. (2017) additionally used the Louvain algorithm with their data model. This algorithm was also used by Glänzel and Thijs (2017) and (in modified form) by van Eck and Waltman (2017) on their data. These comparisons gradually fill a matrix of data models and approaches (see Velden et al. 2017b). Each of the additional experiments improves the comparison of approaches by reducing the number of possible sources of variance.

Third, simply observing commonalities and differences between cluster analyses forced us to question our approaches. Why does one solution construct a cluster in a specific region of the network, while the other does not? Why do approaches seem to 'agree' on topics in some regions of the network but not in others? These questions made us look more closely at the solutions we produced and their relationship to the data model we used. Although we cannot argue with a 'ground truth', we can compare structural properties and contents of clusters. For the latter, we started from an idea Kevin Boyack has applied for a while and we created 'cluster description sheets' that list the distribution of documents over time, the most frequent keywords, journals and authors, and the most highly cited publications. The discussion about *methods to compare* cluster solutions led to two new papers on Thesaurus-based methods (Boyack 2017b) and on labelling (Koopman and Wang 2017). It also further informed the 'comparison paper' by Velden et al. (2017b).

Thus, although an ultimate validation based on a single undisputed ground truth remains impossible, we were able to address the above-described problems. Establishing comparability and reproducibility of clustering exercises is very difficult, and some of the obstacles we discussed are beyond the control of bibliometricians. Nevertheless, meeting these challenges seems crucial for bibliometrics to become a cumulative enterprise.

4. The data set: WoS *Astronomy and Astrophysics* papers 2003-2010

For our exercise, we selected the field of astronomy and astrophysics because its publications are very well covered by journal publication databases (Moed 2005, p. 130) and because it has a reputation as being relatively well separated from other fields. Since some groups projected the data set on their own databases, not all clustering exercises used *exactly* the same data set. Deviations are, however, small. They are described in the contributions on the clustering exercises.

We downloaded all articles, letters and proceedings papers from the Web of Science published 2003-2010 in journals listed in the category *Astronomy and Astrophysics* of the Journal Citation Reports of these years. We chose the period from 2003 to 2010 because several approaches work with direct citation networks and thus require a longer time window. The actual length of the time window was not a sensitive issue for any of the approaches.

Each of the groups had access to the Web of Science. The Berlin group downloaded the data through the web interface and provided a list of UT codes (the Accession Numbers, i.e. the unique publication identifiers used in the Web of Science databases), which were used by the others to match their downloads or to identify the publications in their own database. The latter procedure caused some deviations (Glänzel and Thijs 2017) and an interesting matching of WoS and Scopus datasets (Boyack 2017a).

Since the WoS download does not include the links between citing and cited records, an additional cleaning and matching of references was necessary. In particular, abbreviated journal titles varied significantly across reference lists and had to be standardised with rule-based scripts.

5. Introducing contributions to the special issue

The selection of approaches presented in this special issue results from the history of the meetings, and eventually the availability of groups to contribute to the special issue (Table 1). This is one reason why we see this special issue as the beginning of a specific methodological discourse which hopefully will spread among the bibliometric community.

Table 1: Table of content of the special issue "Same data – different results"

Introduction
Gläser et al.: Same data – different results? Towards a comparative approach to the identification of thematic structures in science
Known methods applied
Boyack (a): Investigating the Effect of Global Data on Topic Detection
Wang and Koopman: Clustering Articles Based on Semantic Similarity
Velden et al. (a): Mapping the Cognitive Structure of Astrophysics by Infomap: Clustering of the Citation Network and Topic Affinity Analysis
Van Eck and Waltman: Citation-Based Clustering of Publications Using CitNetExplorer and VOSviewer
New methods applied
Glänzel and Thijs: Using Hybrid Methods and `Core Documents' for the Representation of Clusters and Topics. The Astronomy Dataset
Havemann et al.: Memetic Search for Overlapping Topics Based on a Local Evaluation of Link Communities
Koopman et al.: Contextualization of Topics (Extended): Browsing through the Universe of Bibliographic Information
Synthesis
Boyack (b): Thesaurus-Based Methods for Mapping Contents of Publication Sets
Koopman and Wang: Mutual Information Based Labelling and Comparing Clusters
Velden et al. (b): Comparison of Topic Extraction Approaches and Their Results

