The application of bibliometric analysis: disciplinary and user aspects

Ellegaard, Ole

Published in:
Scientometrics

DOI:
10.1007/S11192-018-2765-Z

Publication date:
2018

Document version
Accepted manuscript

Document license
Unspecified

Citation for published version (APA):

Terms of use
This work is brought to you by the University of Southern Denmark through the SDU Research Portal.
Unless otherwise specified it has been shared according to the terms for self-archiving.
If no other license is stated, these terms apply:

• You may download this work for personal use only.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.
Please direct all enquiries to puresupport@bib.sdu.dk
The application of bibliometric analysis: Disciplinary and user aspects.

Ole Ellegaard

Abstract  Bibliometric analysis (BA) has been used increasingly as a tool within the scientific community. Interplay is vital between those involved in refining bibliometric methods and the recipients of this type of analysis. Production as well as citations patterns reflect working methodologies in different disciplines within the specialized Library and Information Science (LIS) field, as well as in the non-specialist (non-LIS) professional field. We extract the literature on bibliometric analyses from Web of Science (WoS) in all fields of science and analyze clustering of co-occurring keywords at an aggregate level. It reveals areas of interconnected literature with different impact on the LIS and the non-LIS community. We classify and categorize bibliometric articles that obtain the most citations in accordance with a modified version of Derrick’s, Jonker’s and Lewison’s method (Derrick et al. 2012). The data demonstrates that cross-referencing between the LIS and the non-LIS field is modest in publications outside their main categories of interest, i.e. discussions of various bibliometric issues or strict analyses of various topics. We identify some fields as less well-covered bibliometrically

Keywords  Bibliometric analysis, Policy implications, Publication analysis, Citation analysis.
Background

The use of bibliometric methods in the scientific and professional community goes far beyond the original idea of simple listings of scientific production or citation indexing and shows great variety throughout the professional disciplines. In the academic world, rankings and other productivity measurements are applied routinely based on bibliometric indicators, with economic and political consequences (Harvey 2008; Abbott et al. 2010). Initially, the founder of the Institute for Scientific Information, E. Garfield, described the future prospects of bibliometrics. Citation indexing was particularly discussed in a number of editorials in Current Contents, see e.g. Garfield (1977). The field grew rapidly in the decades that followed the release of electronic versions of the major bibliographic databases (Wilsdon et al. 2015).

The concept ‘systematic review’ relates closely to performing bibliometric analyses. Many scientists consider it as a necessary part and initiation of the scientific process (Tranfield et al. 2003). A systematic review serves a purpose as a complete documentation of the literature in an often narrowly focused field. Consequently, Health Science publish a number of papers that discusses and evaluate this kind of documentation.

In contrast, bibliometrics is an analytical tool used in many widely different contexts. It ranges from traditional measurements of citation impact (Kaur et al. 2013) to identification of problematic substances in the environment (Grandjean et al. 2011).

Using statistical analyses of the increasing amount of data available (Bornmann et al. 2014) makes reliable mapping of science development, cooperation, and rankings possible. A recent literature review by Wouters et al. (2015) listed articles that discussed numerous aspect of bibliometric science.

Hirsch (2005) invented the H-index as a simple way to measure scientific performance. It spurred an interest in research evaluation that went beyond the usual stakeholders in the field. Obviously, a simple parameter that combined productivity and impact into one number appealed to a number of people. The limitations of such a metric gradually became obvious and the discussions that followed in the scientific community were highly beneficial, because they led to a more profound interest in the field of scientometrics (Bormann and Daniel 2005; Glänzel 1996). A number of alternative metrics were introduced that either elaborate on the h-index or are meant to compensate for its deficiencies (Bornmann et al. 2008). Policymakers and science evaluators quickly took advantage of the new tools. Scattered data and input on the outcome of research which had hitherto been qualitative could be quantified (Bollen and Van de Sompel 2008; Weller 2015). Wallin (2005) described in detail the pros and cons of using bibliometric methods for research evaluation. The necessity of a discussion about the application of bibliometric methods and establishing standards was obvious in the research community already at an early stage (Glänzel 1996; Weingart 2005; Van Noorden 2010; Bonell 2016).

Hicks et al. (2015) took all these aspects and their implementation to a more formal level in the Leiden Manifesto. They set out to establish principles that could improve the application of various metrics and indicators in research evaluation. They also encouraged an increasing level of cooperation between bibliometricians in the LIS-community and the recipients of reports (the non-LIS community), and
advocated for more transparency in the process. Fulfilment of the requirements in the Leiden Manifesto could remedy some of the issues raised by Glänzel (1996) and others about the validity of applying bibliometric methods in research evaluation. Despite the reservations mentioned above, bibliometric methods are widely used in this context. Theoretical aspects of these methods as well as implementation of practical details are discussed in the LIS community and, although to a lesser extent, among professional stakeholders outside LIS (Larivière 2012). There is evidence that the two groups define the role and practice of bibliometrics differently (Cox et al. 2017). According to Petersohn (2016), no systematic empirical evidence exists about users of bibliometrics in different fields and their relation and interaction with the academic fields of Scientometrics. If reflected in publishing practices, i.e. users from the different fields preferentially use their ‘own’ journals for studies of bibliometric nature, we can apply publication patterns as a proxy for empirical studies of co-operation between the LIS and non-LIS community.

In the same manner, any overlap in publication patterns or co-authoring of articles between the two groups reflects the mutual interest in making knowledge about bibliometric methods and practices available to the opposite community. Petersohn (2016) demonstrated the need for knowledge about standardization of bibliometric indicators by the professional practitioners at numerous occasions.

In the case of the LIS-community, it is necessary to be aware of the use of bibliometric methods in the many reports now published by policy makers and research analysts. It is important to inform the recipients of reports about any limitations in the methodology, especially the preconditions. In the end, the integrity and validity of the data material should be the basis of the conclusions that are drawn. A recent example is the, at times heated, discussion about ranking of educational institutions and its implications for funding; see e.g. Van Raan (2005) or Hicks (2012).

