Local, Networked or External?

Inclusion of Regional Journals in WoS and its Effect

Valeria Aman

iFQ – Institute for Research Information and Quality Assurance
Schützenstraße 6a, Berlin, Germany
aman@forschungsinfo.de

Abstract

Between 2007 and 2009 Thomson Reuters started to accept journals beyond the international high-impact literature, and included 1,600 regional journals in Web of Science (WoS). The study sets out to analyze the effect on countries that in terms of absolute article numbers profited most from the inclusion process. The comparison of a country's newly included journals with those in existence in WoS before 2007, and the distinction of articles according to their origin (local, networked or external) allows to detect various effects. The reception of these two sets of journal articles is likewise distinguished by the origin of the recipients. Results show that some countries suffered a loss in international visibility and are now represented in WoS more locally oriented than before the altered journal selection process. Other countries experienced a growth in external citations as well as publications from international authors, being everything but regional.

Keywords: Regional journals, Coverage, Productivity, Web of Science

In: F. Pehar/C. Schlögl/C. Wolff (Eds.). Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th–21st May 2015. Glückstadt: Verlag Werner Hülsbusch, pp. 358–369.

1 Introduction

What Gibbs (1995) laconically calls "Lost science in the Third World" refers to the locally published journal articles that are virtually invisible for the international scientific community. The low visibility derives from the fact that the majority of their publications are disseminated through domestic scholarly journals or other publication means with a geographically and linguistically limited range of dissemination (Arunachalam, 2004). Since the inception of the *Science Citation Index* (SCI) by Eugene Garfield more than half a century ago, the goal of the journal selection process was to identify journals that build the core literature of the sciences. These were mostly English-language journals meeting the need of a broad research community. In 1973, journals from the Social Sciences were added to SCI, followed by Arts & Humanities in 1978. Journals included in Web of Science were regarded as the ones with the highest citation impact of all journals published. Over a period of more than 40 years, the set of international highly-cited journals was relatively constant. Between 2007 and 2009 Thomson Reuters changed its journal policy and included around 1,600 so called "regional journals". These journals present research from a regional perspective, while their international impact is secondary. The aim of this paper is to analyze the effect on those countries that profited most from the inclusion process that started in 2007. The distribution of locally publishing, networked and external authors and how it changed with the inclusion of regional journals will be addressed. Likewise, the distribution of the origin of incoming citations, just as the citation counts, indicate the effect of the journal inclusion on the orientation of the countries, thus whether they are interconnected, locally or externally oriented.

2 Background

The community that uses Web of Science (WoS) in their daily work has expanded over the years. Whereas in the past universities and research institutions in the US and Western Europe formed the core, nowadays major universities and research facilities all over the world show broad interest in Web of Science. With the expanse of the circle of users, Thomson Reuters adapted

quickly a strategy to serve the need of the newcomers. They dissociated from the original journal selection process and moved into "the realm of the regional journal literature" (Testa, 2009). According to Testa (2009) regional journals are those published outside the US or UK. They focus on topics of regional interest or present them from a regional perspective. The authors are from the region rather than from international research groups. The importance of regional journals is measured by the content and not simply by their citation impact. Requirements for the inclusion of regional journals in WoS are the publication on time, a formal peer-review and bibliographic information in English (with the exception of Arts & Humanities). Testa calls it a "monumental task", which was initiated by the Editorial Development Department of Thomson Reuters in 2006. Subject editors compiled a list of more than 10,000 journals from all areas of scholarship. Within 12 months they selected 700 journals for coverage in WoS. Another 900 were added by the end of 2009 (Testa, 2011). The number of regional journals covered by WoS will significantly influence any analysis of the country that is affected by the inclusion of regional journals. Basu (2010) examined the effects on a country's scientific productivity, if the SCI-indexed journals covered keep changing with time. He concluded that an apparent increase of a country's productivity may simply result from the inclusion of new journals in the database. There are several previous studies focusing on regional countries, their visibility and the influence of publication language on impact. Van Leeuwen et al. (2001) discussed the consequences of the relative language coverage of journals in the SCI, when comparing national research performance on a global scale. They concluded that the citation rate of a paper depends primarily on the language it was published, with an outstanding advantage of English-language papers. Tijssen, Mouton, van Leeuwen & Boshoff (2006) analyzed South African journals, including international journals indexed in bibliographic databases and local journals. The analysis of more than 200 South African journals in regard to output and citation impact showed that the majority of local journals are invisible for the global scientific community. Collazo-Reyes et al. (2008) studied the publication and citation patterns and growth dynamics of Latin American and Caribbean journals covered in WoS from 1995–2003. They found little inter-citation among local papers, while the highest cited papers by extra-regional authors were those published in English. Lermarchand (2010) determined the journal coverage of 12 Iberoamerican and Caribbean countries and analyzed the evolution of the cooperation networks among them in the 1973–2006 period. What

most of the previous studies have in common is that they take the presence of regional journals in WoS for granted, showing that regional journals are disadvantaged in terms of visibility and impact due to their publication language or outreach. Kosanović & Šipka (2013), and Collazo-Reyes (2014) studied critically the effect of Thomson Reuters change in journal policy. Their analyses of South East European journals and Latin American and Caribbean journals respectively, show that the recently covered journals in WoS are low in citation impact. They conclude that overrepresentation in WoS can neither be the long-term interest of the country affected, nor of Thomson Reuters.

3 Data and methods

The study was conducted by means of an in-house-database version of Thomson Reuters' Web of Science. The journal analyses include the *Science Citation Index* (SCI), *Social Sciences Citation Index* (SSCI) and the *Arts & Humanities Citation Index* (A&HCI). To prevent any bias among countries, the analyses are restricted to journals and articles only. In a first step, journals had to be assigned to the country they are published, and the year of their inclusion into WoS had to be identified. Based on these information, for each of the countries of interest a distinction between "new" journals, those that were by definition included after 2007 and "old" journals, those that have been included prior to 2007 in WoS, was possible. For both of these sets of journals (new and old) articles published between 2007 and 2013 were classified into one of the following three groups, according to the geopolitical location of the author's institution and the journal:

- (L) *Local articles*, where all of the authors affiliated to an institution are situated in the country where the publishing journal is edited.
- (N) *Networked articles,* where at least one of the authors is from the local country, the target journal is edited in and one author whose affiliation belongs to an institution situated in a country other than that where the publishing journal is edited.
- (E) *External articles*, where all of the publishing authors work at institutions located outside the country where the respective journal is edited.

These three sets are disjoint and allow comparisons with respect to the following indicators: Total numbers and percentages of articles published be-

tween 2007 and 2013 in a country, distinguished by the three groups of articles (L, N, E) and the set of journals (old vs. new). The same distinction has been applied for the citation analyses. Local citations are those from the country, where the article is published, networked citations originate from articles where a local author co-operated with an author from abroad. External citations are accumulated by publications where all of the authors are located in a country other than that where the cited article was published.

4 Results and discussion

The countries of interest have been chosen on the basis of the highest number of incoming journal articles, resulting from Thomson Reuters' inclusion of regional journals. Table 1 lists the 20 countries in a descendant order, according to their absolute number of articles published between 2007 and 2013 in newly included journals. On the right part of the table we can see the number of articles that were published in journals that have been already covered in WoS before 2007. A column indicating the number of "new" and "old" journals is provided next to each of the sets of articles.