The attempt to compare our approaches to topic identification presented us with three challenges. A first challenge emerges from the nature of our knowledge. While we know our own approaches well enough to successfully apply them, we also need to be aware of the 'screws and levers' we use in order to produce what we think is a good result. Some of this knowledge is tacit, which means that we

do not (cannot) usually communicate it. This affects the opportunities for others to understand, let alone reproduce, our solutions.

The second challenge concerns the comparison of approaches to topic identification. Owing to the solitary nature of methods development and application in bibliometrics, the need of comparison of approaches has been made explicit very rarely, if at all. All approaches presented in this special issue lead to groups of papers that are supposed to represent topics. However, these paper sets were produced by quite different means. What information must be shared to understand how an approach produced the reported results? How can we understand approaches to topic identification in a comparative perspective?

A third challenge concerns the comparison of outcomes of topic identification approaches. It turned out that such a comparison is a major methodological challenge in its own right. Given that there is no yardstick against which 'better' or 'worse' solutions can be identified: How does one compare solutions that cluster papers? What are important commonalities and differences between solutions? How can these differences be explained by linking them to properties of approaches?

The papers in this special issue address these challenges by describing approaches used for topic reconstruction, identifying decisions that have an impact on the outcome, and comparing solutions produced with different approaches. This special issue includes two types of papers presenting approaches to topic detection. The section 'Known methods applied' contains papers which apply existing methods to our dataset. Their approaches have already been published, and so the papers recapitulate the information that is necessary for integrating them in the comparative perspective.

Kevin Boyack (2017a) matches our Astro-dataset to his model of the full Scopus database. This makes it possible not only to discuss the quality of the subject delineation but also to compare local and global perspectives of the dataset. Boyack demonstrates convincingly that partitioning a subset of a database forces the creation of some topics that cannot be considered as such if their environment of documents from neighbouring disciplines is taken into account. His observations challenge the use of journal-based document sets to identify the structure of scientific fields.

Shenghui Wang and Rob Koopman (2017) apply and extend their method for building semantic matrices to the Astro data set (as detailed in *Koopman et al.* 2017) and use two algorithms, namely K-means and Louvain, to construct topics. They compensate for a major weakness of K-means – the necessity to determine the number of clusters *ex ante* – by building a 'ground truth' from four other clustering solutions presented in this volume and determining the number of clusters of which the K-means solution best fits this 'ground truth'. The comparison of their solutions to each other and to the other solutions provided in this volume suggest that clustering solutions depend more strongly on the data models than on the algorithms used.

Theresa Velden et al. (2017a) construct the direct citation network of the Astro data-set, to which they apply the Infomap algorithm. Since this algorithm produces nearly 2000 clusters, it is applied a second time to these clusters and thus leads to a solution consisting of 21 clusters. Due to its stochastic nature, the algorithm produces different solutions in each run, which, however, are largely consistent. In their discussion of the two consecutive runs of their algorithm, Velden et al. point out the impact of this decision on the size of the topics finally obtained, and raise the interesting question how the 'right' level of an algorithm's resolution can be determined.

Nees Jan van Eck and Ludo Waltman (2017) use CitNetExplorer, a recently developed tool that is also available online, for clustering the direct citation network and exploring the solution. The clustering is produced by their smart local moving algorithm, which is an improvement of the Louvain algorithm. They produce clustering solutions at four different resolution levels with more than 400 clusters at the

highest resolution level and 22 clusters at the lowest level. The solution with 22 clusters is used in our comparative exercise and is analysed in detail in the paper using CitNetExplorer and VOSviewer. Both tools can be used for an exploration of the clustering solution at the aggregate level and for "drilling down" to individual publications and their citation relations respectively term maps of individual clusters.