In a recent paper, Leydesdorff et al. (2016a) mentioned four main groups of actors: 1. Producers of bibliometric products, 2. Bibliometricians, 3.Managers, and 4. Scientists. We will only take into account the last three players in the field but mention that group 1, Producers, now encompasses local repositories, i.e. University Research Information Systems (CRIS) with facilities for performance assessments and networking. Many scientists as well as managers are directly involved in using these systems, and it is important that they understand the outcome of research reports within the limits of applied bibliometric methodology. Without the involvement of bibliometricians at any stage in the process, research managers can download such research reports from these repositories. In this way, the presentation as well as discussions of bibliometric methods in the literature becomes increasingly relevant to the scientific community.

Derrick et al. (2012) analyzed thoroughly the characteristics of bibliometric articles both outside and inside the LIS field. They found, based on the number of co-publications, an increasing amount of cooperation between LIS and the professional (non-LIS) community. On the other hand, bibliometric researchers cite bibliometric papers published in non-LIS journals relatively less than similar papers in LIS journals. The authors’ main hypothesis is that the purpose of publishing bibliometric articles is different within the two communities. A publication in a non-LIS journal will most likely target a specific audience
with a professional interest in the outcome. Further, they concluded that in general, readers of non-LIS bibliometric articles are less likely to cite them, because they are outside the scope of their professional interest. LIS journals publish mostly theoretical bibliometric work that includes discussions about methods and their interpretations. These types of work are probably less known outside the LIS community (Jonkers and Derrick 2012). In contrast, Prebor (2010) found that only one third of the research tagged on the ProQuest Digital Dissertations database under ‘Library Science’ or ‘Information Science’ was actually conducted at LIS departments. Her study could indicate a gradual move towards more interdisciplinary work in the LIS field. Ellegaard and Wallin (2015) elaborated further on these points. They followed citation patterns over a long period and calculated, in analogy with the Journal Impact Factor (JIF), a normalized impact. They defined it as the yearly average number of articles citing a corpus of articles on bibliometric analysis in the fields of natural science and medicine published before a given year. A normalized impact of the non-LIS part of the articles showed an almost linear increase for non-LIS citations as compared to LIS citations. The similar numbers of normalized LIS citations show a more constant level. The sluggish growth suggests that bibliometrics is regarded as a mature field by LIS.

If we consider the use of bibliometrics in a broader perspective, it is likely that countries with long traditions of doing bibliometric science (hereafter named ‘Western countries’) have different publication patterns than similar countries without this tradition (hereafter named ‘up-and-coming countries’). Both the methodology used and the fields that are studied could be influenced. A sensible question could be if the bibliometric work reflect the prevalent scientific studies in the relevant countries. The division of countries in two main groups makes it possible to find out whether newcomers to the field of bibliometrics tend to investigate certain mainstream subjects. It could also happen that they turn to more pressing issues representing concrete local problems. Furthermore, it is likely that scientists from ‘up-and-coming’ countries are less aware of the ‘nuts and bolts’ of bibliometric analysis. It may increase the risk of misinterpretations and skewed results. On the other hand, bibliometricians from ‘western countries’ may not be fully aware of the methodological problems faced by members of the other group. Co-operation of any kind between different groups: LIS or non-LIS, experienced or non-experienced countries, would be highly beneficial to the field of bibliometrics.

The extent to which the two categories of countries could benefit from each other may show up in an analysis of citation patterns. One may perform this analysis in variety of ways: 1. Affiliation of authors with the possibility of contribution from both LIS and non-LIS teams/persons in the same paper (Jonkers and Derrick 2012). 2. Multidisciplinary article content (Jonkers and Derrick 2012; Ellegaard and Wallin 2015). 3. Acknowledgement via citation or other types of metrics to publications between different fields. It is possible that some high-profile scientific fields with a large literature basis are undernourished from a bibliometric point of view. This leads up to the more profound question about usability and to what end the work is initiated. The characteristics of bibliometric work, i.e. analysis of fields, topics, people or organizations, may be preferred by particular groups and at various times. Discussions of bibliometric
theory, improvements of methods, or policy implications, may be more prominent in other groups and periods. It is probable that the subjects under investigation play a prominent role. In fact to shed light on e.g. political or environmental issues bibliometric analysis are often published (Van Raan 2005; Liu et al. 2011). The LIS community publishes more theoretical, analytic papers that discuss issues of importance to the advancement of the bibliometric field. The non-LIS community places more emphasis on factual analysis or statistics that are strictly relevant to the fields investigated. Examples could be literature such as ‘A bibliometric analysis of global Zika research’ (Martinez-Pulgarin et al. 2015) or ‘Publish and perish? Bibliometric analysis, journal ranking and the assessment of research quality in tourism’ (Hall 2011).

Could this division of subjects spill over to the publications from countries that have published numerous fundamental or applied papers on bibliometric issues in contrast to those with less and only newer publications in the field?

In the present paper, we plan to corroborate the findings of Jonkers and Derrick (2012) and earlier results by Ellegaard and Wallin (2015) on these co-operation patterns between the LIS and the non-LIS community. The subject fields studied as well as a carefully conducted keyword analysis in the bibliometric literature may shed further light on this issue. Citation patterns may also contribute with knowledge about visibility between the two groups and the way they use and acknowledge the work of each other.

The growing number of publications seems to fulfil the demand for reports and analytical statistics in an increasing number of scientific communities. Still, it is important that the LIS community keeps its focus not only on elaborating on bibliometric theory, indicators, and methods, but also carefully considers the use of bibliometrics in a practical context (McKechnie and Pettigrew 2002; Braun et al. 2010). The focus should especially be on how, when, and why a given method or indicator is applied. It may lead to obvious questions about these procedures. We will consider in more detail what characterises the content of the publications as well as the method of analysis, and look specifically at the impact of each category.

Investigation of the engagement between the two communities, LIS and non-LIS, is vital in order to uncover areas with a potential for further cooperation.

To this end, we apply studies of citation patterns in combination with keyword analysis and examining the characteristics of the bibliometric literature as our preferred analytical methods.

