JSCIRES RESEARCH ARTICLE

The bibliometric community as reflected by its own methodology

Dirk Tunger*, Julia Wilhelm[1]

Central Library, Forschungszentrum Jülich GmbH, 52425 Jülich, [1]KIT-Library, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

ABSTRACT

The field of bibliometrics is currently enjoying a growth in popularity. At least, this is the impression we get if we estimate the number of bibliometrics-related publications, conference papers and studies that appear in the form of grey literature. For a more accurate picture, we must examine the bibliometric community more closely. One way of doing this is to carry out a bibliometric analysis.

Keywords: Bibliometric community, development of the bibliometric community, network of the bibliometric community, scientometric community

INTRODUCTION

The bibliometric community analyses scientific publications with a set of indicators under different scopes. It is the aim to recognize scientific trends, to validate statements about scientific development and to look on actors and institutions in the scientific landscape.

One aim of this paper is to carry out a bibliometric analysis about the bibliometric community[1].[1] Questions to be answered in this paper are: "What is the developing trend on bibliometric publications" and "what changes can be seen in the co-publication activitiy?" As well there is the question, what changes on a topic level are visible between 1980 and 2011.

The paper aims to analyze the scientific community of bibliometrics and scientometrics publications. The field of bibliometrics is currently enjoying a growth in popularity. At least, this is the impression we get if we estimate the number of bibliometrics-related publications, conference papers and studies that appear in the form of grey literature. For a more accurate picture, we must examine the bibliometric community more closely. One way of doing this is to carry out a bibliometric analysis.[2]

METHODS

The first step in performing a bibliometric analysis is to formulate a suitable search query. An important factor to consider in this respect is the principle of precision and recall, which ensures that the final search results contain only relevant hits, while also ensuring that as many relevant hits as possible are returned.

When creating the search query, it must be noted that while some search terms are clearly associated with bibliometrics, others can just as often appear in

1 Julia Wilhelm studied the bibliometric community as part of her diploma thesis at Cologne University of Applied Sciences, Faculty of Information Studies and Communication Sciences. She examined publication behaviour, analysed networks and discussed individuals in detail. Wilhelm, Julia: Bibliometrische Analyse der Wissenschaftlichen Community "Bibliometrie und Scientometrie;" submitted to Prof. Dr. Simone Fühles-Ubach and Dr. Dirk Tunger, Central Library, Forschungszentrum Jülich, Cologne 2011.

***Address for correspondence:**
E-mail: d.tunger@fz-juelich.de

Access this article online

Quick Response Code:

Website: www.jscires.org

DOI: 10.4103/2320-0057.128990

other contexts. For this reason, it is advisable to base the search query on the topics covered by the key journals (Scientometrics and Journal of Informetrics). With this in mind, the scientific publications that appeared in these two journals in the last 5 years were searched to find suitable keywords. The search of possible keywords was limited to the time period 2007-2011, because this kind of work is very time consuming. While using the core journals and very often used terms such as "bibliometric*" or "scientometric*," a majority of publications will be found. Adding the keywords in #1 and #2, nearly all bibliometric papers (that are relevant for this study) in Web of Science are found.

The keywords were then divided into two sets: the first set of keywords (#1) is very specific to the topic, whereas the second set (#2) only produces useful results when combined with specific keywords. These two sets of keywords form the basic framework of the search query used. A third set (#3) was added to these, comprising all of the publications in the two bibliometrics journals, Scientometrics and Journal of Informetrics.

The parts of the search query were formulated as follows:

#1 (Specific Keywords)

TS = ("self-citation correction" or "site interlinking" or "uncitedness" or "web impact factor*" or bibliometric* or scientometric* or "citation age data" or "citation database*" or "citation distribution*" or "citation habit*" or "citation index*" or "citation metrics" or "citation network analysis" or "citation patterns" or "citation perspective" or "citations analyze*" or "co-citation cluster*" or "co-word analys*" or "disciplinary citation impact" or "generalized impact factor*" or "Google PageRank algorithm*" or "greater research impact" or "h-Index sequenc*" or "Hirsch-core" or "Hirsch-index" or "Hirsch-type index*" or "in-formetric*" or "internet citation*" or "Jin a-index" or "journal influence" or "journal qualimetric*" or "journal rank*" or "journal self-citation*" or "Kosmulski's h (2) - index" or "long-term citation impact" or "Lotkaian informetric*" or "Matthew core journal*" or "publication delay*" or "publication output" or "scientific-research output" or informetric* or webometric*).

The field tag "TS" allows the user to search for topic terms in fields such as "Title," "Abstract" or "Keywords" in Web of Science.

#2 (Less Topic-Specific Keywords, Which Were Searched for in Combination with More Specific Keywords)

TS = ("impact factor" or "peer review" or "cited papers" or "co-citation" or "evaluating research" or "g-index" or "journal peer review" or "journal-citation-report" or "multiple authorship" or "publication productivity" or "research performance" or "r-index" or "research trend" or "Scopus" or "self-citation*" or "word analys*" or "Zipf's law") and TS = ("author self-citation*" or bibliometric* or citation* or egghe* or "Hirsch index" or "Hirsch-core" or "Hirsch-type indic*" or "impact factor*" or informetric* or "journal influence*" or "Lotka*" or "Matthew core journal*" or "patent citation analys*" or "research output index" or "scientific research output" or scientometric* or "sleeping beaut*" or "uncitedness" or webometric* or ranking).

The search operator "and" was used in the search query to search for a combination of two keywords. A document was only returned in the hits if it contained keywords from both the first and the second set of brackets in #2.