The next section 'New methods applied' presents new approaches that have not yet been fully described elsewhere. The first of these is the paper by *Wolfgang Glänzel and Bart Thijs* (2017), who introduce a new method of hybrid clustering that combines bibliographic coupling and lexical coupling. Their data model enables to vary the relative influence of bibliographic and lexical coupling on the clustering. Furthermore, they use two different lexical approaches (single term and noun phrases). From the many possible combinations of these parameters, they compare pure bibliographic coupling to one hybrid of bibliographic and single-term lexical coupling and one hybrid combining bibliographic and noun-phrase lexical coupling. The Louvain algorithm is applied to each of these data models with two resolution levels. Core documents are used to further investigate the cluster structure. The comparison of the six solutions shows that they are consistent, and that the solutions at different resolution levels form a hierarchy in each case.

Frank Havemann et al. (2017) argue that from a theoretical point of view, topics are overlapping thematic structures, and that reconstructing this particular feature of topics requires going beyond hard clustering. They cluster links instead of nodes and develop a memetic algorithm that combines random and deterministic searches. Since the algorithm grows a large number of communities independently from each other rather than partitioning the network, it explicitly burdens the analyst with decisions that sometimes remain hidden in other algorithms. These include explicit decisions on the minimum size of a community that justifies considering it as representing a topic and on the minimum dissimilarity of communities to be considered as representing different topics.

Rob Koopman et al. (2017) apply ideas from information retrieval for building a system that supports the exploration of a semantic space, in this case a semantic space constructed from the Astro data set (LittleAriadne). Their approach differs from lexical approaches in bibliometrics in its inclusion of all components of an article. The authors demonstrate how this system can be used to explore the content of articles, clusters and whole clustering solutions. They also use LittleAriadne for comparing cluster solutions of the Astro data set provided by the other contributors to this volume, thereby supporting the understanding of commonalities of and differences between approaches.

The paper by Koopman et al. already introduces the topic of comparing cluster solutions. The last section of the issue contains under the heading 'Synthesis' three papers, namely two shorter papers that propose new methods for comparing cluster solutions and a joint paper of a larger group of authors of this issue that compares the cluster solutions produced by the group. *Kevin Boyack* (2017b) uses a domain specific Knowledge Organization System, the Unified Astronomy Thesaurus, to index documents in topical clusters, to develop a backbone map how the Thesaurus unfolds in our dataset, and eventually uses this Thesaurus map as a reference to place and thus compare the different cluster solutions. *Rob Koopman and Shenghui Wang* (2017) apply Mutual Information measures to compare distributions of terms across clusters, and cluster solutions.

The final 'comparison paper' by *Velden et al.* (2017b) starts from an overview of data models and clustering algorithms used by the participants, and then applies a variety of tools (including the new ones introduced by the two preceding papers) to compare the solution from different perspectives. In its comparison of all solutions, pairwise comparisons of solutions and in-depth comparisons of particular clusters created by some of the solutions, the paper advances strategies and methods for comparison. The central challenge remains comparing the constructed clusters as representations of

thematic structures without using an understanding of the research content, which bibliometricians usually do not have for domains they apply their methods to. Responding to the challenge, the paper and the accompanying explanations of the methodologies demonstrate how we can achieve at least some understanding of the reasons why different approaches construct different topics.

6. Looking forward

The articles in this special issue detail what we have learned about our approaches, the solutions they provide, and critical decisions in a topic reconstruction exercise. Where do we go from here? We would like to emphasise three lessons we consider particularly important and to invite the bibliometrics community to join us in moving forward the comparative approach to topic reconstruction.