**Method.**

We extract a corpus of articles that apply or discuss bibliometric analytical methods from the literature and use them for further analysis. It includes works published within the natural, medical and social sciences as well as humanities. The publication patterns are somewhat different within social science and the humanities but here we choose to include these subject areas in order to get a more complete picture. Articles published throughout 1964-2016 are included in the present work and represent the completely active period of publication in this field.

According to Larivière et al. (2006) and Leydesdorff (2016a) one may use databases such as WoS by Clarivate Analytics or Scopus by Elsevier legitimately for bibliometric purposes in natural and life science.
albeit to a lesser extent in social science. The method that uses WoS subject categories has become an established practice in evaluative bibliometrics (Leydesdorff and Bornmann 2016b). Publication patterns govern this categorization. Further, keyword analysis provides patterns more detailed with regard to subjects and this includes the possibility of network analysis.

Google Scholar has a better coverage in all fields but their sources of statistics and materials are largely unknown (Leydesdorff 2016a). Only WoS apply the topic area of ‘Library and Information Science’ (LIS) as an independent entity in the database and the field delineation is rigorously handled using JCR categories in the database.

Jonkers and Derrick (2012) describe a topical search method that applies the database WoS and we will repeat only a few details here. Our search profile is a slightly extended version of the one used by Jonkers and Derrick (2012). Their profile carefully deselected non-bibliometric items. The procedure produced a representative sample of documents for further analysis:

$$TS=(bibliometric* OR scientometric* OR webometric* OR altmetric* OR informetrics*)$$

Indexes: SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH.

Only genuine articles, letters, proceeding papers and reviews are included in the analysis. WoS does not have comprehensive coverage, but the database indexes the major journals in the field. WoS characterizes the individual articles as either LIS or belonging to the non-LIS group. The basis of this categorization is Thompson-Reuters JCR’s (Journal Citation Report) tagging of the journals. The only exception is ‘Proceedings of the International Conference on Scientometrics and Informetrics’. WoS does not identify correctly this major conference in bibliometrics and it belongs definitely to the LIS field.

In the LIS group, we do not attempt to subdivide the articles further into theoretical, or methodological studies and applied studies as was done earlier by Ellegaard and Wallin (2015). A few journals and hence articles get more than one attribute assigned to them. A LIS classification is probably only applied to papers with a significant contribution of LIS aspects and methodical considerations in the article. In case a LIS attribute is assigned, we consider the LIS aspect as most dominant and designate the article as LIS. This classification is parallel to that used by Jonkers and Derrick (2012), although these authors introduced an additional classification based on author affiliation. As a part of our search profile, we study primarily articles and assign them to different research areas based on WoS classifications and we choose to examine the 25 most popular fields as indicated by number of publications.

We apply a citation window that covers the whole period, including 2016. Citations indicate article impact although individual articles are, of course, not comparable due to different exposure time since their publication. The citing articles can be subdivided further into one of the two categories (LIS or non-LIS), publication year (citing year), or authors’ country of affiliation. In the same way, we analyse articles on
bibliometric analysis and their publication by certain countries and their citations as well. To this end, we examine countries with most publications in the field of bibliometrics and place them in in two groups. Group 1 is ten ‘Western’ countries from Europe, North America or Australia, which presumably have a long tradition in the bibliometric field. Group 2 is ten countries that represent ‘up-and-coming countries’ and, in recent years, assigned a large number of publications (Appendix A).

Placed in its own group are publications with authors from both groups of countries. This group is, of course, also interesting because of the possibility of cooperation between authors from ‘up-and-coming’ and ‘Western’ countries.

In order to extract data on the most important terms occurring in publications, we use the software tool VOSviewer that analyses and maps bibliographic data. It shows these terms organized in clusters and their co-occurrence relations (Van Eck and Waltman 2017). In a recent overview by Van Eck and Waltman (2010) further details on the software can be found.

Using VOSviewer, we analyse all the bibliometric articles in our portfolio and visualize networks in the data. A co-occurrence of keywords indexed by WoS and extracted from the articles can reveal clusters of articles on mutual interests. A keyword occurs only in one cluster. In this case, the program can also quantify data on the number of occurrences of interlinked keywords. 18080 keywords are downloaded from the VOSviewer program. We choose the 200 most abundant keywords that connect with 11260 mutual links. The program then identified eight different clusters of keywords. Separately, our analysis identifies the frequency of occurrence of the cluster keywords in both the LIS and the non-LIS literature. This method has the potential of revealing any field dependent difference in interests between the LIS and the non-LIS community and with special attention on the theoretical and methodical aspects of the articles.

Other methods characterize the literature on bibliometric analysis as well. They encompass co-author analysis (Jonkers 2012), co-citation analysis or bibliographic coupling schemes but our method that uses co-occurrence of keywords seems to be the most feasible to apply in our case.

Further, in order to investigate citation patterns, we analyze in total the 400 most cited articles with 200 in each category (either LIS or non-LIS). The articles are further subdivided into 8 different groups (A-H), as seen in Table 1, according to a modified version of a scheme applied earlier by Derrick et al. (2012), hereafter referred to as the DJL (Derrick, Jonkers and Lewison) scheme. We assign all publications to one or more groups (a maximum of three is applied) based on their content or the general methodology used. The method using descriptors makes it possible to investigate and compare the individual characteristics of bibliometric articles in detail, as well as their citing articles in different contexts. This provides information about the way they influence the LIS and non-LIS communities. We choose the most cited articles because of their impact, and they add most items to the pool of citing articles used for further analysis.
Table 1 List of descriptors used to characterise bibliometric papers, with single-letter codes. Based on a slightly modified version of the scheme used by Derrick et al. (2012).