#3 (Bibliometrics Journals)

SO = ("Journal of Informetrics" or "Scientometrics")

#3 records all of the publications of these two journals focusing on the topics of bibliometrics/scientometrics, irrespective of keywords used.

These parts were linked by "or" to form the overall search strategy, which was the basis of all further evaluations.

In comparison to a similar search strategy, created by Bar-Ilan,[3] the underlying strategy of this article seems to be more precise, because less topic-specific keywords are combined with more specific keywords. For example, if one uses the term "impact factor" as a search criterion, he will as well retrieve results like the following: "[…] the prestressing impact-factor method was adopted to deduce semiempirical equations for cracking torque in such composite beams […]."[4]

The number of journals in the underlying search strategy is not restricted to "Journal of Informetrics" or "Scientometrics," but these two journals are the publication basis for the complete analysis, because they are the scientific core journals of the topic "bibliometrics/scientometrics." Based on the search strategy with topic related keywords, all documents are found, that match with the used keywords in #1 and #2.

Publication Development

Figure 1 compares the publication numbers of the bibliometric community, those of the two bibliometrics journals Scientometrics and Journal of Informetrics and the total publication numbers of the Web of Science database. In 1980, the bibliometric community published a total of 50 bibliometrics publications; from 1980 to 2009 the community's publication output increased about 17 times. It can also be clearly seen that there was a sharp increase in the community's publication numbers in 2005 compared to the previous year. The number of publications rose from 265 in 2004 to 453 in 2005. Furthermore, the number of publications doubled again from 2005 to 2009. Compared to the overall developments in Web of Science (where total output "only" increased 2.5 times), we can therefore state that the increase in bibliometrics publications was disproportionately high. There was also a disproportionate growth in the publication output of the two bibliometrics journals, but it is much smaller than the growth of the bibliometric community as a whole. This indicates that a large proportion of the bibliometric community's publications, particularly in recent years, were published in journals, which are not directly associated with the community itself. In other words, bibliometrics is also generating interest outside its own community. Bibliometrics has ultimately become a tool that is used by the entire (natural) scientific world and has found a place in the world beyond its own specialist community.

How can this be explained? The development may be linked to the fact that greater attention is being paid to quantitative elements in research output. For instance, 2005 saw the launch of the excellence initiative in Germany. One of the objectives of the excellence initiative is to provide funding for outstanding research at German universities through the federal government and the federal states. The excellence initiative supports top-level academic research and raises its international profile, creates exceptional conditions for young scientists at universities, intensifies cooperation between disciplines and institutions, strengthens international research networks, promotes equal opportunities for men and women in science, boosts scientific competition in Germany and improves the overall quality of Germany as a scientific research location.[5]

Prior to the launch of the initiative, the German Council of Science and Humanities published a press release on 12 November 2004 on the subject of increasing transparency in the research sector through benchmarking.[6] This press release called for transparency in research activities in both universities and non-university research institutions through the use of a research ranking system.

The excellence initiative is just one example from Germany and one example for metric-based evaluation systems. Similar examples can be given for many different countries and in many different forms.[7] Often they are connected to the foundation of research groups that are dealing with

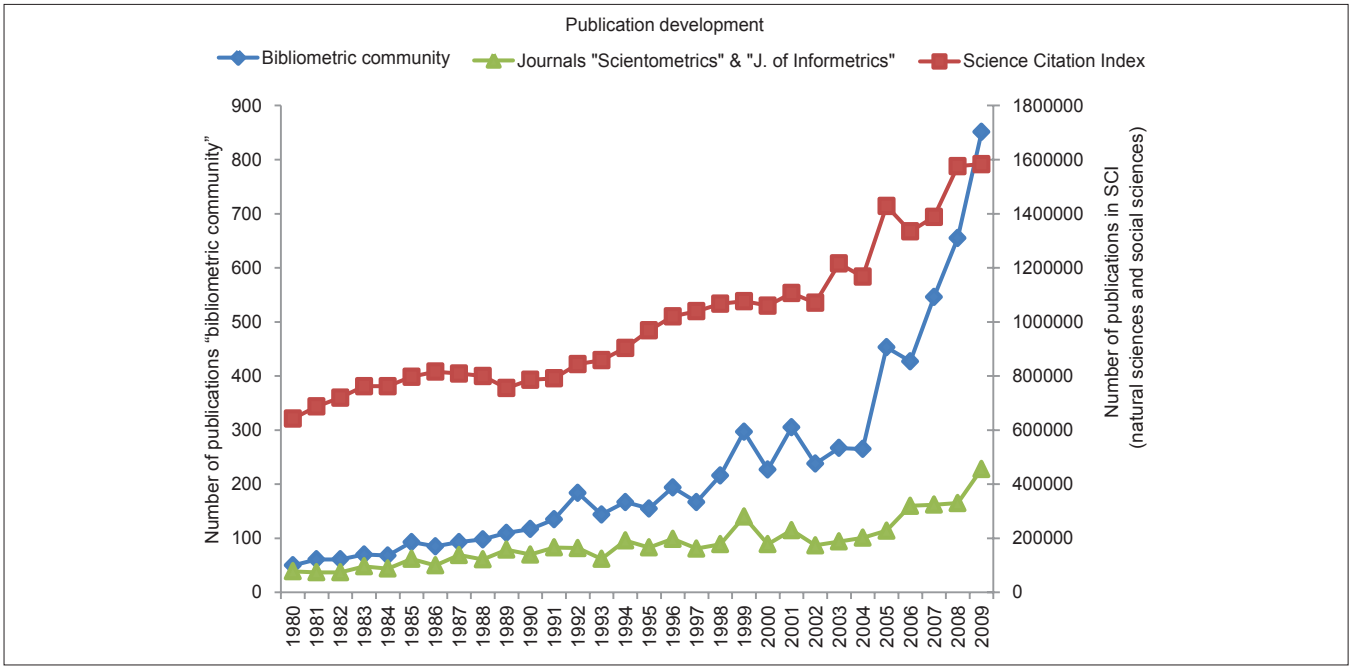

Figure 1: Publication development

bibliometrics and that are also active in the scientific work. This means, that these groups also publish scientific papers.