We can infer from the table that the lion's share of articles in newly included journals are of US-origin, followed by Great Britain. Brazil is on rank 3, followed by Poland and Germany. According to Testa (2011) UK, US, Germany and the Netherlands faced the greatest increase in journal coverage between 2005 and 2010. We can see that Turkey's number of journals represented in WoS has grown by a factor of ten. The ratio of articles and journals results in the indicator "Journal Packing Density" (JPD), which is defined as the average number of papers in journals for a given country (see Basu, 2010). The example of Pakistan shows that the JPD is extremely high for the two journals included prior to 2007. On average, the ratio results in 280 articles per journal per year. In contrast, Pakistan's JPD for the recently included journals is much lower. To illustrate the countries that profited most from the journal inclusion, the following figure provides a bar chart. It visualizes the percentages of articles published between 2007 and 2013 in journals accepted after 2007, in relation to articles in journals that have been covered in WoS before 2007.

Table 1. Overview of the effect of the journal inclusion process on article numbers and journal numbers published between 2007 and 2013 in SCI-E, SSCI, and A&HCI for countries with the highest growth in absolute articles numbers.

Country	No. of articles in journals incl. after 2007	No. of journals incl. after 2007	No. of articles in journals incl. prior to 2007	No. of journals incl. prior to 2007
USA	256,814	766	3,287,377	4,251
GB	146,719	622	1,422,221	2,988
Brazil	44,290	98	30,414	35
Poland	31,349	84	35,581	68
Germany	31,264	141	275,294	510
China	31,037	67	94,650	68
Netherlands	30,968	141	730,690	701
South Korea	27,365	63	41,357	39
India	23,195	63	43,056	52
Turkey	22,244	63	2,620	6
Romania	21,046	44	8,442	11
Switzerland	18,776	73	171,143	184
Italy	17,310	74	36,972	126
Japan	14,995	49	101,822	145
Spain	13,022	79	21,106	65
Pakistan	12,278	12	3,916	2
Iran	11,139	44	1,371	8
France	9,545	52	95,897	208
Australia	8,437	33	13,762	58
Croatia	8,317	40	4,871	17

Depending on the number of journals that were already in presence in WoS before 2007, the percentages vary immensely among countries. Evidently, Turkey, Iran and Pakistan are the countries who owe their today's presence in the database to the inclusion of regional journals. More than 50% of all articles published in Romania, Croatia and Brazil between 2007 and 2013 arise from the recently included journals. Although the number of journals from USA and GB has grown in absolute terms, effects on the proportion of all articles published in these countries are rather weak.

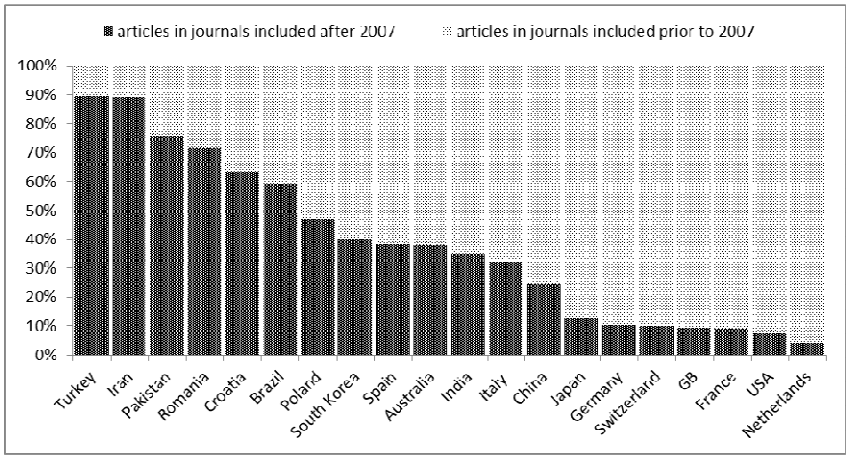

Figure 1. Countries that profited most from the journal inclusion process and their relation of article numbers published between 2007 and 2013, distinguished by articles in journals included after 2007, and those included prior to 2007.

To show the effects of the implemented journal selection process, it is now of interest to distinguish between "old" and "new" journals and the origin of the published article. Therefore, the following figure presents a bar chart for the 20 countries of interest. The left bar represents the "old" journals (included before 2007), whereas the right bar represents the set of newly included journals. Each of the bar consists of three groups of articles, according to their origin (L, N, E). The countries in figure 2 are not in an arbitrary order, instead they are arranged in accordance to the highest increase of the share of locally published articles. Starting with Pakistan, we can see that in the 2 journals already included in WoS before 2007, 70% of the articles published between 2007 and 2013, are exclusively by local authors. The right bar for Pakistan with the newly absorbed ("regional") journals shows that these journals are not as regional as expected. Only 36% of articles are local, whereas 60% of articles are published exclusively by authors from abroad. Thus, the regional journals included are evidently of relevance for the international community, since many external authors place their articles in these journals.

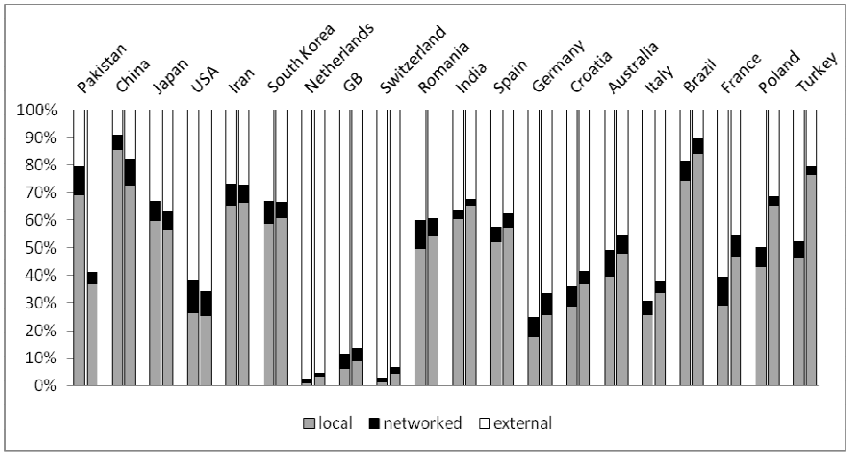

Figure 2. Distribution of the origin (L, N, E) of the articles published between 2007 and 2013. The left bar for each country represents articles in journals included prior to 2007, whereas the right bar represents the articles in journals included after 2007.

The same effect becomes visible for China. The journals included prior to 2007 show the highest rate of locally published articles (85%) in figure 2. The 67 Chinese journals included after 2007, show a rate of 72% of local articles. The recently included journals from Japan and USA also show a lower rate of locally published articles. For Iran and South Korea a small increase of locally published articles is visible, with around two-thirds of their articles beeing local. The Netherlands, GB and Switzerland show different characteristics. Together with UK and Germany they are centers of international scholary publishing and experienced the greatest increase in journal coverage in WoS on a routine yearly basis (Testa, 2011). The journals published in these countries have a high influx of external publications. According to table 1 Romania, India and Spain faced an increase in regional journals. The bars in figure 2 show that the journals included after 2007 consist of more than 50% of local articles. A high number of regional journals was also included in Germany and Croatia. Just as China, Brazil is outstanding as a counry with a high share of local articles – in absolute article numbers, and in relation to all articles published in the country. Different from China, the bars show that the newly included journals from Brazil are to a higher share local than those covered in WoS before 2007. The highest increase in the share of local articles becomes evident for Turkey. Whereas articles in "old" journals are to a rate of 50% external, the inclusion of regional journals let

this share sink to 20%. The number of networked papers has diminished as well. In terms of article counts, we can observe different effects on counries that were affected by the inclusion of 1,600 regional journals. It is now of interest to study the reception of articles published in these countries. A citation window of three years is applied. Thus, only those citations are counted, which were received in the year of article publication and the two following years. Figure 3 can be read as figure 2, with the difference that each bar represents the share of the origin (L, N, E) of the incoming citations. Whereas the two journals from Pakistan, included prior to 2007, received 35% of their citations from external authors, the newly included journals owe 68% of their citations to this group. Pakistan marks the highest increase in external citations from the countries displayed in figure 3. The results show that the recently absorbed journals are not as regional as stated in Testa's documentation (2011). Iran, Romania, India and China too, receive to a higher degree external citations in comparison to the journals that have been already integrated in WoS before 2007.