The first lesson is that our comparative exercise once more highlighted the importance of notions of 'topic' underlying the approaches to topic identification. Our conviction that we can identify topics is based on a tacit understanding of what a topic is. We need to discuss this more explicitly. Several contributions to this special issue identified problems that might require a more explicit definition of the concept, be it because the minimum size of a structure that can represent a topic needs to be determined (Velden et al. 2017a; van Eck and Waltman 2017; Havemann et al. 2017) or because topics look differently if the environment of the paper set is included (Boyack 2017a). Furthermore, all approaches reported in this special issue require decisions that have consequences for the number of topics constructed. While only the K-means algorithm required the exact number of topics to be determined ex ante, the other algorithms required indirect decisions on numbers of topics. In other cases, the number of clusters followed from the combination of a data model with resolution levels of the algorithm (Louvain and the smart moving local algorithm), the decision to run the algorithm twice in order to obtain a manageable number of clusters (Infomap), or the selection of a cut-off point at which the network is sufficiently covered and structured by topics (memetic algorithm). Do we assume that different decisions would have produced different but equally valid solutions, i.e. that any partitioning of a set of papers into topics returns a valid set of topics? Can we make explicit our understanding of topic on which the answer to this question is based?

Our understanding of topic is closely linked to the purpose of a topic reconstruction exercise. This points to a limitation of our comparison exercise and thus to a task for future comparative research. The assessment of validity needs to be complemented by an assessment of appropriateness. For example, exploring the thematic structure of the disciplines astronomy and astrophysics for sociological purposes might lead to the construction of hundreds of overlapping or hierarchically organized topics, while mapping the field for science policy or management clients would require a much simpler structure. Detecting emerging topics would be a different matter again because few small special topics would need to be identified. In this special issue, we focused on technical aspects of topic reconstruction, thereby disregarding for the moment that different approaches fit different purposes. The purposes for which the authors of this special issue commonly use 'their' approaches are only implicitly present in authors' methodological choices. *Systematically* linking approaches and decisions to purposes of topic reconstruction would be an important step forward. This step must include the reconstruction of topics from outside perspectives (e.g. applied or political ones), as was illustrated by Leydesdorff et al. (2012), who used medical subject headings to construct a perspective on the dynamics of medical innovations. Again, the major task ahead of us is a systematic comparison of links between purposes, approaches, and their outcomes in terms of topic reconstruction. Our current contribution to this future work is suggesting a framework and language for a comparative

approach. We also want to avoid normative myopia that recommends the 'best' approach or at least a best-for-that-purpose approach.

As part of this framework, we need to open up the question of algorithms. While we are in control of our data models, we tend to leave algorithms as they are provided by computer scientists or physicists. To be sure, we 'shop around' and identify weaknesses of algorithms that make them unsuitable for the task, which we try to overcome (see Wang and Koopman 2017 on K-means, van Eck and Waltman 2017 on transforming Louvain, Velden et al. 2017a on running Infomap repeatedly) or use as decision criteria (see Glänzel and Thijs 2017 on abandoning Ward in favour of Louvain). But the larger problem is highlighted rather than solved by these workarounds. It also goes beyond the problem of 'bad habits' of algorithms mentioned earlier in this article. The algorithms provided by other disciplines are not neutral but have built-in assumptions about network properties and 'ground truths' expressed by these properties. We need to explore these assumptions and find algorithms whose in-built assumptions match the properties of bibliometric networks and topics.

The joint comparison of approaches to topic identification that is documented in this special issue is only a first step. We will continue to explore our own solutions in order to learn more. In particular, we want to talk to experts from astronomy and astrophysics. This poses an interesting new challenge in itself because the interaction between bibliometrics and field experts now has a new purpose. Instead of 'validating' the outcome of one bibliometric exercise (or finding the 'most true' one in our case), the collaboration with experts would now be aimed at identifying perspectives from which certain (aspects of) cluster solutions make sense in astronomy and astrophysics.