These are just a few examples of the increased public interest in research output and its quantitative development. Many more examples can be found in both university and non-university research institutions in Germany and further afield. Thus, a possible explanation for the rising number of scientific publications on topics relating to bibliometrics could be the heightened public interest in this issue, coupled with an increase in the number of actors who not only carry out commissioned work, but are also employed in the field of science.

Another phenomenon is the 2005 published "h-Index" that led also to lots of publications in a very short time. There are approximately 530 Publications dealing with this indicator in the time period from 2005 to 2011 in the given search strategy. These publications are part of the publication increase, but they are not the cause.

The increased number of people publishing on bibliometrics-related topics is not only evident in the number of participants attending international conferences, but also in the network of the bibliometric community.

The Network of the Bibliometric Community

Definition of networks

"Network: Term used to describe interhuman relations in a social system. In a graphical representation, points designate people and lines designate their mutual relationships. The whole of society in a given area can be broken down into social networks and analyzed in this framework."[8]

Functions and aims of author networks

One of the main functions and aims of an author network is to describe and analyze the relationships between actors. These actors may be scientific authors, institutions, countries, etc., Relationships between actors may be scientific publications, contacts, memberships or exchange partnerships. Author networks make it possible to show based on strategic positioning, whether an author is particularly relevant in a network and to classify the author in a thematic structure. A graphic representation of a network illustrates the structure of information links within scientific disciplines at a given time. By ascertaining which actors are at the center of an author network, we also gain a better idea of who the experts in that particular community are. Hence it is important to know for reading the network, that the position of a person in the network says something about his status and his relevance for the displayed community. The more central a person is in the network, the more important he or she is for the analyzed community. The networks do not have any axes. Persons that often publish together are also arranged closely together on the diagram.

Using the search query outlined at the beginning of this paper, a co-publication network analysis was conducted for the periods 2000-2004 [Figure 2] and 2005-2009 [Figure 3]. These two time periods were chosen to make actual trends of the bibliometric community visible. They should visualize the current activity in the first decade of this millennium.

The program used for this task was Network Workbench. This program was developed by Indiana University (USA) in 2005 for performing network analyses. It is based on Cyberinfrastructure Shell and is available as open-source software. The advantage of Network Workbench is that it allows the user to create a network – both online and offline – from a large volume of data within a reasonable amount of time.

The next step was to load the hits into Network Workbench so that an author network could be generated from these raw data. Information such as the author's name (label), number of citations (times cited), number of publications (number of works) and a unique node number (unique index) were recorded in an Excel file. This unique node number is important when cleaning author name data. It helps to conflate names that have been spelled incorrectly with those that have been spelled properly, as there may be occasions where there are two versions of a single author's name in the table – one correct and one incorrect. In order to conflate both versions, the unique node number of the incorrect version is overwritten with that of the correct version. If only the name were to be corrected, the two same names would appear later in the network. To prevent this, both versions of an author's name must be conflated under one unique node number. All associated data are also conflated along with the author's name.

This network consists of 1805 nodes, which equates to a community of 1805 authors in this 5 year period. Out of these 1805 nodes, 232 are isolated. This means that 13% of the authors in the community did not have any connections to other authors within the specified period. Furthermore, there is evidence of a greater readiness to co-authorship in that 87% of publications were produced in collaboration with at least one other author. It should be noted that 25%