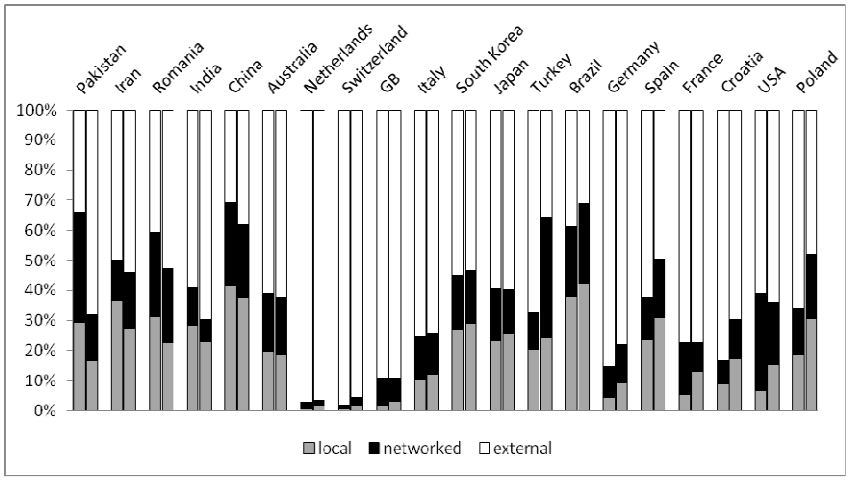

Figure 3. Distribution of the origin of the citations to articles published between 2007 and 2013 in the country as indicated. The left bar for each country represents articles in journals included prior to 2007, whereas the right bar represents articles in journals included after 2007.

The central countries of publishing, Netherlands, Switzerland and GB show high external citation rates. South Korea and Japan show with ca. 30% a stability in their share of national self-citations. Turkish journals that are

recently covered in WoS, are to a high degree cited by networked publications. One-third of all incoming citations results from Turkish authors, who have collaborated with authors from other countries than Turkey. The countries on the right of figure 3 show an increase in national self-citations in journals covered in WoS from 2007 on. These countries are accompanied by lower external citation rates. The increase of local citations in US should be interpreted with caution, because a size effect is at stake. Since the US-output is relatively large (table 1), US articles constitute a large citation "target" (see Moed 2005: 293). Having presented the origin of citations, we can now turn to the citation impact. The following figure indicates the citation rates to the group of articles distinguished (L, N, E). Pakistan on the left shows that more than 50% of all citations received in the 2007–2013 period point to the external publications in Pakistan's journals. We can conclude that the recently included journals from Pakistan are not regional, since the majority of cited articles are those being external.

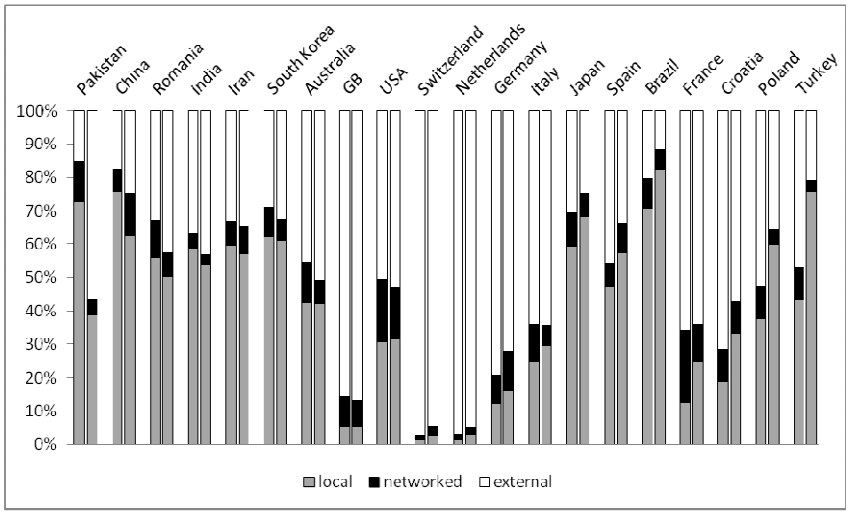

Figure 4. Distribution of the origin of a country's articles that receive most of the citations between 2007 and 2013. The left bar for each country represents articles in journals included prior to 2007, whereas the right bar represents articles in journals included after 2007.

China, Romania, India, Iran and South Korea show that with the newly included journals a higher share of the external articles accounts for citations, than in the journals included prior to 2007. At the same time a lower share of

local articles accounts for incoming citations. This means, that these countries became less regional despite the inclusion of regional journals. On the opposite, we can see on the right the countries who faced an increase of citations to locally published articles. Local articles account for more than 80% of all citations to articles from Brazil. Articles of French origin and published by local authors account for higher citation rates at the expense of networked articles. Three quarter of all citations to Turkish articles in newly included journals, target to local articles. This is in accordance with the high share of local articles in Turkey (fig. 2).

5 Conclusion

Thomson Reuters extended the journal coverage by focusing on regional journals all over the world. The goal was to enrich the collection of international journals with those whose focus is on specific regional themes. Countries such as Turkey, Iran, Pakistan and Romania grew from very narrow representation to significant coverage. By the end of 2010, 87 countries were represented in Web of Science on the journal level, 14 of which for the first time (Testa, 2011). As Michels and Schmoch (2012) state, this growth in the articles number in WoS should not be interpreted as an increased scientific activity. It is primarily an artefact of the inclusion of regional journals since 2007, and the tolerance to accept journals regardless of their impact. The distinction by local, networked and external articles shows that some countries are now represented as more locally than before the journal expansion. The results of the citation impact indicate nevertheless that regional studies are referenced from external publications and do not go unnoticed in the international community of researchers. To conclude, the inclusion of regional journals since 2007 had different effects on the visibility and the reception of these journal articles for each of the countries presented.

References

Arunachalam, S. (2004). Science on the periphery: bridging the information divide. In: H. Moed et al. (Eds.). *Handbook of Quantitative Science and Technology Research.* Dordrecht/Boston/London: Kluwer Academic Publishers.

Basu, A. (2010). Does a country's scientific 'productivity' depend critically on the number of country journals indexed? *Scientometrics*, 82 (3), 507–516.

Collazo-Reyes, F. (2014). Growth of the number of indexed journals of Latin America and the Caribbean: the effect on the impact of each country. *Scientometrics*, 98, 197–209.

Collazo-Reyes, F. et al. (2008). Publication and citation patterns of Latin American & Caribbean journals in the SCI and SSCI from 1995 to 2004. *Scientometrics*, 75 (1), 145–161.

Gibbs, W. W. (1995). Lost science in the Third World. *Scientific American*, 273, 76–83.