<table>
<thead>
<tr>
<th>Code</th>
<th>Characteristic</th>
<th>Code</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Analyses a field or topic</td>
<td>E</td>
<td>Analyses collaboration, networks or author behaviour</td>
</tr>
<tr>
<td>B</td>
<td>Analyse journals or databases</td>
<td>F</td>
<td>Analyses, develops, discusses or improves bibliometric indicators, methods, theory or law</td>
</tr>
<tr>
<td>C</td>
<td>Analyse countries</td>
<td>G</td>
<td>Discusses policy implications, the merits (or not) of bibliometrics, peer review issues or evaluation systems</td>
</tr>
<tr>
<td>D</td>
<td>Analyses a researcher, group or organisation</td>
<td>H</td>
<td>Analyses patents</td>
</tr>
</tbody>
</table>

Results

In total, for further analysis, we extract 4637 articles in the non-LIS category and 4215 articles in the LIS category. The two groups obtained 40700 citations (85% of the citing articles classified as non-LIS and 15% as LIS) and 57862 citations (51% non-LIS citing articles and 49% LIS citing articles) respectively during the period 1964-2016. (Citation data extracted from WoS Ultimo February 2018). Our data confirm the same tendency already reported by Jonkers and Derrick (2012) that on average non-LIS articles on bibliometric analyses are cited less often than LIS articles in the same field (8.8 versus 13.7 times). It is evident that non-specialists who publish in the non-LIS literature are aware of and cite the literature on bibliometrics. In contrast, LIS researches cites less often non-LIS literature about bibliometric analysis although it is difficult to state absolute numbers because of the smaller numbers of persons involved.

Absolute and relative number of non-LIS bibliometric articles.

The number of bibliometric articles that analyse a given field or subject is important (Online resource 1, Table S1). In figure 1 with black bars, the subjects have been placed in decreasing order based on absolute number of publications in the field (e.g. Business economics has 649 publications and has been set to one while Sociology in the other end only gets 61 publications. In order to judge the true impact of the fields, we choose normalised numbers related to the entire production of publications as determined from WoS within the different fields. The data show the analysis of the relative number of non-LIS bibliometric
articles (bibliometric-articles/total production of articles about a topic) for the 25 most popular subjects (blue bars). The largest normalized value of ‘public administration’ has been set to one.

Fig. 1 Relative number of articles on non-LIS bibliometric analysis. Most popular subjects. 1964-2016. Data from Online resource 1, Table S1.

‘Business economics’ is the most studied subject from a bibliometric point of view. While the subjects ‘engineering’, ‘computer science’, and ‘science and technology’ draw a large number of bibliometric analyses, the picture is different when we consider the relative numbers. Studies of ‘public administration’ and ‘social science’ with bibliometric methods are numerous while ‘medicine’, ‘physics’, and ‘chemistry’ represent lesser studied issues.

The ratio between the numbers of articles produced by different communities on bibliometric analysis about a subject reflects their difference in interests. Fig. 2 shows this number in case of the ten ‘up-and-coming’ countries versus the ten similar ‘Western’ countries with most publications on a subject. The number of articles in each subject and country category has been normalized to the total number of articles on bibliometric analysis in each country category respectively. A relative ratio around one indicates that a subject is equally popular between the two groups of countries. A number greater than one indicates that bibliometric analyses on this subject obtain more publications in ‘up-and-coming’ countries compared to ‘Western’ countries. Figure 2 further demonstrates that articles on ‘energy fuels’ and ‘operations research management science’ but also a number of other topics are overrepresented in ‘up and coming’ countries. Medical topics, ‘psychology’ and ‘sociology’ are clearly underrepresented.
The impact of non-LIS articles on bibliometric analysis

With regard to impact as measured by the average number of citations obtained by an article in the most popular subjects, we observe especially high impact for ‘public administration’, energy fuels and business economics, while the impact is low for computer related subjects. In general, bibliometric analysis of health related studies are cited at an intermediary level. (Fig.3). There seems to be a slight linear correlation between the normalised number of articles on bibliometric analyses and citation impact of a given subject ($r^2=0.39$, $p=8.5E-4$ in Online resource 1, Table S2). A priori, we might expect that bibliometric analysis of well-investigated subjects also obtain a larger number of citations.
Fig. 3  Non-LIS Bibliometric analysis. Most popular subjects. Average number of citations per article within each field. Data from Online resource 1, Table S1

Figure 4 shows the relative interest in a given field between actors who participate directly and the LIS community. It displays the relative fraction of citations given by LIS to non-LIS articles on bibliometric analysis within a given field. Specialized fields like ‘education’, nursing and ‘surgery’ show high ratios. On the low side is the productive fields with a large number of publications like ‘computer science’ and ‘engineering’. Health science draws an almost equal interest from the two communities.
Fig. 4 Non-LIS Bibliometric analysis. Most popular subjects. LIS fraction of all articles that cite articles on a given field. Data from Online resource 1, Table S1

*Keyword analysis of the articles on bibliometric analysis. Occurrence of keywords in the LIS and the non-LIS literature.*

Based on data from the program *VOSviewer* on co-occurrence of keywords in bibliometric articles, eight different clusters in the LIS and non-LIS literature combined could be identified (Fig. 5). This figure shows the most pregnant keywords with interconnected links and the coloring illustrates the different clusters.
Figure 5. Clusters of articles on bibliometric analysis based on co-occurrence of keywords.

Online resource 2, Table S3, shows the different keywords along with their partition in clusters. As an example, Appendix B lists the keywords that represent the ‘technology and innovation cluster’.

We assign tentatively, in Table 2, all clusters to their pertinent fields. The data summarize for each cluster, the total number of keywords that occur in either the LIS or the non-LIS part of the literature. We multiply or normalize the LIS occurrence with a factor based on the number of articles non-LIS/LIS: 4637/4215=1.1. It allows us to compare the frequency of occurrence directly and by this, we can determine if some keywords in the cluster are relatively more prevalent in one of the two groups.
Table 2. Clusters of keywords with assigned categories and their occurrence in the literature on bibliometric analysis.