To make the most of our comparative exercise, we would like to include as many colleagues as possible. The permission by Clarivate Analytics to make the dataset of our exercise available with a user-friendly license enables a topic identification challenge (Boyack et al. 2017): Dear colleagues, we invite you to apply your approach to topic identification to our data, and to use the comparison to our results for achieving a deeper understanding of your own approach. We offer our own solutions for comparison and some tools to support comparison at www.topic-challenge.info.

As a conclusion to this article, we would also like to invite the bibliometrics community to consider further bibliometric problems to which a comparative approach like ours should be applied. Topic identification is an important task in bibliometrics but by no means the only one which could benefit from comparative approaches. Too much bibliometric research suffers from scholarly isolation. While attempts to create standards for bibliometrics (Glänzel and Schoepflin 1994; Glänzel 1996; Glänzel et al. 1996; Sirtes et al. 2013) advance only slowly and do not go unchallenged, we should at least agree on the importance of being able to explain and compare what we did. This is a pre-condition to the also pending challenge to produce re-producible results, and to develop datasets for benchmarking. Price ascribed cumulative growth to the 'hard' sciences and hoped that bibliometrics would move towards this growth pattern (Solla Price 1986 [1963]). Developing comparative approaches to our own research will support this by increasing consistency and coherence of our research.

Acknowledgements

References

Bauin, S., Michelet, B., Schweighoffer, M. G., & Vermeulin, P. (1991). Using bibliometrics in strategic analysis: "understanding chemical reactions" at the CNRS. *Scientometrics, 22*(1), 113−137.

Blei, D. M. (2012). Probabilistic topic models. *Commun. ACM, 55*(4), 77−84.

Boyack, K. W. (2017a). Investigating the Effect of Global Data on Topic Detection. In Gläser, J., Scharnhorst, A. & Glänzel, W. (eds), Same data – different results? Towards a comparative approach to the identification of thematic structures in science, Special Issue of *Scientometrics*, volume (issue), pages.

Boyack, K. W. (2017b). Thesaurus-based methods for mapping contents of publication sets. In Gläser, J., Scharnhorst, A. & Glänzel, W. (eds), Same data – different results? Towards a comparative approach to the identification of thematic structures in science, Special Issue of *Scientometrics*, volume (issue), pages.

Boyack, K. W. Glänzel, W., Gläser, J., Havemann, F., Thijs, B., Van Eck, N. J., Velden, T. & Waltmann, L. (2017) Topic Identification Challenge. In Gläser, J., Scharnhorst, A. & Glänzel, W. (eds), Same data – different results? Towards a comparative approach to the identification of thematic structures in science, Special Issue of *Scientometrics*, volume (issue), pages.

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.

Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics, 64*(3), 351−374.

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*(2), 191−235.

Cambrosio, A., & Keating, P. (1988). "Going Monoclonal": Art, Science, and Magic in the Day-to-Day use of Hybridoma Technology. *Social Problems, 35*(3), 244−260.

Chubin, D. E. (1976). The Conceptualization of Scientific Specialties. *Sociological Quarterly, 17*(4), 448−476.

Collins, H. M. (1975). The Seven Sexes: A Study in the Sociology of a Phenomenon, or the Replication of Experiments in Physics. *Sociology, 9*, 205−224.

Collins, H. M. (1974). The TEA set: tacit knowledge and scientific networks. *Science Studies 4*, 165 −186.

Crane, D. (1972). *Invisible Colleges: Diffusion of Knowledge in Scientific communities*. Chicago: The University of Chicago Press.

Crawford, S. (1971). Informal Communication Among Scientists in Sleep Research. *Journal of the American Society for Information Science*, 301−310.

Edge, D., & Mulkay, M. J. (1976). *Astronomy Transformed: The Emergence of Radio Astronomy in Britain*. New York: John Wiley & Sons, Inc.

Fortunato, S., & Barthélemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences, 104*(1), 36−41.

Glänzel, W. (1996). The need for standards in bibliometric research and technology. *Scientometrics, 35*(2), 167−176.