Glänzel, W., Leta, J. & Thijs, B. (2006). Science in Brazil. Part 1: A macro-level comparative study. *Scientometrics*, 67 (1), 67–86.

Kosanović, B., & Šipka, P. (2013). Output in WoS vs. Representation in JCR of SEE Nations: Does Mother Thomson Cherish All Her Children Equally? *Centre for Evaluation in Education and Science*, 125–137.

Lemarchand, G. A. (2010). The long-term dynamics of co-authorship scientific networks, Iberoamerican Countries (1973–2006). http://arxiv.org/abs/1001.2837.

Michels, C. & Schmoch, U. (2012). The growth of science and data bases coverage. *Scientometrics*, 93, 831–846.

Moed, H. F. (2005). *Citation analysis in research evaluation*. Springer, Berlin.

Testa, J. (2008). A regional perspective on science. *KnowledgeLink newsletter*, July 2008.

Testa, J. (2009). Regional content expansion in Web of Science®: opening borders to exploration. GlobalHigherEd.

Testa, J. (2011). The globalization of Web of Science 2005–2010. http://wokinfo.com/media/pdf/globalwos-essay.pdf.

Tijssen, R. J. W., Mouton, J., van Leeuwen, T. N. & Boshoff, N. (2006). How relevant are local scholarly journals in global science? A case study of South Africa. *Research Evaluation*, 15 (3), 163–174.

Van Leeuwen, T. N., Moed, H., Tijssen, R. J. W., Visser, M.S., Van Raan. A. F. J. (2001). Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics*, 51, 335–346.

Bibliometric Analysis of the Field of Folksonomy Research

Tomislav Ivanjko, Sonja Špiranec

University of Zagreb, Faculty of Humanities and Social Sciences, Department of Information and Communication Sciences
Ivana Lučića 3, 10000 Zagreb, Croatia
{tivanjko, sspiran}@ffzg.hr

Abstract

The area of researching folksonomies is still in development, so theoretical perspective and research methods are still being defined. This study conducts a webometric and bibliometric analysis of the folksonomy research in the Library and Information science (LIS) field by collecting data from Web of Science (WOS), SCOPUS and Google Scholar in July 2014. It utilizes a total of 346 papers with 2660 citations from WOS and 1581 papers with 8848 citations from SCOPUS. In addition, Google Scholar database search was also included for providing a wider coverage of works published in conference proceedings, books and to include a wider journal base. Based on these results, research identifies most influential papers and authors across all three databases.

Keywords: Folksonomies, Social tagging, Collaborative tagging, Bibliometric analysis, Webometrics

1 Introduction

With the rise of Web 2.0, a new wave of user participation in creating and describing online resources instigated a new approach in knowledge repre-

In: F. Pehar/C. Schlögl/C. Wolff (Eds.). Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th–21st May 2015. Glückstadt: Verlag Werner Hülsbusch, pp. 370–379.

sentation – folksonomy. Folksonomy relies on the process of collaborative tagging, where many users add metadata in the form of keywords to shared content (Golder & Hubermann, 2006; Mathes, 2004). The totality of these user-generated keywords (tags), gathered around any different platform or resource creates a folksonomy (Peters, 2009). Within this framework, different approaches are possible, where only one of the elements can be analyzed (for example, analyzing the linguistic characteristics of a chosen tag corpus) or, more often, the relationship between two elements is investigated (such as the relationship between tags and resources, identifying possible differences in tagging different types of resources). Since the coining of the term the new research topic emerged in the field of Information Science dealing with the structure, use and application of folksonomies in the field of knowledge organization and representation and information retrieval. This paper aims to explore the body of literature currently present on the topic of folksonomies inside the field of Information Science by using webometric tools and methods to identify key concepts and bibliometric methods to identify key authors and papers in the field.

2 Identifying key concepts

Since the coining of the term folksonomy (Vander Wal, 2004) different competing terms emerged to describe the field of research. Peters (2009) provided an exhaustive literature review regarding the terminology use and listed the most prominent ones "ethnoclassification", "communal categorization", "democratic indexing", "mob indexing", "social classification system", "social indexing", "user-generated metadata", "collaborative tagging", "social tagging" and "folksonomy". Following the methodology from our previous work (Lasić-Lazić, Špiranec & Ivanjko, 2014) where the focus was on the content analysis of the field, a webometric analysis of the competing terms was conducted using the tool Webometric Analyst 2.0[1]. Following the method from Thelwall (2013) a cross-domain web impact assessment via web mentions was conducted in July 2014 including the most mentioned terms. Web impact assessment (WIA) is the evaluation of the "web impact" of documents or ideas by counting how often they are mentioned online. The

1 http://lexiurl.wlv.ac.uk

underpinning idea is that, other factors being equal, documents or ideas having more impact are likely to be mentioned online. The tool returns a number of different metrics, the most reliable being the number of domains due to the possibility that text or links are copied across multiple pages within a web site (Thelwall, 2013). The results are presented in table 1.

Table 1. Cross-domain web mentions of competing terms

Term	Result
folksonomy	575
user-generated metadata	473
social tagging	453
collaborative tagging	298
social classification system	240
social indexing	188
ethnoclassification	166
democratic indexing	51
mob indexing	50
communal categorization	40

As we can see from the analysis the most widely used term is “folksonomy” with 575 cross-domain mentions, followed closely by “user-generated metadata” and “social tagging”. By examining the results in detail it became obvious that the terms “social classification system” and “ethnoclassification” yielded such high results not because they relate to a concept found in the literature but its origin derives from sociology where they denote a completely unrelated notions so it was clear they should be excluded from any literature search as it would generate a lot of false results not related to our field of interest.

3 Identifying key authors and papers

The results of the webometric analysis gave us a starting point for constructing a Boolean query (folksonom* OR "social indexing" OR "social tagging" OR "user-generated metadata" OR "collaborative tagging") in order to in-

clude all the relevant concepts when searching the databases. Three different sources included in the search were Web of Science[2], SCOPUS[3] and Google Scholar[4]. In addition to searching the standard bibliographic databases in the field, Google Scholar was also included in order to provide a better insight into publications outside high impact journals, such as works published in conference proceedings, books and to include a wider journal base as suggested by Harzing (2008). Some studies have shown that although Google Scholar ranking algorithm weighs heavily on articles' citation counts (Beel & Gipp, 2007), top ranked articles are not necessarily those with the highest citation count so the search of Google Scholar database was conducted using software Publish or Perish 4 (Harzing, 2007) to identify relevant papers.