<table>
<thead>
<tr>
<th>Item category</th>
<th>Cluster</th>
<th>Occurrence all</th>
<th>non-LIS occurrence</th>
<th>LIS occurrence</th>
<th>LIS occurrence normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research trends and output. Health and environmental science</td>
<td>1</td>
<td>6180</td>
<td>4080</td>
<td>2100</td>
<td>2356</td>
</tr>
<tr>
<td>Impact. Coverage in databases</td>
<td>2</td>
<td>5822</td>
<td>2855</td>
<td>2967</td>
<td>3264</td>
</tr>
<tr>
<td>Productivity and performance in different fields</td>
<td>3</td>
<td>2460</td>
<td>949</td>
<td>1511</td>
<td>1662</td>
</tr>
<tr>
<td>Information science. Co-citation</td>
<td>4</td>
<td>4065</td>
<td>1771</td>
<td>2294</td>
<td>2523</td>
</tr>
<tr>
<td>Technology and innovation</td>
<td>5</td>
<td>3624</td>
<td>2085</td>
<td>1539</td>
<td>1693</td>
</tr>
<tr>
<td>Indicators and rankings</td>
<td>6</td>
<td>3206</td>
<td>1255</td>
<td>1951</td>
<td>2146</td>
</tr>
<tr>
<td>Collaboration and network</td>
<td>7</td>
<td>1811</td>
<td>694</td>
<td>1117</td>
<td>1229</td>
</tr>
<tr>
<td>Webometrics and Infometrics</td>
<td>8</td>
<td>1298</td>
<td>410</td>
<td>888</td>
<td>977</td>
</tr>
</tbody>
</table>

It is evident that keywords relevant for the two clusters: ‘Research trends and output. Health and environmental science’ and ‘Technology and innovation’ are most dominant in the non-LIS literature. The keywords that relate more to discussions of indicators, collaboration, networking as well as infometrics are more abundant in the LIS-literature. It seems also that ‘Investigations of productivity and performance in different fields’ are relatively more dominant in LIS. On the other hand, we observe, in all the clusters investigated, a significant contribution from both the LIS and the non-LIS group.

In order to investigate the impact of literature that discusses aspects of bibliometric theory and methods in the two communities, we extract relevant and pregnant keywords from the predominant LIS clusters. These keywords from the clusters 3, 4, 6-8 are shown in Online Resource 2, Table S3 and they are used to find theoretical and methodical articles among all the LIS and non-LIS literature on bibliometric analysis.

A search in the title of the articles finds 1100 LIS articles and 607 non-LIS articles. The numbers are, after evaluating the content of the individual articles, reduced to 540 LIS articles and 276 non-LIS articles that match the subject of our investigation. We summarize the overall findings in Table 3:
Table 3. Number of LIS or non-LIS articles on theoretical or methodical aspects of bibliometric analysis and the respective numbers of citing articles in the two communities.

<table>
<thead>
<tr>
<th>Community</th>
<th>Articles on theoretical or methodical aspects</th>
<th>LIS-citing articles</th>
<th>%</th>
<th>Non-LIS citing articles</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-LIS</td>
<td>276</td>
<td>806</td>
<td>33,8</td>
<td>1782</td>
<td>66,2</td>
</tr>
<tr>
<td>LIS</td>
<td>540</td>
<td>3387</td>
<td>61,1</td>
<td>2159</td>
<td>38,9</td>
</tr>
</tbody>
</table>

It is evident that LIS journals publish most of the articles in the above categories. A fair percentage (34%) of our sample is published in non-LIS journals and it indicates that the discussion about interpretation and improvement of bibliometric methods is widespread among the non-professionals in the field. If we look at the citations, it is clear that LIS-professionals are not particular aware of the methodic discussions that take place outside their own field. More surprisingly, the non-LIS community is responsible for almost 39% of all articles that cites LIS publications with this content but because the possible numbers of citers outside the LIS field are much greater, it is impossible to compare absolute numbers. In fact, a percentage based on normalized numbers should be much lower.

If we instead consider all articles on bibliometric analysis published in LIS journals, the same number is somewhat higher 51%.

Characteristic of the most cited articles on bibliometric analysis. LIS or non-LIS publications.

Next, we focus on the 200 most cited articles of either non-LIS or the 200 most cited articles of LIS origin (Online resource 3, Table S5-S8). We use the modified DJL scheme (Table 1) to characterize the articles and show the data on occurrence in each category in Fig. 6. It is evident from the figure that the dominant categories in non-LIS publications are #A and #G. They include numerous analyses of different topics and to a lesser extent articles on ‘discussions of policy implications and the merits (or not) of bibliometrics’. In contrast, LIS publications dominate the groups: #E-#G that are to a greater degree concerned with collaboration, author behavior and the methods and merits of bibliometrics.
Fig. 6  The 200 most cited articles on bibliometric analysis in 1964-2016. The figure shows the number of articles in characteristic DJL categories. Data from Online resource 3, Table S5 and S8.

All articles as well as the most cited articles on bibliometric analysis. Countries of origin.

Table 4 compares the 200 most cited non-LIS articles with all non-LIS articles on bibliometric analysis. It demonstrates that ‘up-and-coming’ countries publish a significantly smaller part of the most cited publications compared to all non-LIS articles on bibliometric analysis. In fact, authors from the ten most dominant ‘up-and-coming’ countries publish almost one quarter of all non-LIS articles on bibliometric analysis. The same number is only 7.5% for the 200 most cited articles. In case of LIS articles on bibliometric analysis, one observes the same trend although the numbers are slightly lower. It confirms that authors from ‘Western’ countries publish the most cited articles. In fact, authors from the two groups cooperate in 4.0%-5.7% of all articles on bibliometric analysis. The overall distribution between the three categories 10-‘West’, 10-‘up-and-coming’ and 10-‘up-and-coming’ and 10-‘West’ cooperation’ is dependent on whether we consider the non-LIS or LIS group ($\chi^2<4.2E-8$ in Table S9, additional material). The ten most productive countries from the ‘West’ produce a relatively larger number of LIS articles on bibliometric analysis.
Table 4  Articles on bibliometric analysis (BA). 1964-2016. 10-‘West’ and 10-‘up-and-coming’ indicate the ten countries with most publications on bibliometric analysis in each group. 10-‘up-and-coming’ and 10-‘West’ corp. indicates articles where authors from both groups contribute.