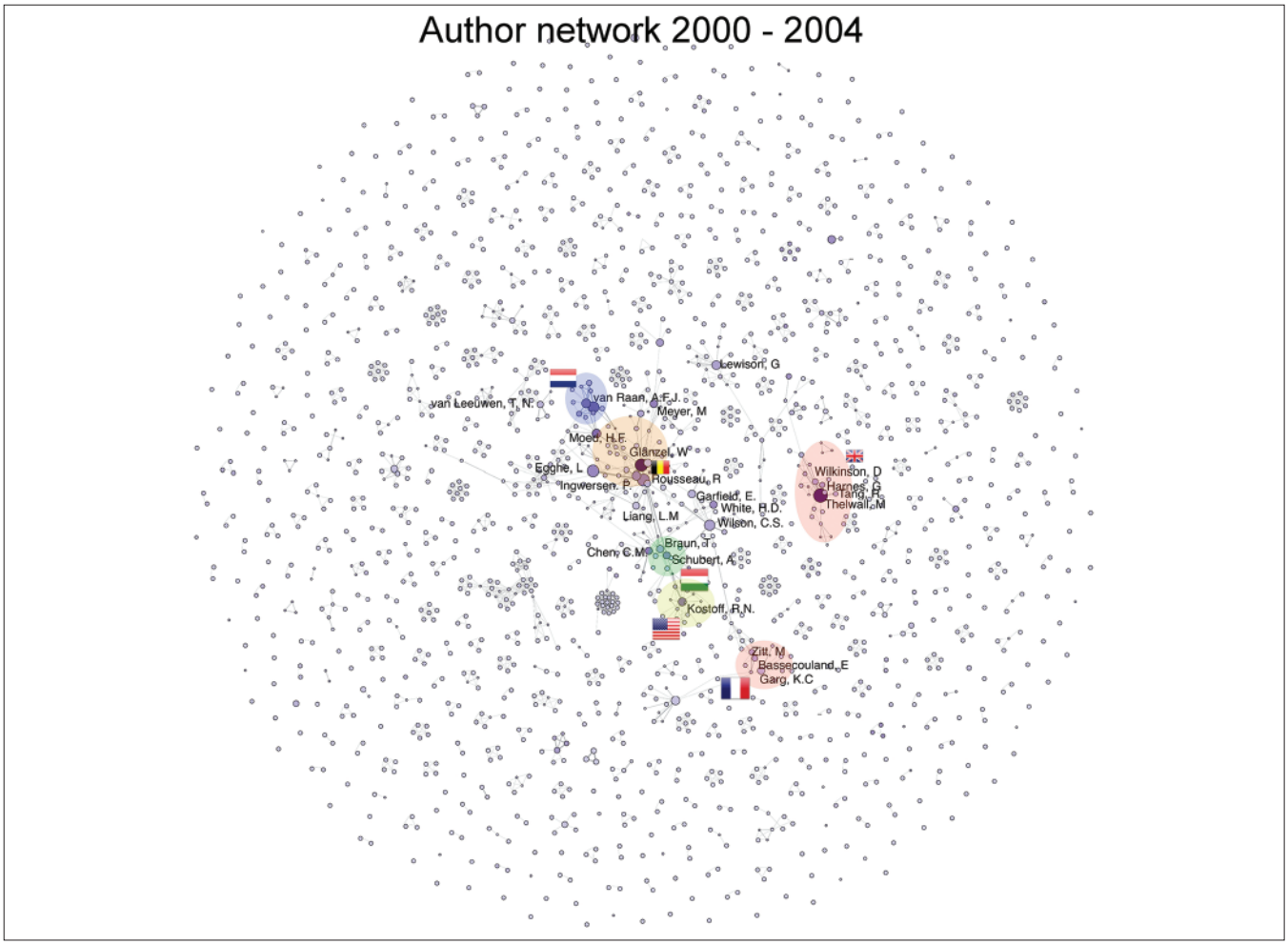

Figure 2: Author network for the period 2000-2004

of the authors had a connection to one other co-author. A further 25% of the scientists in the study collaborated with two co-authors.

In this author network, Michael L. Callaham had connections to 21 other co-authors. He is the author with the most relationships with other co-authors. Anthony F. J. van Raan and Thed N. van Leeuwen (both from the Centre for Science and Technology Studies in Leiden) collaborated most intensively, publishing 10 papers together from 2000 to 2004.

This author network is comprised of 4692 nodes, representing a community of 4692 authors. 8.3% of the nodes in the network do not contact with other nodes and can be considered to be isolated. Thus, 4302 authors had direct contact with other authors. It should be noted that in this network, 21% of the authors had a connection to only one other co-author. A further 23% collaborated with two co-authors in the period under review.

The strongest relationship in this author network is between Hans-Dieter Daniel and Lutz Bornmann. Their collaboration produced a total of 23 publications. Another author pair with an intensive relationship is Wolfgang Glänzel/Bart Thijs. The two authors published 17 papers together in the 5 year period 2005-2009.

In each network representation, the best-known authors in the community are shown in different colors. Just by looking at both networks, it is clear that the authors in the 2005-2009 periods are grouped much more closely together. This is a sign of the increased level of collaboration between them.

What Other Changes can be Observed?

Firstly, it is striking that there are significantly more nodes in the 2005-2009 network than in the 2000-2004 network. The first network (2000-2004) contains a total of 1805 authors, while the second network (2005-2009) consists of 4692 authors overall. This represents a 2.5-fold growth.