Table 2. Summarized data on sources included in the analysis

Database	No. of papers	No. of citations	h-index
WOS	346	2660	21
SCOPUS	1581	8848	41
Google Scholar	1000+	31234	80

As we can see, fewest papers on the topic are published through WOS, with the lowest h-index. As for Google Scholar, the software Publish or Perish 4 is limited to processing the first 1000 results so the total number of articles could not be calculated but instead first 1000 results were analyzed. These results show that there is a notable interest in the field of research with an already respectable number of published articles in high impact journals. To get some insight into the most influential papers and authors in the field 20 most cited articles from WOS and SCOPUS were compared and a total of 7 articles were found cited both in top 20 WOS and SCOPUS. If we take Google Scholar into account, then only 3 papers are present in top 20 for all 3 databases (*Usage patterns of collaborative tagging systems* (2158); *Information retrieval in folksonomies: Search and ranking* (725); *Ontologies are us: A unified model of social networks and semantics* (619)). It is clear from the results that the paper published by Golder and Huberman (2006) (*Usage patterns of collaborative tagging systems*) is by far the most cited paper in the field, having attracted most citations across all three databases. Also it

2 http://wokinfo.com/products_tools/multidisciplinary/webofscience/

3 http://www.info.sciverse.com/scopus

4 http://scholar.google.com

should be noted that the paper from Mika (2008) (*Ontologies are us*) is present in both WOS and SCOPUS in two slightly different versions (a conference paper from 2005 was rewritten as a journal article in 2007) but both papers share the same basic concepts and ideas, so from the intellectual point of view they should be regarded as one article. If we take that into account then the citation number for that paper raises significantly making it the second most cited article across analyzed databases. Although there is a fairly large amount of papers published, it is obvious from the results that the field is very heterogenic, with only several papers being present as top cited in all the databases. Since both WOS and SCOPUS provide access based on subscription fees and Google Scholar is free to access, researchers in the field trying to get insight into the topic could start with very different papers based on their institution financial power with only three articles being present in the top 20 most cited articles across all three databases. When we look at the categories from which the journals with the most citations stem, there are two main fields that are interested in the topic of folksonomies: Computers Science and Library and Information Science. Articles written from a Computer Science perspective are concerned mostly with using folksonomies in exploring the ways in which user tags can improve the effectiveness of different systems and information retrieval (for example, extracting meaningful data for creating partial ontologies as a basis for the Semantic Web). On the other hand, Library and Information Science field is more interested in researching user motivations for tagging (to enable better communication with its patrons) and the potential of user tags in enhancing resource description (to complement standard KOS methods). A more detailed content analysis of the approaches in the field can be found in the work published by Lasić-Lazić, Špiranec and Ivanjko (2014).

4 Co-citation analysis

One of the basic methods of bibliometrics is counting co-citations, a method for identifying influential authors and displays their interrelationships from the citation record (White & McCain, 2009). In order to provide that kind of insight in the field of folksonomy research, a co-citation analysis of the papers from both WOS and SCOPUS was carried out. From the SCOPUS data-

base a total of 1581 articles with 8848 citations were ana-lyzed. The analysis was carried out using the software Bibexcel[5] a bibliometric toolbox for most types of bibliometric analysis (Persson, Danell & Wiborg Schneider, 2009). Bibexcel was used for processing the data, while Pajek[6] was used for visualization of the data as used in Batagelj and Mrvar (2003). Figure 1 shows the co-citation graph from SCOPUS records where the size of vertices indicates the number of citations while the thickness of lines indicates the number of co-citations between authors. To reduce the complexity of the visualization, figure 1 shows only authors that have more than 20 co-citations.

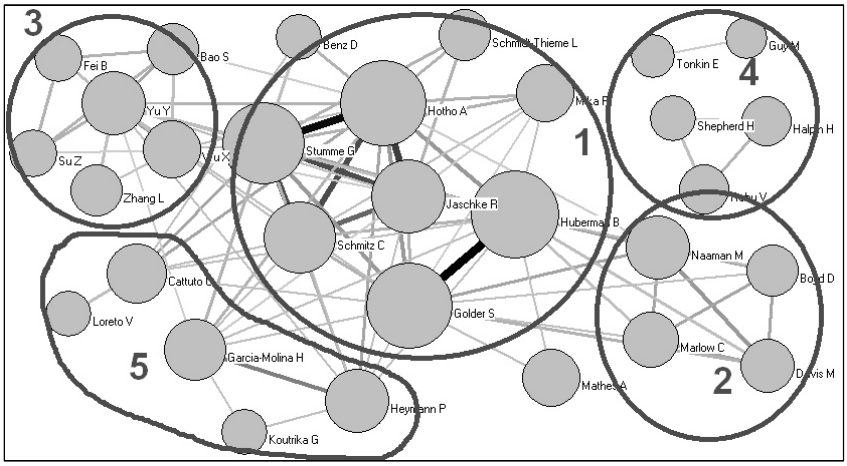

Figure 1. Co-citation graph based on 8848 citations from SCOPUS database

As we can see from the graph, there are roughly 5 main clusters of authors that are interconnected with high number of co-citations. Again, in the centre of the graph (1) there are authors of the two most cited papers (*Golder and Hubermann; Hotho*, *Jäschke, Schmitz, and Stumme*) that have the strongest co-citation links. Then there are three clusters of authors that are on the outskirts of that centre cluster (2, 3, and 5) that have strong mutual connections and are also strongly connected to the central cluster. And finally there are authors that have a high number of citations but are not that strongly co-referenced by other authors (4). This analysis revealed some new influential authors and papers in the field such as Marlow, Naaman, Boyd and Davis

5 http://www8.umu.se/inforsk/Bibexcel/

6 http://pajek.imfm.si/doku.php

(*HT06, tagging paper, taxonomy, Flickr, academic article, to read*) and Xu et al. (*Exploring folksonomy for personalized search*) but it also confirmed previous results, identifying the authors previously mentioned. The final co-citation analysis was conducted on 2660 citations extracted from the WOS database. This time, data was analyzed not only to identify co-citation clusters but also included publication year and shortened journal names for the cited articles so that a time and origin component is added to the analysis enabling better overview of field development.

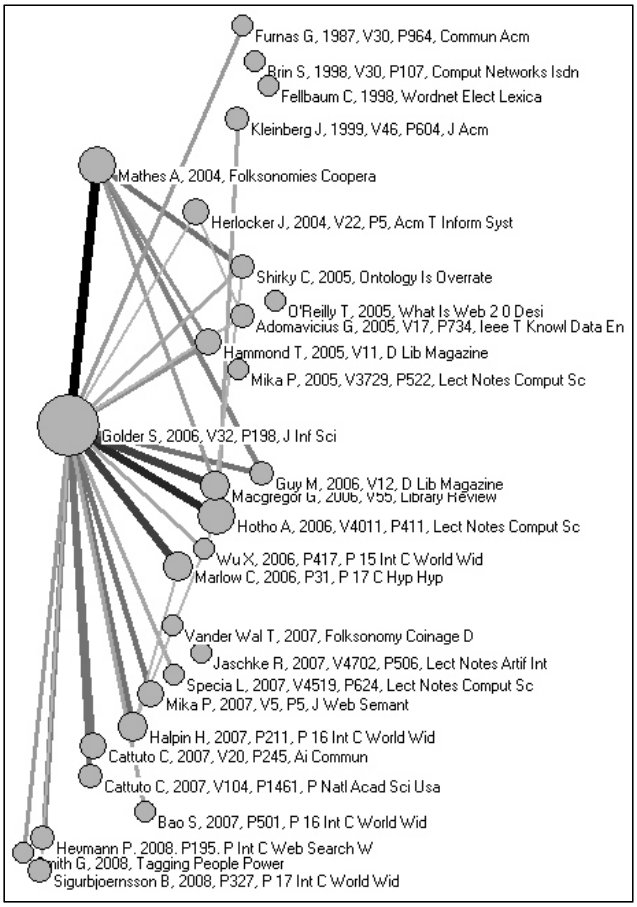

Figure 2.
Co-citation graph based on 2660 citations from WOS sorted by publication year

Figure 2 shows the co-citation graph from WOS records where the size of vertices indicates the number of citations while the thickness of lines indicates the number of co-citations between authors. To reduce the complexity of the visualization, figure 2 shows only authors that have more than 10 co-citations.