<table>
<thead>
<tr>
<th></th>
<th>non-LIS</th>
<th>%</th>
<th>200 most cited</th>
<th>%</th>
<th>LIS</th>
<th>%</th>
<th>200 most cited</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td></td>
<td></td>
<td>non-LIS BA</td>
<td></td>
<td></td>
<td></td>
<td>LIS BA</td>
<td></td>
</tr>
<tr>
<td>All articles</td>
<td>4637</td>
<td>100</td>
<td>200</td>
<td>100</td>
<td>4215</td>
<td>100</td>
<td>200</td>
<td>100</td>
</tr>
<tr>
<td>10-West</td>
<td>2361</td>
<td>50.9</td>
<td>146</td>
<td>73.0</td>
<td>2406</td>
<td>56.9</td>
<td>157</td>
<td>78.5</td>
</tr>
<tr>
<td>10-‘up-and-coming’</td>
<td>1140</td>
<td>24.6</td>
<td>15</td>
<td>7.5</td>
<td>867</td>
<td>20.5</td>
<td>8</td>
<td>4.0</td>
</tr>
<tr>
<td>10-‘up-and-coming’ and 10-‘West’ corp.</td>
<td>266</td>
<td>5.7</td>
<td>9</td>
<td>4.5</td>
<td>205</td>
<td>4.8</td>
<td>8</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Characteristics of the most cited articles on bibliometric analysis. Total number of citations.

The distribution of the relative number of citations (citations per item in each category) shows a more complicated pattern for the most cited non-LIS articles on bibliometric analysis (Fig. 7). The citations to the 200 most cited non-LIS articles account for 42 % of all citations given to the 200 most cited articles either non-LIS or LIS. Apparently, articles in the categories #A-#C that analyze specific fields, authors or countries are cited less than articles that analyse and discuss bibliometric methods, i.e. has a lower impact. All articles in the categories #D-#G that analyze author behavior or discuss bibliometric methods are highly cited. This is especially valid for non-LIS articles in category #F and the numbers of citations are even larger than for LIS articles in the same category. This shows clearly that the non-LIS community takes advantage of papers that discuss the implications of using bibliometric methods. If we consider instead the 58% citations that is allocated to the 200 most cited LIS articles on bibliometric analysis, the categories #D: ‘analyses a researcher, group or organisation’ and #E: ‘analyses collaboration, networks or author behaviour’ have most citations per article. #H (patents) has a large citation value but based on very few articles it is not statistically significant.
Fig. 7 The 200 most cited articles on bibliometric analysis in characteristic DJL categories either non-LIS or LIS. 1964-2016. The figure shows the number of citing articles per item in each category. Data from Online resource 3, Table S8.

Characteristics of the most cited articles on bibliometric analysis. Origin of citations.

Figure 8 shows the mutual interest in citing articles from the opposite field. If we first consider the 200 most cited non-LIS articles on bibliometric analysis in the different categories, the fraction of articles from the LIS field that cites is especially small for category A: ‘analyses a field or topic’. On the other hand, the LIS community contributes with a significant fraction of citations to non-LIS articles in all other categories. Next, we look at the same numbers for the fraction of non-LIS articles that cite the 200 most cited LIS articles on bibliometric analysis in different categories. It is evident that the non-LIS community cites both theoretical as well as practical aspects of LIS articles on bibliometric analysis. The fractions are larger simply because the sheer numbers of non-LIS articles are significantly higher than the similar number of LIS articles.
**Fig. 8** The 200 most cited articles on bibliometric analysis in characteristic DJL categories either non-LIS or LIS. 1964-2015. The figure shows the fraction of articles that cite articles from the opposite field in each category. Data from Online resource 3, Table S8

**Discussion and conclusion**

It is evident that information obtained by bibliometric methods plays an increasing role in evaluation of search. The bibliometric analysis provides an overview and contributes to information about the literature in a given field. Sophisticated statistical methods or metrics could provide a more profound analysis of performance and productivity measures. As an example, the bibliometric literature can point out the directions different fields move scientifically and inform about the current discussion regarding the assessment of science and its practitioners.

This is especially true with regard to the use of different metrics that can quantify many otherwise intangible concepts. Discussion of different indicators and normalization procedures takes place in the literature as well.

Bibliometric indicators are often the basis for funding and financing of research. Research administrators and politicians consider extracts from bibliometric indicators published in different rankings when they are making decisions. The rankings themselves and their preconditions are obviously not always transparent to those involved.
The recipients of reports need a clear insight into bibliometric methodology and methods. This, of course, necessitates that the LIS community understand the requirements and needs of their intended customers. Bibliometricians need to describe the pros and cons of using the different databases and, not least, that in many cases one should interpret the results based on the underlying dataset. A straightforward example could be an H-index calculated from an author’s publications and citations. The result depends strongly on the indexing in various databases (Bar-Ilan 2008).

Methodology must be applied that can be understood with basic prerequisites. LIS professionals must be aware of changing publication patterns as well as methods used by the stakeholders of bibliometric analyses. There are various ways to accomplish border crossing between LIS and non-LIS. It is important, if LIS researchers aim to have a wider societal impact of their research, that either the two groups cooperate directly or LIS researchers publish their results in the non-LIS literature (Derrick et al 2012). In this way, LIS researchers can point at problems of which the other group must be aware. Further, they can gain insight into the use of different methods by non-professionals in the bibliometric field.

If we consider things from the opposite perspective, professional stakeholders with minor knowledge of the field use increasingly ready-made bibliometric solutions by large database providers, e.g. WoS or Scopus (Cox et al. 2017). The risk exists that non-transparent aspects of bibliometric investigations such as coverage of the literature in different fields and databases, normalization or use of different metrics are misinterpreted. Generic types of bibliometric analyses that involve publication in or growth of research fields apparently pose only few problems. In contrast, development of indicators and research impact measures are concepts that are more difficult to get hold on. As a step forward that may improve standards, we observe an increasing number of editorials published in non-LIS journals that discuss implementation of bibliometric measures in their field.

The field has seen a large growth in LIS as well as non-LIS literature on bibliometric analysis over of the last few decades as already documented by Jonkers and Derrick (2012) and Ellegaard and Wallin (2015). Our data provides valuable information on how different practitioners as well as library personnel support the use of bibliometrics. Until now there has been a significant lack in contributions to the understanding of professional roles in this context (Cox et al. 2017).