Glänzel, W., & Czerwon, H. J. (1996). A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level. *Scientometrics, 37*(2), 195−221.

Glänzel, W., Katz, S., Moed, H., & Schoepflin, U. (1996). Preface. *Scientometrics, 35*(2), 165−166.

Glänzel, W., & Schoepflin, U. (1994). Little Scientometrics, Big Scientometrics ... And Beyond. *Scientometrics, 30*(2-3), 375−384.

Glänzel, W. & Thijs, B. (2017). Using hybrid methods and `core documents' for the representation of clusters and topics. The astronomy dataset. In Gläser, J., Scharnhorst, A. & Glänzel, W. (eds), Same data − different results? Towards a comparative approach to the identification of thematic structures in science, Special Issue of *Scientometrics*, volume (issue), pages.

Gläser, J., Heinz, M., & Havemann, F. (2015). Epistemic Diversity as Distribution of Paper Dissimilarities. In A. A. Salah, Y. Tonta, A. A. A. Salah, C. Sugimoto & U. Al (Eds.), *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference, Istanbul, Turkey, 29 June to 3 July, 2015* (pp. 1006−1017). Istanbul: Boğaziçi University Printhouse.

Glenisson, P., Glänzel, W., Janssens, F., & De Moor, B. (2005). Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing & Management, 41*(6), 1548−1572.

Griffith, B. C., Small, H. G., Stonehill, J. A., & Dey, S. (1974). The Structure of Scientific Literatures II: Toward a Macro- and Microstructure for Science. *Science Studies, 4*(4), 339−365.

Havemann, F., Gläser, J. & Heinz, M. (2017). Memetic search for overlapping topics based on a local evaluation of link communities. In Gläser, J., Scharnhorst, A. & Glänzel, W. (eds), Same data − different results? Towards a comparative approach to the identification of thematic structures in science, Special Issue of *Scientometrics*, volume (issue), pages.

Healey, P., Rothman, H., & Hoch, P. K. (1986). An experiment in science mapping for research planning. *Research Policy, 15*(5), 233−251.

Hric, D., Darst, R. K., & Fortunato, S. (2014). Community detection in networks: Structural communities versus ground truth. *Physical Review E, 90*(062805), 1−19.

Janssens, F., Glänzel, W., & De Moor, B. (2007). Dynamic hybrid clustering of bioinformatics by incorporating text mining and citation analysis. *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. San Jose, California, USA, ACM, 360−369.

Jarneving, B. (2001). The cognitive structure of current cardiovascular research. *Scientometrics, 50*(3), 365−389.

Katz, J. S. (1999). The self-similar science system. *Research Policy, 28*(5), 501−517.

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.

Klavans, R., & Boyack, K. W. (2015). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? http://arxiv.org/abs/1511.05078 (accessed 21 December 2016).

Koopman, R. & Wang, S. (2017). Mutual Information Based Labelling and Comparing Clusters. In Gläser, J., Scharnhorst, A. & Glänzel, W. (eds), Same data − different results? Towards a comparative approach to the identification of thematic structures in science, Special Issue of *Scientometrics*, volume (issue), pages.

Koopman, R., Wang, S. & Scharnhorst, A. (2017). Contextualization of Topics (Extended): Browsing through the Universe of Bibliographic Information. In Gläser, J., Scharnhorst, A. & Glänzel, W. (eds), Same data − different results? Towards a comparative approach to the identification of thematic structures in science, Special Issue of *Scientometrics*, volume (issue), pages.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.

Law, J., Bauin, S., Courtial, J.-P., & Whittaker, J. (1988). Policy and the Mapping of Scientific Change: A Co-word Analysis of Research into Environmental Acidification. *Scientometrics, 14*(3-4), 251−264.

Leydesdorff, L. (1986). The Development of Frames of References. *Scientometrics, 9*(3-4), 103−125.

Leydesdorff, L. (1987). Various Methods for the Mapping of Science. *Scientometrics, 11*(5-6), 295−324.