As we can see from the graph, the starting point for field development is the 2004 article by Mathes (*Folksonomies-cooperative classification and communication through shared metadata*) and the central article is again the 2006 paper by Golder and Hubermann (*Usage patterns of collaborative tagging systems).* Such visualization that includes a time component is a great start for a possible reading list for new researchers in the field where the development of the topic is clearly outlined with closely 30 key papers in the field. We can see that a large amount of most cited papers is from 2006–2007 where the field of research was defined and when the scientific debate was at its peak. Again, here we can see that the journal that published most cited articles are from the field of Computer Science and Library and Information Science.

5 Conclusion

This paper aimed to provide insight into the field of research on the topic of folksonomies by combining webometric and bibliometric tools in analysis of the data found in the most prominent databases in the field of Information Science. Since the field is fairly new with terminology and methods still being discussed, first analysis used the webometric method of counting cross-domain web mentions of competing terms. The results have shown that the most commonly used terms when describing the field of research are folksonomy, user-generated metadata, social tagging and collaborative tagging with the term folksonomy being the single most used term in use.

Based on these insights Web of Science, SCOPUS and Google Scholar were queried with a Boolean query including all the commonly used terms to ensure all the relevant papers were reached. These queries resulted with 346 papers with 2660 citations from WOS, 1581 papers with 8848 citations from SCOPUS and 1000 papers with 31234 citations from Google Scholar. Such

numbers clearly showed that the field is already well developed with a respectable number of papers published on the topic.

Next, the top 20 most cited articles from each database were compared in order to identify key paper and authors. It was shown that only 7 papers are present in both WOS and SCOPUS top 20, and when taking into account Google Scholar that number falls down to only three articles. This has shown that the field is very heterogenic, with only several papers being present as top cited in all the databases. When we examined the journals where the most cited papers were published, two main subfields of Information Science that are interested in the topic of folksonomies arose: Computer Science and Library and Information Science. Computer Science perspective was concerned more with using folksonomies and tags to improve the effectiveness of different systems and information retrieval, especially in the domain of Semantic Web, while Library and Information Science papers were more interested in researching user motivations for tagging and the potential of user tags in enhancing resource description.

Finally, a co-citation analysis was conducted on the citation data from both SCOPUS and WOS databases. The data from SCOPUS has given insight into the most influential authors in the field and their mutual connections, while the data from WOS included a time component that enabled the tracking of the field development. The best identification of key papers and authors is achieved in figure 2 which gives a chronological reading list for all new researchers in the field trying to explore the heterogenic field of folksonomy research.

This analysis confirmed that the field of folksonomy research is a relevant topic inside the Information Science field, with a respectable number of papers published in the most prominent databases for the field. It identified key authors and papers, as well as provided a chronological list of key papers and their mutual connections by conducting a co-citation analysis. Further analysis should include a topic analysis of the papers published on the topic in recent years thus providing insight into the current state of research and possible future directions.

References

Batagelj, V., Mrvar, A. (2003). Pajek – analysis and visualization of large networks. In: Jünger, M., Mutzel, P. (Eds.). *Graph Drawing Software*, Berlin: Springer, pp. 77–103.

Beel, J. & Gipp, B. (2007). Google Scholar's ranking algorithm: the impact of articles' age: an empirical study. *Proceedings of the 6th International Conference on Information Technology: New Generations*. pp. 160–164.

Golder, S.A. & Huberman, B.A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32 (2), 198–208.

Harzing, A. W. (2007). Publish or Perish. http://www.harzing.com/pop.html <22.1.2015>.

Harzing, A. W. (2008). Google Scholar: a new data source for citation analysis. http://www.harzing.com/pop_gs.htm <22.1.2015>.

Lasić-Lazić, J.; Špiranec, S. & Ivanjko, T. (2014). Tag-Resource-User: a review of approaches in studying folksonomies. *QQML Journal*, 3, 683–692.

Mathes, A. (2004). Folksonomies – cooperative classification and communication through shared metadata. *Computer Mediated Communication*, 47, 10.

Persson, O. D., Danell, R. & Wiborg Schneider, J. (2009). How to use Bibexcel for various types of bibliometric analysis. In: Åström, F. et al. (Eds.). *Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday*, International Society for Scientometrics and Informetrics pp. 9–24.

Peters, I. (2009). *Folksonomies: indexing and retrieval in Web 2.0*. Berlin: De Gruyter.

Thelwall, M. (2013). Webometrics and Social Web research methods. http://www.scit.wlv.ac.uk/~cm1993/papers/IntroductionToWebometricsAndSocialWebAnalysis.pdf <22.1.2015>.

Vander Wal, T. (2007). Folksonomy coinage and definition. http://vanderwal.net/folksonomy.html <22.1.2015>.

White, H. D., & McCain, K. W. (2009). 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.

Economic and Business Studies Journals and Readership Information from Mendeley

Kaltrina Nuredini, Isabella Peters

ZBW – German National Library of Economics
Düsternbrooker Weg 120, 24105 Kiel, Germany
{k.nuredini, i.peters}@zbw.eu

Abstract

We present Mendeley-readership information for 30 journals from the German Handelsblatt ranking for Economics and Business Studies from 2010/2012. We use readership data to characterize both fields by journals with over twenty years of publication activity. The analysis focusses on journal output, reader counts, scientific disciplines, academic status as well as geographic origin of readers. The results show that Mendeley provides relatively good coverage of research articles for both disciplines. The majority of readers are PhD students in Business Administration from USA and Germany. Moderate correlations are found between journals' reader numbers and impact factors. The results suggest that Mendeley readership data on journal level adds useful information to research evaluation and journal rankings and helps economists to publish in the best journal according to the intended target groups.

Keywords: Altmetrics, Readership information, Journal ranking, Economics

In: F. Pehar/C. Schlögl/C. Wolff (Eds.). Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th–21st May 2015. Glückstadt: Verlag Werner Hülsbusch, pp. 380–392.

1 Introduction

Today social media are a good believe that it will have an impact towards different fields of study, especially Science 2.0 and scholarly communication. Within these web-based environments, it is important to have tools that help researchers to evaluate, conduct and publish their research. Hence, a large number of economic researchers in Germany often use social media, especially Wikipedia, DropBox, and Facebook to communicate, collaborate, share literature, disseminate works, sharp ideas with other people, and identify research opportunities (Siegfried, Mazarakis, & Peters, 2014). They also deal with many alternative publication formats, e.g., working papers or blogs, to disseminate their findings. These new scientific workflows lead to two main challenges: 1) the variety of publications in the web makes it difficult to decide what is important to read or where research findings should be present, which is closely connected to 2) traditional bibliometric means do not necessarily apply to these new (social) web-based publication formats and practices. As Haustein (2012) states, scientific output is growing day by day and determining the number of journals on the market of scientific publications is not easy. In bibliometric studies the basic indicator for the absolute size of published research output is the number of publications *P* given for a discipline, a journal, an author, or any other level of aggregation. Haustein (2012) evaluated 45 journals in physics with a total of 168,109 papers. She found out that the number of scholarly journals and average output per journal increases in some particular years and decreases in others resulting in a large amount of publications available making it difficult for researchers to decide what to read. Also, in the (social) web the notion of trustworthy high-quality publication outlets, as typically reflected by peer-reviewed journals in the offline scientific world, changes. Moreover, journal and research rankings which assist readers and evaluators in quality assessment have been criticized for their narrow scope.