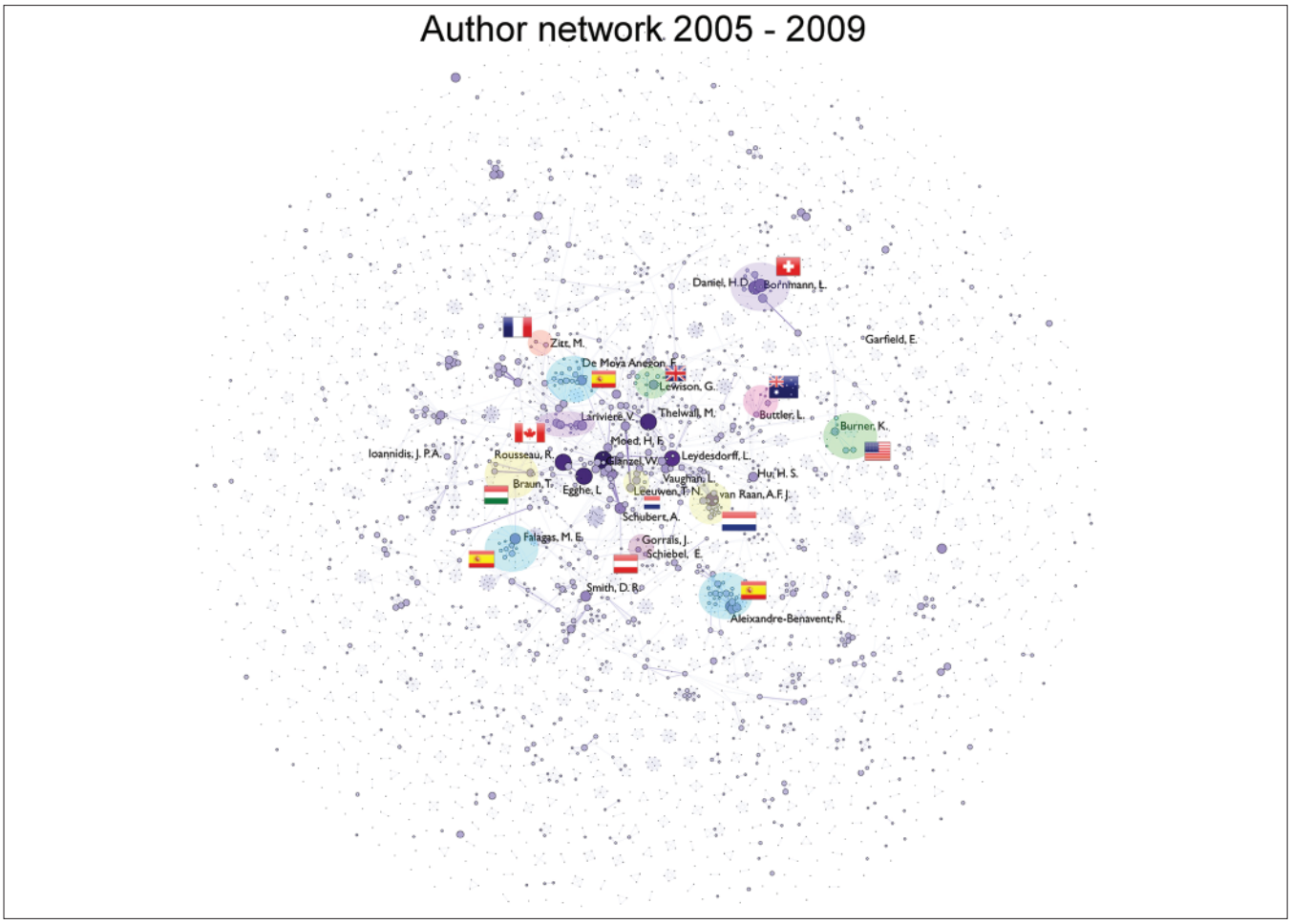

Figure 3: Network representation for the period 2005-2009

Furthermore, there was a marked increase in the number of publications in the two periods: the community produced 1755 publications in the first period and 2932 in the second. As stated earlier, interest in bibliometric questions has grown substantially within the wider scientific community, particularly in the last few years – a trend clearly reflected in these results.

The authors in the second time period worked together more intensively; this is indicated by the proportion of isolated nodes in relation to the total nodes, as well as by the increase in the total number of edges. If we compare the number of isolated nodes in the first and second networks, we see that a total of 13% of the nodes in the first network have no connection to any other node. In contrast, only 8% of the nodes in the second network have no connections. This demonstrates that the community became increasingly willing to collaborate with other authors in the years from 2005 to 2009. Although collaboration is a common trend in scientific publishing, it is very interesting to identify this trend not only in the decreasing share of single authored papers, but also in the compressed network.

Statements about Individual Authors

The network analysis identifies the central actors in the bibliometric community. It might be assumed that the next logical step would be to examine individual actors in detail and so continue the bibliometric study at the level of individuals. But perhaps we should not be so hasty.

Before doing anything of the sort, we must ensure that the individuals connected to each other can actually be compared with each other and that they have a sufficient number of publications (n greater than 30) and have published consistently over a period of time if statistical evaluations are to be effective. The individuals should also hold a similar scientific position, as this has a considerable effect on the publication options open to them. Moreover, it is important not to overlook the ages of the individuals being compared. When considering all of these factors,

it must be borne in mind that because the publication numbers are lower, there will be much more variation than for institutions. Bibliometrics can reveal the trend here:

Two bibliometricians (the names are not important), both of whom have long been active in the field and have a very central position in the network, have roughly the same number of publications. They also have almost the same number of uncited publications, around 20%. A big difference, however, is the proportion of papers that were the product of co-operations. This number is 85% for the first author, which means that almost every paper was produced in collaboration with at least one other author. 37% of the second author's publications were collaborative papers, amounting to over one third of the total. The first author's citation rate is twice as high as the second's. Does collaborating lead to success, then? It would be going too far to say that this is universally true – but it certainly did not hurt.

Thematic Developments in the Bibliometric Community

A thematic evaluation of bibliometrics publications was then conducted with the aid of the "VOSviewer" tool. This evaluation was based on the "strictly" bibliometrics publications in the bibliometrics journals Scientometrics and Journal of Informetrics. The bases for the thematic evaluation were the publications' keywords and abstracts. Results were obtained for the years 1980-1989, 2000-2004 and 2007-2011. Thematic clusters (shown in different colors) and the frequency of the basic terms were illustrated. It is clear that the term "science" is the most frequently used the term in the first period under investigation and that it occupies a relatively central position along with the term "journal". Other important terms include "quantitative study," "scientific activity," "scientist," and "publication output." These terms are today considered to be the fundamental pillars of bibliometrics.

In the second period investigated, the subject areas are more diversified. Four larger topic threads can be identified: "Scientific communication" (including "self-citation"), the more classic "scientometrics," "patentmetrics" and the fundamental indicator debate (e.g. "impact factor"). The number of terms raised comparing to the first time period, although the minimum number of occurrences was set from 3 to 8 times. This stresses the thesis of thematically diversification.

This thematic diversification continues to develop. In the final period under review, a new topic, "h-index," appears for the first time. There is also an enhanced focus on the topics of "network analysis" and "webmetrics." "Impact factor" and with it discussions of indicators, methods and data bases are still important as well as "Patent analysis." As regards policy-related topics, there is an increasing trend to use bibliometric analyses as the basis for funding decisions; relevant keywords in this context are, e.g. "university," "research impact" or "research evaluation."