Leydesdorff, L. (2004). Clusters and maps of science journals based on bi-connected graphs in Journal Citation Reports. *Journal of Documentation, 60*(4), 371−427.

Leydesdorff, L., & Bornmann, L. (2011). How fractional counting of citations affects the impact factor: Normalization in terms of differences in citation potentials among fields of science. *Journal of the American Society for Information Science and Technology, 62*(2), 217−229.

Leydesdorff, L., & Cozzens, S. E. (1993). The Delineation of Specialities in Terms of Journals Using the Dynamic Journal Set of the SCI. *Scientometrics, 26*(1), 135−156.

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology, 60*(2), 348−362.

Leydesdorff, L., Rotolo, D., & Rafols, I. (2012). Bibliometric perspectives on medical innovation using the medical subject Headings of PubMed. *Journal of the American Society for Information Science and Technology, 63*(11), 2239−2253.

Leydesdorff, L., Wagner, C., & Bornmann, L. (2015). Replicability and the public/private divide. *Journal of the American Society for Information Science,* forthcoming.

Leydesdorff, L., Wouters, P., & Bornmann, L. (2016). Professional and citizen bibliometrics: complementarities

and ambivalences in the development and use of indicators—a state-of-the-art report. *Scientometrics*, 1−22.

Marshakova, I. V. (1973). A system of document connection based on references (in Russian). *Scientific and Technical Information Serial of VINITI,* 6(2), 3−8.

Moed, H. F. (Ed.) (2005). *Citation Analysis in Research Evaluation*. Dordrecht: Springer.

Moed, H. F., Burger, J. M., Frankfort, J. G., & Raan, A. F. J. van (1985). The application of bibliometric indicators: important field- and time-dependent factors to be considered. *Scientometrics,* 8(3-4), 177−203.

Mulkay, M. J. (1977). Sociology of the Scientific Research Community. In I. Spiegel-Rösing & D. de Solla Price (Eds.), *Science, Technology and Society: A Cross-Disciplinary Perspective* (pp. 93−148). London: Sage.

Mulkay, M. J., Gilbert, G. N., & Woolgar, S. (1975). Problem Areas and Research Networks in Science. *Sociology,* 9(2), 187−203.

Noyons, E. C. M. (2004). Science maps within a science policy context. In H. F. Moed, W. Glänzel & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* (pp. 237−256). Dordrecht: Kluwer Academic Publishers.

Noyons, E. C. M., & Raan, A. F. J. van (1998). Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research. *Journal of the American Society for Information Science,* 49(1), 68−81.

Noyons, E. C. M. (2001). Bibliometric mapping of science in a science policy context. *Scientometrics,* 50(1), 83−98.

Peel, L., Larremore, D. B., & Clauset, A. (2016). *The ground truth about metadata and community detection in networks*. Cornell University Library. arXiv:1608.05878. https://arxiv.org/abs/1608.05878 (accessed 21 December 2016).

Raan, A. F. J. van (1990). Fractal dimension of co-citations. *Nature,* 347, 626.

Raan, A. F. J. van (2000). On growth, ageing, and fractal differentiation of science. *Scientometrics,* 47(2), 347−362.

Raan, A. F. J. van, & Tijssen, R. J. W. (1993). The Neural Net of Neural Network Research: An Exercise in Bibliometric Mapping. *Scientometrics,* 26(1), 169−192.

Rip, A., & Courtial, J.-P. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics,* 6(6), 381−400.

Schiminovich, S. (1971). Automatic classification and retrieval of documents by means of a bibliographic pattern discovery algorithm. *Information Storage and Retrieval,* 6(6), 417−435.

Sirtes, D., Waltman, L., Archambault, É., Glänzel, W., Hornbostel, S., & Wouters, P. (2013). Bibliometric Evaluation Standards Debate: Introductory Presentation and Panel Discussion. In S. Hinze & A. Lottmann (Eds.), *Translational twists and turns: Science as a socio-economic endeavor* (pp. 373−377). Berlin: Institute for Research Information and Quality Assurance.