Although Seiler and Wohlrabe (2014) found the impact factor-based ranking of economic journals to be relatively robust, even when the top 5 and top 10 most cited articles were excluded, and therefore considered the impact factor (IF) a reliable quality indicator, there are good reasons for the boycott of research rankings in Economics (Berlemann & Haucap, 2012). For example, rankings often focus on articles from highly ranked journals and neglect new, but important articles which either did not make it into the top-journals

because of their timeliness or have not had time to accumulate citations yet. However, social media and web-based publication formats can provide new approaches for research ranking and evaluation as has been discussed by a.o. Schlögl et al. (2013). Thelwall et al. (2013) found that some social media-mentions, i.e., altmetrics, correlate with citations. Since citations require some time to accumulate with altmetrics researchers would get to know almost right away after publication what their research impact is, meaning how their research is used, communicated, and shared via social media tools.

The usage of social media-based metrics in science evaluation is still in its experimentation phase and it has not yet been necessarily validated to fully complement the research evaluation toolbox (e.g., regarding applicability to different disciplines; Schlögl et al., 2013). Also, it is still open whether altmetrics reflect quality or popularity of research products. Therefore, this research will focus on Economics' and Business Studies' publication outlets (i.e., journals), their coverage and usage in Mendeley and will provide more evidence on the potential of altmetrics in terms of number, geographical information, discipline, and academic status of readers. We exploit the information found on user profiles to determine the value of altmetric indicators and to add a further layer to traditional research evaluation which often only considers citations but not the characteristics of citers. For example, Haustein and Larivière (2014) have shown that the majority of users in Mendeley are PhD students and Postdocs. Also, by understanding who reads economic journals, in terms of finding the appropriate target group, researchers can better choose where to publish.

2 Methods and data sources

The top 30 journal names from the Handelsblatt (handelsblatt.com) ranking are selected of which 15 come from Business Studies (from the Handelsblatt journal ranking in 2012) and the other half is from Economics (from the 2010 Handelsblatt journal ranking). The Handelsblatt ranking integrates three sources, i.e., Erasmus Research Institute of Management (EJL), German Academic Association for Business Research (VHB-JOURQUAL 2.1), and Social Science Citation Index (SSCI). In our study the journals that are shared in both disciplines are considered only once and are allocated in Busi-

ness Studies (BS) whereas Economics (E) journals are replaced with journals following right after the top 15. The Handelsblatt journal ranking evaluates the impact of journals from E, BS and also considers multidisciplinary journals like *Nature*. *Nature* is ranked among the top 15 journals in the Handelsblatt ranking but because of its comparably large number of articles published (66,813) that would bias the results *Nature* was excluded and replaced with *Quarterly Journal of Economics*.

Mendeley.com is a social reference management system that allows users to search for papers, add them to their libraries, apply tags, and organize them in folders for better retrieval. Its catalog contains more than 40 million entries coming from user libraries which are merged into a single Mendeley research catalogue (Hammerton et al., 2012). As such Mendeley might be called a crowed-sourced library, since the study of Zahedi, Costas and Wouters (2013) shows that Mendeley has the highest coverage of readership information compared to Twitter, Wikipedia and Delicious. Every publication has readership counts reflecting its popularity within the Mendeley community, i.e. number of readers having a particular publication in their libraries. Moreover, users have profile pages with personal information like their discipline, research interests, academic status, and geographical information. We will provide readership information on journal level, especially regarding what researchers are active on Mendeley. Articles are categorized in disciplines only on the basis of the user information gathered from user profiles. Academic status informs about target groups such as undergraduate, postgraduate, professor, researcher, etc. of particular journals. User country metric saves geographical location of the users.

Since users are allowed to add papers to their libraries this sometimes results in duplicates within Mendeley. Additionally, Hammerton et al. (2012) stated that there might be papers without unique identifiers added and therefore could not be used for duplicate detection and removal. According to Bar-Ilan et al. (2012) and based on our self-testing, searching Mendeley by title or keywords is problematic since it does not support special character search. Moreover, if a journal name is searched to retrieve all the papers that are published in that journal, instead, this search will retrieve all entries that have a minimum of two words in common with that specified journal title.

Hence, in order to avoid data duplication, missing values, and search issues, we used DOIs for gathering readership data from Mendeley and the CrossRef API to retrieve the DOIs for all publications of the 30 journals published in the years 1994–2013. Data were collected in September 2014

and added to a MySQL database for further analyses. When searching for articles via DOIs in Mendeley, it is still possible that a DOI has been misspelled by the users and results in no hits. Or, some DOIs point to the same article but are written differently. As for example the DOI "10.5465/AMJ.2008.33665124" is found in CrossRef for the *Academy of Management Journal* but in Mendeley and at the journal webpage itself it appears as 10.5465/AMR.2008.33665124. Since by now there has been no useful way of searching Mendeley for maximal recall (i.e., combination of textual and DOI-queries) we work with DOI-based searches by knowing that results of these searches might be an underestimation of actual readership.

3 Results

We identified 51,473 papers from CrossRef for the publication years 1994–2013 of which 39,937 articles have a DOI in Mendeley. BS holds 24,439 papers of which 4,083 DOIs (17%) are not found either because of missing DOIs or because they are not of interest to the Mendeley community (table 1).

Table 1. Total number of DOIs in Mendeley

discipline	total number of papers from WoS	total number of papers from Crossref	total number of papers with DOIs in Mendeley
Business Studies (BS)	22,036	24,439	20,356
Economics (E)	24,573	27,034	19,581
Total	46,609	51,473	39,937

In E 7,453 DOIs (28%) are not retrieved from Mendeley, hence, BS journals are better covered in Mendeley although there are more E articles in CrossRef. *American Political Science Review* (*vapsr*) leads with 3,588 papers published in 20 years in twelve issues per year (see table 2).

Table 2. Description of journal output for 30 journals

journal	short	Impact Factor	total number of papers (P)	number of issues per year
Academy of Management Journal	amj	4.974	1,502	6
Academy of Management Review	amr	7.917	1,9	4
Administrative Science Quarterly	asq	2.394	999	4
American Economic Review	aer	3.305	3,33	11
American Political Science Review	vapsr	3.844	3,588	5
Annals of Statistics	vas	2.44*	1,744	6
Econometrica	eco	3.504	1,595	6
Economic Journal	vej	2.587	2,645	8
European Economic Review	veer	1.364	1,984	8
Games and Economic Behaviour	vgeb	1.025	2,154	6
Information Systems Research	isr	2.322	689	4
International Economic Review	vier	1.415	1,056	4
International Organization	vio	2.6	665	4
Journal of Accounting and Economics	vjae	2.833*	791	5
Journal of Business and Economic Statistics	vjbes	2.32	1,326	4
Journal of Business Research	vjbr	1.306	3,039	12
Journal of Consumer Research	jcr	2.783	1,394	4
Journal of Econometrics	vje	1.533	2,495	12
Journal of Economic Theory	vjet	0.919	2,544	6
Journal of Finance	jf	6.033	2,489	6
Journal of Financial Economics	jfe	3.769	1,815	12
Journal of Health Economics	vjhe	2.254	1,381	5
Journal of Labor Economics	vjle	1.979	681	4
Journal of Marketing	jm	3.819	1,025	6
Journal of Marketing Research	jmr	2.66	1,253	6
Journal of Monetary Economics	jme	2.065	1,649	8
Journal of Political Economy	jpe	3.617	986	6
Management Science	ms	2.524	2,83	12
Quarterly Journal of Economics	qje	5.966	983	4
Review of Economic Studies	vres	3.235	941	4

Figure 1 shows the comparison of the total number of articles for each of the 30 journals found in Crossref and Mendeley. *Information Systems Research* (*isr*) is best covered in Mendeley, while *vapsr* has the worst ratio of articles covered and number of articles published.