The previous bubble charts containing the thematic development are based on the core publications in the bibliometric journals only. That means, in Figures 4-6 the publications in the scientific subject journals dealing with bibliometric topics are missing. To know about the difference, Figure 7 shows a thematic network for the last time period 2007-2011 based on all bibliometric publications.

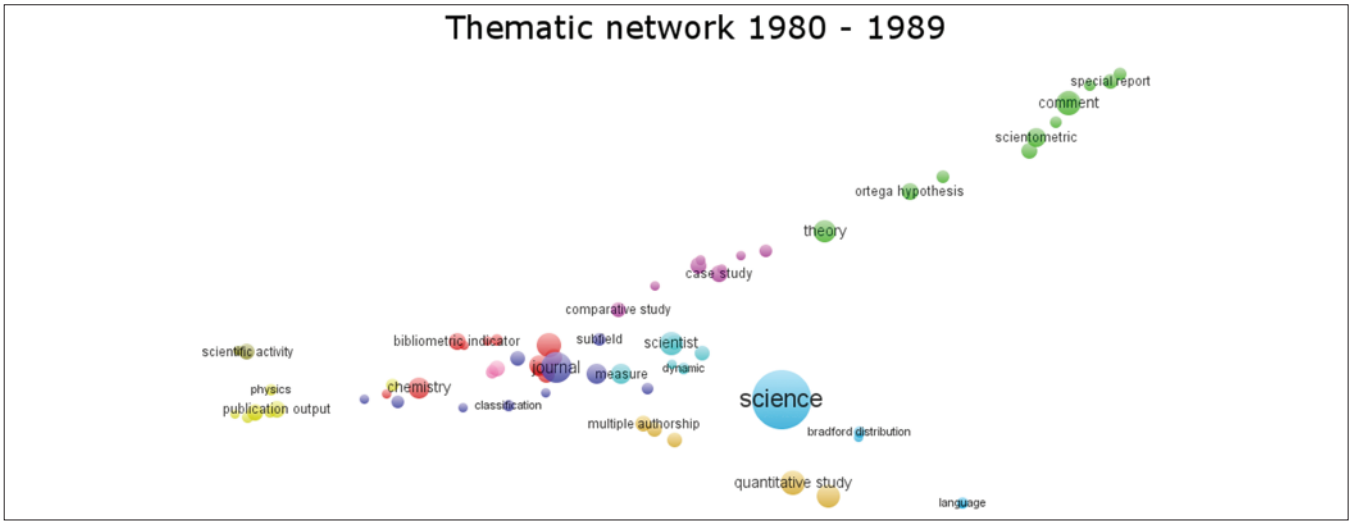

Figure 4: Thematic network 1980-1989

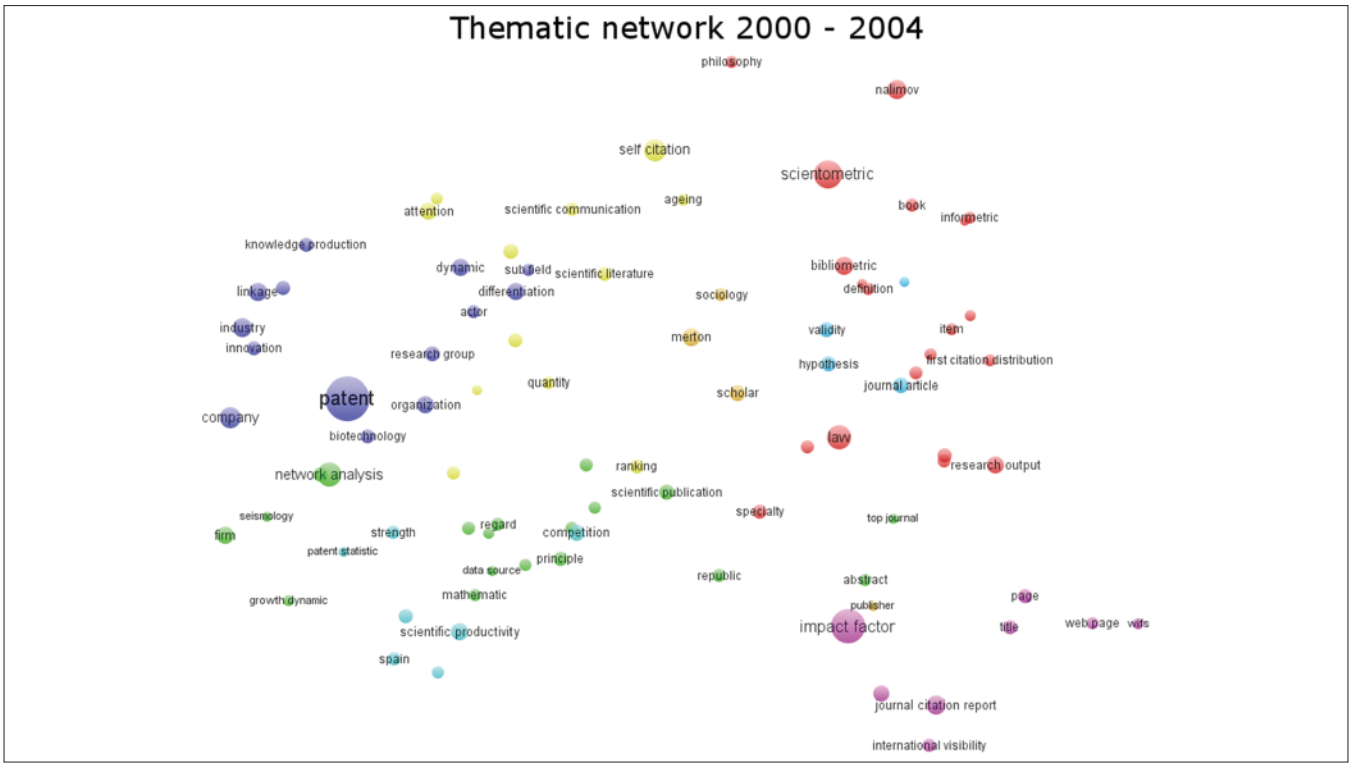

Figure 5: Thematic network 2000-2004

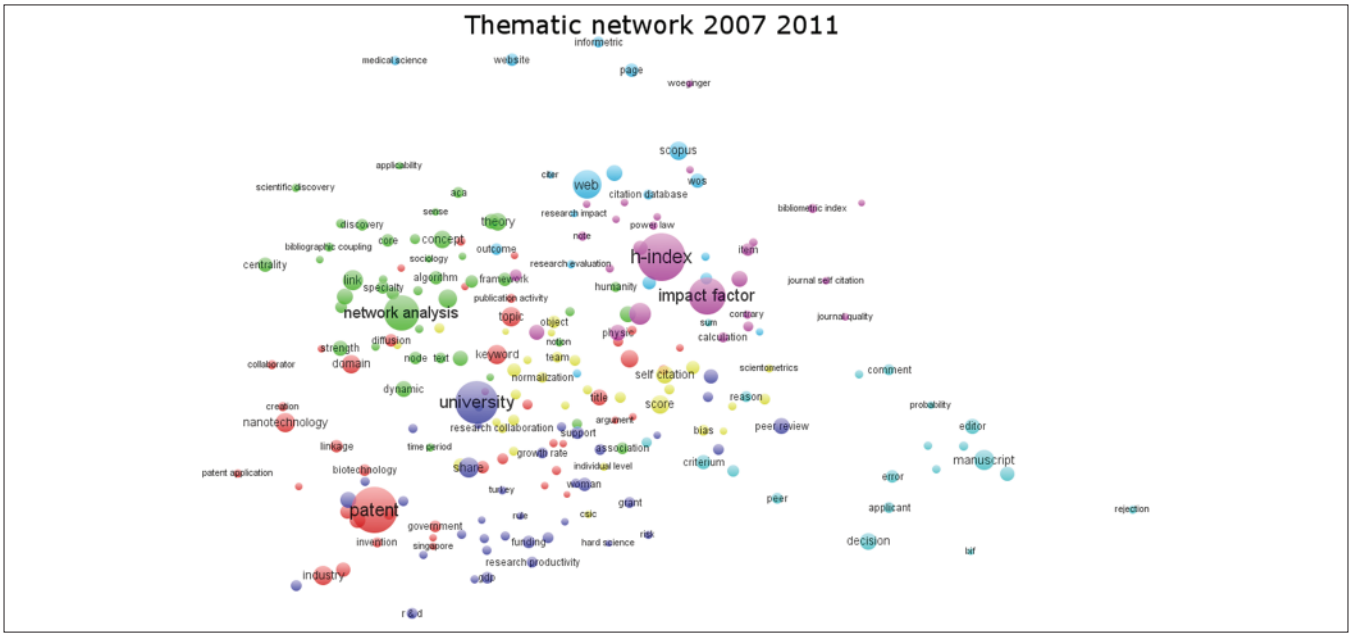

Figure 6: Thematic network 2007-2011

The network has a higher density caused by a higher number of publications. By analyzing the main terms and topics, it is getting visible, that they are less specific than the terms in Figure 6. Main terms in Figure 7 are "journal," "author," "measure," "citation," "science citation index (SCI)" etc., These terms are relevant to describe bibliometrics for the scientist of other scientific fields. That is what these publications are published for: They include specific analyses and for these, they use bibliometrics as a tool. In Figure 6 you find main terms like "impact factor," "h-index," "university" etc., describing specific bibliometric indicators or fields of interest in bibliometrics. The terms may be on a similar level, but what they represent is different. It is the aim of this publication to describe the