Small, H. G. (1973). Co-citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents. *Journal of the American Society for Information Science,* 24(4), 265−269.

Small, H. G. (1977). A Co-Citation Model of a Scientific Specialty: A Longitudinal Study of Collagen Research. *Social Studies of Science,* 7(2), 139−166.

Small, H. G., Boyack, K. W., & Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy,* 43(8), 1450−1467.

Small, H. G., & Griffith, B. C. (1974). The Structure of Scientific Literatures I: Identifying and Graphing Specialities. *Science Studies,* 4(1), 17−40.

Smiraglia, R. P. (2015). Domain Analysis of Domain Analysis for Knowledge Organization: Observations on an Emergent Methodological Cluster. *Knowledge Organization,* 42(8), 602−611.

Solla Price, D. de (1986 [1963]). *Little science, big science ... and beyond*. New York: Columbia University Press.

Šubelj, L., Eck, N. J. van, & Waltman, L. (2016). Clustering Scientific Publications Based on Citation Relations: A Systematic Comparison of Different Methods. *PLoS ONE,* 11(4), e0154404.

Tijssen, R. J. W. (1992). A Quantitative Assessment of Interdisciplinary Structures in Science and Technology: Co-Classification Analysis of Energy Research. *Research Policy,* 21(1), 27−44.

Van Eck, N.J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. In Gläser, J., Scharnhorst, A. & Glänzel, W. (eds), Same data − different results? Towards a comparative approach to the identification of thematic structures in science, Special Issue of *Scientometrics*, volume (issue), pages.

Velden, T., Yan, S. & Lagoze, C. (2017a). Mapping the Cognitive Structure of Astrophysics by Infomap: Clustering of the Citation Network and Topic Affinity Analysis. In Gläser, J., Scharnhorst, A. & Glänzel, W. (eds), Same data − different results? Towards a comparative approach to the identification of thematic structures in science, Special Issue of *Scientometrics*, volume (issue), pages.

Velden, T., Boyack, K. W., Gläser, J. Koopman, R., Scharnhorst, A. & Wang, S. (2017). Comparison of Topic Extraction Approaches and Their Results. In Gläser, J., Scharnhorst, A. & Glänzel, W. (eds), Same data – different results? Towards a comparative approach to the identification of thematic structures in science, Special Issue of *Scientometrics*, volume (issue), pages.

Verspagen, B., & Werker, C. (2003). The Invisible College of The Economics of Innovation and Technological Change. *Estudios de Economía Aplicada,* 21(3), 393-419.

Waltman, L., & Eck, N. J. van (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology,* 63(12), 2378−2392.

Wang, S. & Koopman, R. (2017) Clustering Articles Based on Semantic Similarity. In Gläser, J., Scharnhorst, A. & Glänzel, W. (eds), Same data – different results? Towards a comparative approach to the identification of thematic structures in science, Special Issue of Scientometrics, volume (issue), pages.

Wolfram, D. (2015). The symbiotic relationship between information retrieval and informetrics. *Scientometrics*, 102(3), 2201−2214. doi:10.1007/s11192-014-1479-0

White, H. D., & Griffith, B. C. (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. (1984). The development of management studies as a fragmented adhocracy. *Social Science Information,* 23(4/5), 775−818.

Woolgar, S. (1976). The Identification and Definition of Scientific Collectivities. In G. Lemaine, R. Macleod, M. Mulkay & P. Weingart (Eds.), *Perspectives on the Emergence of Scientific Disciplines* (pp. 233−245). Paris: Mouton.

Yau, C.-K., Porter, A., Newman, N., & Suominen, A. (2014). Clustering scientific documents with topic modeling. *Scientometrics,* 100(3), 767−786.

Zitt, M., & Bassecoulard, E. (2006). Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information Processing & Management,* 42(6), 1513−1531.