The present work demonstrates a marked difference between the numbers of published articles on bibliometric analysis in various fields. This trend is especially clear if we normalize the number of non-LIS articles on bibliometric analysis with data for the total amount of publications in each field. Our investigation shows that e.g. ‘management’ and ‘public administration’ articles seem overrepresented in the literature. In absolute numbers, the coverage of medical and health areas from a bibliometric point of view is above average, though not when we compare to the complete corpus of literature on these subjects. The ratio between the number of articles on bibliometric analysis of ‘up-and-coming’ origin and those from ‘Western’ countries shows that analyses of ‘Operation Management’ and ‘Energy Fuels’ dominate in the former. Relatively more frequently published in ‘Western’ countries are ‘Health Care Science’ and
‘Sociology’ related issues. These considerations indicate that authors from ‘up-and-coming’ countries to some degree prefer analysing literature in other fields than authors from ‘Western’ countries. Apparently, the former authors focus on industrial and business applications while ‘Western’ countries dominate the literature on more human-related issues.

Jonkers and Derrick (2012) as well as Ellegaard and Wallin (2015) found that non-LIS literature has generally low visibility in the form of citation impact. Probably, many analyses published in specialized, non-bibliometric journals are cited mainly by colleagues in the field and to a lesser degree by LIS researchers (Jonkers and Derrick 2012). As observed in fig. 3, the citation impact is also field dependent, but apart from ‘public administration’, which shows a high citation impact, the difference is not very large. This relatively large value is most likely due to the large number of bibliometric analyses published on the subject. The fraction of LIS-citations may indicate the relative impact of a given subject in the two communities. As shown in Fig. 4, specialised fields such as ‘Energy Fuels’ and ‘Specialized Health Services’ obtain low fractions, while analyses related to ‘Sociology’ and ‘Public Administration’ are well cited by the LIS community. Obviously, analyses of the literature about these issues seem to be particularly important to this community.

An analysis of co-occurrence of keywords visualizes the way the literature on bibliometric is interlinked. By doing this, we obtain detailed knowledge about the content of bibliometric articles. Then we can assess on an informed basis about deficiencies, interpretations or collaborative relationships in this type of work. This may benefit the use of bibliometric methods or contribute to the correct understanding of the results. The software program VOSviewer allows us to quantify the occurrence of keywords in different subject clusters. We find some subject fields represented more often in the non-LIS literature whereby one observes a skewed distribution of keywords between the LIS and non-LIS literature. Subjects such as ‘Research trends and output’ especially with relation to ‘Health and environmental science’ and ‘Technology and innovation’ are more frequently published in the non-LIS literature. The LIS literature dominates the more traditional subjects related to interpreting and improving bibliometric methods. In addition, collaboration and networking are mainly LIS-topics.

Our keyword analysis demonstrates the representation of all subjects in both types of literature. Both groups seem to communicate results to an audience outside their traditional field of interest. The impact of such border crossing between the LIS and non-LIS is traditionally evaluated using citation analysis. Our keyword analysis demonstrated that relative few from the LIS community actually cited bibliometric articles on theoretical and methodical issues published outside their own domain.

A fair percentage of articles that cite LIS articles in the above category actually are from non-LIS. It indicates a certain awareness of methodical issues but the number is still below the average percentage of non-LIS citations to all LIS articles on bibliometric analysis.

In order to obtain detailed information, we considered the characteristics of the 200 most cited non-LIS or LIS publications according to the DJL-scheme. Articles on ‘Analyses of a field or topic’ occur most frequently. These articles compromises ~ 68% of all articles and they are well above the result obtained by
Derrick et al. (2012) ~25%, but their data include all articles about bibliometric analyses. The number of papers in #G, ‘Discusses policy implications, the merits (or not) of bibliometrics...’, is significant as well. This finding demonstrates that non-LIS authors take part in the discussions about bibliometric methods and results. Then it is more likely that non-professionals in the bibliometric field can fulfill some of the demands advocated in the Leiden Manifesto (Hicks et al. 2015). It is evident that one may find only a few well-cited non-LIS papers on bibliometric analyses outside the categories earmarked for practical bibliometric analyses. In contrast, the most cited papers on bibliometric analyses published in LIS journals deal with theoretical aspects, including discussion and evaluation of the field.

The different focus of the specialized LIS field and the recipients of bibliometric information have been on the agenda since Glänzel (1996) advocated the need for standards in bibliometric research.

The rising tide of literature in the field of bibliometric analyses, and especially that published by ‘up and coming’ countries, necessitates a discussion of the implementation of the fields’ standards and methods. In the present work, we analyse one part of this problem area by studying, distributed into different categories, the subjects of papers on bibliometric analyses as well as their citation impact. The data shows that with regard to citation impact, in most categories, non-LIS papers are evenly distributed. The most significant result is a generally low citation impact for the applied categories #A- #D but, more importantly, the fact that the LIS community consistently cites non-LIS papers in the categories #E- #G that deal with the merits of bibliometrics as well as improvement of bibliometric methods. This confirms a certain awareness of these types of papers that could prove important to the discussion of bibliometric evaluations in the two communities. In the same manner non-LIS authors seem to cite authors that publish LIS articles on bibliometric analysis in all categories. It demonstrates that many authors that publish on bibliometric analyses in the specialist non-LIS literature are aware of and know the LIS literature.

The same discussion takes place from the viewpoint of non-LIS publications on bibliometric analyses by authors from ‘Western’ and ‘up-and-coming’ countries, as well as those countries where mixed author teams cooperate and publish articles. One must see these discussions in the light of the fact that only 15% of the 200 most cited non-LIS publications on bibliometric analysis have an author team from ‘up-and-coming’ countries. The same number for all non-LIS articles on bibliometric analysis 24.6%. The number for mixed teams from the two tiers is 5.7%. (Table 3). This number increases to 9% for the 200 most cited articles and proves the benefit of cooperation.