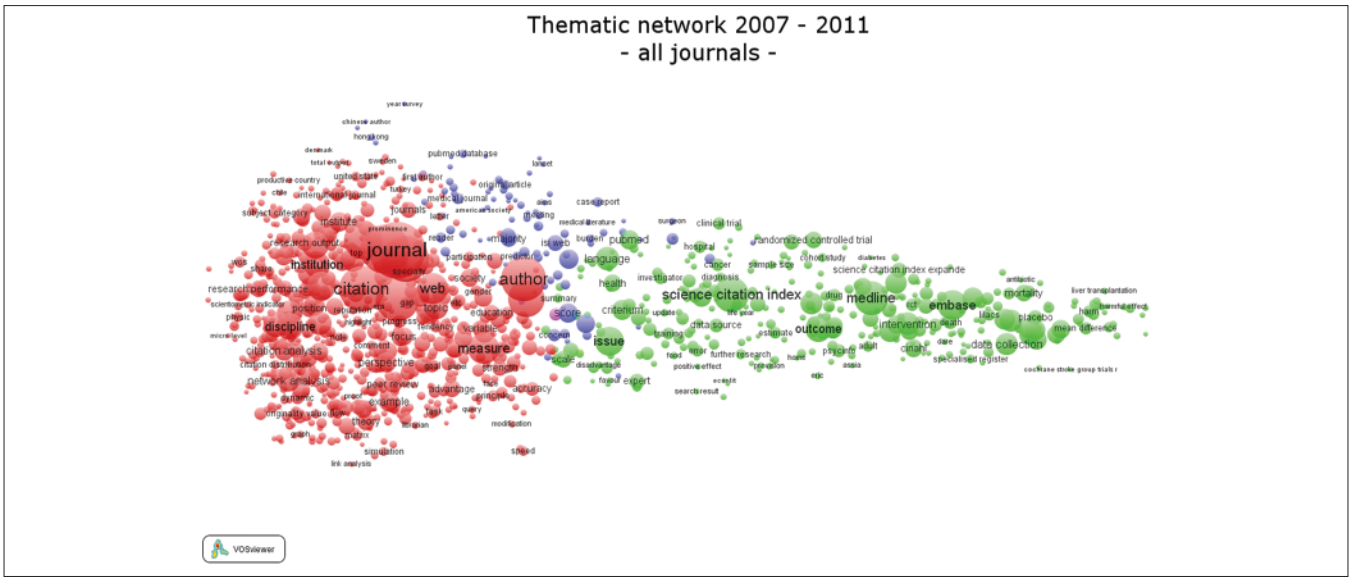

Figure 7: Thematic network 2007-2011, all journals

development in the field of bibliometrics. That is why the publications in the subject journals have not been considered for the last analysis.

SUMMARY

What can be learnt from the bibliometric analysis and the results presented here? It was shown that the development of a particular subject area can be traced using thematic search queries. Details of the development were examined more closely using the appropriate methods, providing a better understanding of the subject area in question. Key players were identified and changes in a community's network structure were documented.

It was proven that interest in the field of bibliometrics has clearly grown in recent years. This is not only reflected in the number of papers published, but also in the number of authors dealing with the topic and the degree of interaction between them. It was shown that the network of the bibliometric community as an author network has become much more tightly-knit over the past few years than it was around 2000. Bibliometrics has diversified considerably in terms of topics; what was initially a set of individual thematic aspects has developed into a scientific discipline in its own right, with an extensive thematic network.

Bibliometrics has found its place in the (natural) sciences: As shown in Figure 1, many bibliometric or scientometric publications are published in Journals that are not seen as core journals of the bibliometric community. The share of publications outside bibliometric journals is rising more intensive than those ones in the core journals or the SCI.

The author networks allow us, to study the structure of the bibliometric community, to see changes in co-publication-activity. By comparing the networks with the most active actors on conferences and in the publication landscape, the expected persons can be found in the networks. This underlines that the search strategy works and also the other results based on this strategy are valid.

The results of the thematic maps, e.g. can be compared with the program of past conferences:[9-13] This comparison shows, that the main topics, shown in clusters on the maps, are identical with the session topics of international bibliometric conferences. The development of the number and depth of these topics allows the conclusion, that the diversity of topics is actual rising.

More general: A rising publication output, more co-publication activity and a rising thematically diversification allows the statement, that the whole community is rising. This might be affected by a rising interest in bibliometric indicators in the context of metric based evaluation.

REFERENCES

1. Wilhelm J. Bibliometric analysis of the scientific community 'bibliometrics and scientometrics'. Cologne/Juelich: Diploma Thesis at Cologne University of Applied Sciences, Faculty of Information Studies and Communication Sciences; 2011.
2. Tunger D. Spieglein, spieglein an der Wand.-Die bibliometrische community im spiegel ihrer eigenen methodik. B.I.T. Online 2012;15:127-33.

3. Bar-Ilan J. Informetrics at the beginning of the 21st century-A review. J Informetr 2008;2:1-52.
4. Hu SW, Chen L. Test analysis on prestressed concrete composite beams with steel boxes subjected to torsion and combined flexure and torsion. In: Science China-Technological Sciences. Vol. 55, no. 12; 2012. pp. 3302-10.
5. Wissenschaftsrat. German Council of Science and Humanities, 2013. The Excellence Initiative. Available from: http://www.wissenschaftsrat.de/en/fields-of-activity/excellence-initiative.html. [Last retrieved on 2013 Apr 05].
6. Kling-Mathey C. More transparency provided by bench marking systems in research, 2013. Available from: http://www.wissenschaftsrat.de/download/archiv/pm_3304.pdf. [Last retrieved on 2013 Apr 05].
7. Hasitzka K, Gumpenberger C, Gorraiz J, Wieland M. The "sound of bibliometrics": An analysis of its level of awareness and acceptance in Austria. In: Proceedings of the ISSI 2011 Conference: 13th International Conference of the International Society for Scientometrics and Informetrics, Durban; 2011.
8. Brockhaus. Brockhaus encyclopedia (30 volumes). 21st ed. Leipzig: 2006. p. 13.
9. ISSI 2011. Proceedings of the ISSI 2011 Conference: 13th International Conference of the International Society for Scientometrics and Informetrics, Durban; 2011.
10. S and TI 2006. Book of Abstracts, 9th International Conference on Science and Technology Indicators, Leuven; 2006.
11. S and TI 2008. Excellence and Emergence. A New Challenge for the Combination of Quantitative and Qualitative Approaches; Book of Abstracts, 10th International Conference on Science and Technology Indicators, Vienna; 2008.
12. S and TI 2010. Science and Technology Indicators "Creating Value for Users": Book of Abstracts, 11th International Conference on Science and Technology Indicators, Leiden; 2010.
13. S and TI 2012. Proceedings of STI 2012: 17th International Conference on Science and Technology Indicators, Montreal; 2012.

How to cite this article: Tunger D, Wilhelm J. The bibliometric community as reflected by its own methodology. J Sci Res 2013;2:92-101.

Source of Support: Nil, **Conflict of Interest:** None declared