These numbers are not very different in the case of LIS articles on bibliometric analysis. Jonkers and Derrick (2012) investigated already mixed teams, which are due to different author affiliations, i.e. a mix of authors that belong to LIS as well as the non-LIS community. They found a significant increase in the number of mixed team publications compared to other types, and took this as evidence of a gradual decrease in mutual misconceptions between the two sides.

We have tried to synthesize information about production, impact, areas of interest and general characteristics of bibliometric analyses and interpreted the findings from the standpoint of those who either produce or use the results. Some specific fields are less developed bibliometrically, and bibliometric
analyses of computer-related science as well as exact sciences often have low impact. Many articles analyses topics, countries, researchers, or institutions and are probably read by many, but cited to a minor degree. Theoretical and methodical articles but also articles on analyses of persons or organizations are the most cited articles by both the LIS and non-LIS community, a fact that could secure better interpretation of results as well as improvements of standards in the field.

The general and rather crude division into either ‘Western’ or ‘up-and-coming countries’ provided some insight into different spheres of interest among the two groups. Bibliometric analyses from ‘Western’ countries often invoke theoretical and methodical papers as well as health science in a broader context. Analyses published by authors from ‘up-and-coming’ countries show an overweight of papers that relate to issues on industrial applications. Apparently, until now there has been relatively little focus on more actual or acute problems such as the exhaustion of resources, environmental degradation and the consequences of global warming, i.e. problems that could be highly relevant to policy makers, administrators as well as local communities. The number of primary publications on these issues is growing steadily and with increasing bibliometric coverage. It might be quite interesting to see whether e.g. authors that publish in social science journals would prioritise the subjects of bibliometric analyses differently from authors in natural science.

A more thorough analysis of the characteristics (groups, #A-#L) of bibliometric analysis, from the perspective of both primary articles and their citing articles, may shed further light on some of these problems. Targeting the limited resources of bibliometricians could also be highly beneficial in order to meet the needs of the scientific and professional communities. These latter groups as well as research administrators will then be more likely to benefit from the outcome of bibliometric methods and analyses.

In conclusion, our data has provided valuable information on how different practitioners as well as information specialists and library personnel support the use of bibliometrics. Establishing and expanding a knowledge base about publishing behavior of those involved is obviously vital in this respect. The bibliometric literature reflects the position and status of the field. In the present analysis, we investigated who took advantage of the different methods and why we observe an increasing number of publications in the field. The ongoing discussion of the role of metrics and standards in research assessment makes it even more urgent to know the foundation on which the bibliometric literature rests.

Acknowledgement The author wishes to thank Ph.D. Mette Bruus and anonymous referees for valuable comments and suggestions for improvement of the article.
### Appendix A Countries with most publications on bibliometric analysis 1964-2016.

<table>
<thead>
<tr>
<th>“Western”</th>
<th>“Up-and-coming”</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>China</td>
</tr>
<tr>
<td>Spain</td>
<td>Brazil</td>
</tr>
<tr>
<td>England</td>
<td>India</td>
</tr>
<tr>
<td>Germany</td>
<td>Taiwan</td>
</tr>
<tr>
<td>Netherland</td>
<td>South-Korea</td>
</tr>
<tr>
<td>Italy</td>
<td>South-Africa</td>
</tr>
<tr>
<td>Canada</td>
<td>Mexico</td>
</tr>
<tr>
<td>France</td>
<td>Iran</td>
</tr>
<tr>
<td>Belgium</td>
<td>Malaysia</td>
</tr>
<tr>
<td>Australia</td>
<td>Chile</td>
</tr>
</tbody>
</table>
Appendix B  Keywords represented in cluster 5: ‘Technology and innovation’ and their occurrence in the literature on bibliometric analysis.

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Occurrence all</th>
<th>Non-LIS occurrence</th>
<th>LIS occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>bibliometrics</td>
<td>1934</td>
<td>1201</td>
<td>733</td>
</tr>
<tr>
<td>technology</td>
<td>219</td>
<td>108</td>
<td>111</td>
</tr>
<tr>
<td>innovation</td>
<td>219</td>
<td>132</td>
<td>87</td>
</tr>
<tr>
<td>knowledge</td>
<td>209</td>
<td>127</td>
<td>82</td>
</tr>
<tr>
<td>nanotechnology</td>
<td>101</td>
<td>52</td>
<td>49</td>
</tr>
<tr>
<td>growth</td>
<td>92</td>
<td>48</td>
<td>44</td>
</tr>
<tr>
<td>system</td>
<td>79</td>
<td>52</td>
<td>27</td>
</tr>
<tr>
<td>text mining</td>
<td>79</td>
<td>47</td>
<td>32</td>
</tr>
<tr>
<td>interdisciplinarity</td>
<td>72</td>
<td>27</td>
<td>45</td>
</tr>
<tr>
<td>industry</td>
<td>68</td>
<td>45</td>
<td>23</td>
</tr>
<tr>
<td>research-and-development</td>
<td>67</td>
<td>11</td>
<td>56</td>
</tr>
<tr>
<td>dynamics</td>
<td>64</td>
<td>28</td>
<td>36</td>
</tr>
<tr>
<td>patents</td>
<td>58</td>
<td>32</td>
<td>26</td>
</tr>
<tr>
<td>statistics</td>
<td>57</td>
<td>31</td>
<td>26</td>
</tr>
<tr>
<td>diffusion</td>
<td>52</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>biotechnology</td>
<td>50</td>
<td>23</td>
<td>27</td>
</tr>
<tr>
<td>nanoscience</td>
<td>45</td>
<td>19</td>
<td>26</td>
</tr>
<tr>
<td>science policy</td>
<td>43</td>
<td>19</td>
<td>24</td>
</tr>
<tr>
<td>database tomography</td>
<td>42</td>
<td>33</td>
<td>9</td>
</tr>
<tr>
<td>science-and-technology</td>
<td>38</td>
<td>10</td>
<td>28</td>
</tr>
<tr>
<td>organization</td>
<td>36</td>
<td>14</td>
<td>22</td>
</tr>
<tr>
<td>All</td>
<td>3624</td>
<td>2085</td>
<td>1539</td>
</tr>
</tbody>
</table>
References


Van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics, 111*(2) 1053-1070