When investigating the readership numbers for each publication year and the number of articles findable on Mendeley, we can see that there is a bias towards newer articles (see fig. 2). Mendeley-readers add more current papers to their libraries resulting in a good coverage of newer research. However, even that there are less older articles available on Mendeley, those articles gain comparably high reader counts.

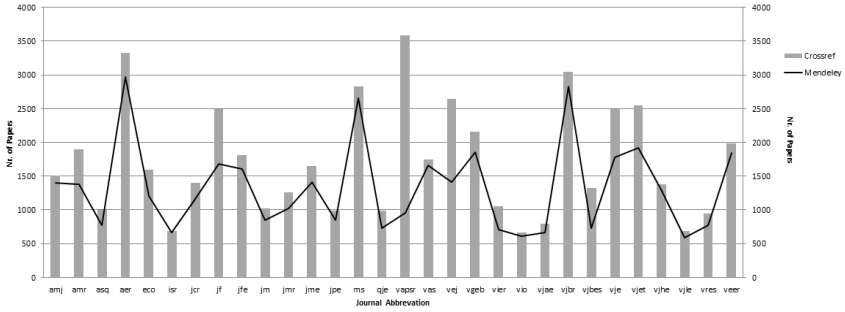

Figure 1. Comparison of DOI availability and Mendeley coverage

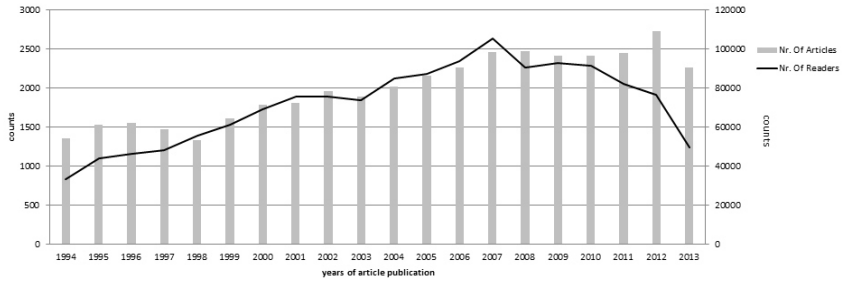

Figure 2. The year-wise distribution of articles and readers in Mendeley

3.1 Reader count based on Mendeley

The most often read article in our dataset is "The Coming of Age for Qualitative Research: Embracing the Diversity of Qualitative Methods", with a reader count of 10,125. The readership information also reveals that from BS, *Academy of Management Journal* (*amj*) has the largest reader count number and *vjbr* from E. The most popular article from *amj* was published in 2007 and the readership number for that journal increases for papers published from 2006 onwards.

Figure 3 shows a clear difference between the journals of the two disciplines. The 15 journals corresponding to BS (black) appear to be heavily read as opposed to the other half from E (grey) in terms of readership numbers. Journals that are shared from both disciplines are considered once (patterns). Based on our data, BS journals are the most read journals in 2007.

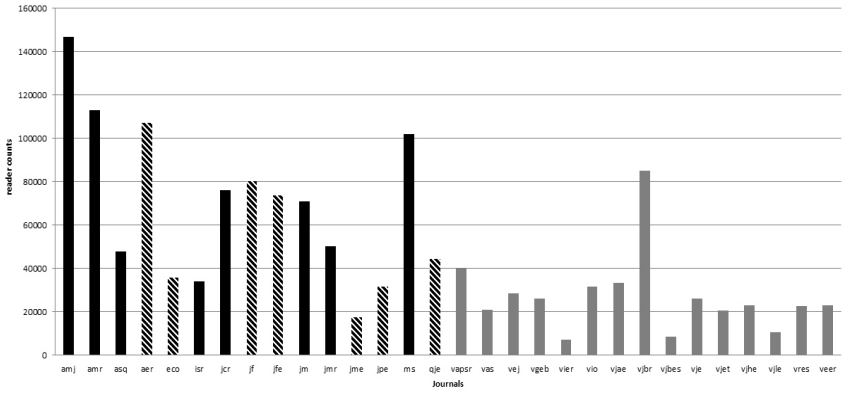

Figure 3. Journal-wise distribution of reader numbers

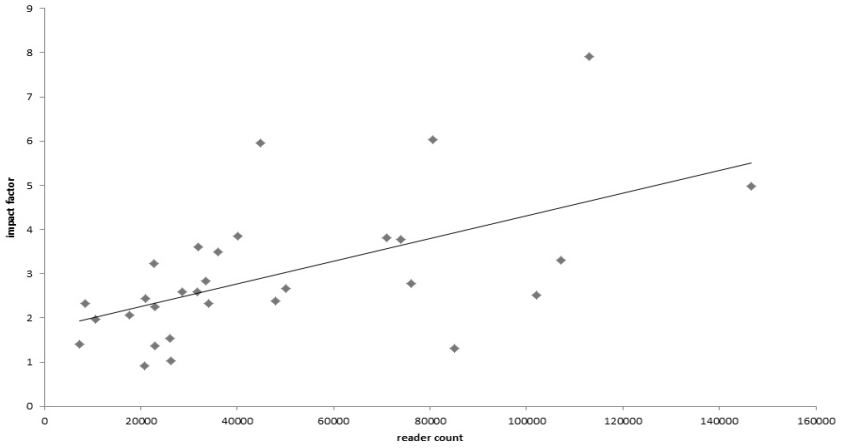

Figure 4.
Scatter plot for total number of readers for 30 journals and their Impact Factors

Traditionally the importance of a journal is determined by the IF based on Web of Science citation data. The IF values for the 30 journals are given in table 2 and come from the 2013 Journal Citation Reports Social Science Edition (JCRSSE). The IF of two journals (marked with * in table 2) had to be obtained from the journals' website since they were not included in the 2013 JCRSSE. To determine whether there is a symmetric relation between the IF and readership counts on journal level we applied both Spearman rank correlation p and Pearson correlation r (see fig. 4) resulting in $p = .628$ and $r = .574$. We compare readership counts and IF since both can be used as a

criterion for journal evaluation. People are more likely to make quality judgments based on these indicators because this information is available.

We found that reader numbers of both disciplines, BS and E, are positively, but only moderately correlated with the IF of the journals. This becomes also visible when comparing the journal ranking by reader count and by IF: *amj* has the highest reader count whereas the *Academy of Management Review* (*amr*) has the highest IF. Mohammadi and Thelwall (2014) received similar results when comparing Mendeley's reader counts with Web of Science citations for articles from social sciences (.516) and from humanities (.428). Hence, as shown before (Haustein et al., 2014), reader counts and IFs reflect different impact on users of scholarly literature and add another means to journals evaluation.

3.2 User discipline based on Mendeley

Table 3. User's discipline for articles from BS and E

discipline of users	Nr. of readers of Business Studies (BS) articles	Nr. of readers of Economics (E) articles
Arts and Literature	315	408
Astronomy/Astrophysics/Space Science	78	77
Biological Sciences	644	2,02
Business Administration	**35,819**	**10,688**
Chemistry	68	651
Computer and Information Science	4,136	3,017
Design	359	260
Earth Sciences	104	221
Economics	27,623	14,181
Education	1,232	1,552
Electrical and Electronic Engineering	201	555
Engineering	1,653	1,294
Environmental Sciences	558	986
Humanities	343	627
Law	416	489
Linguistics	89	119
Management Science / Operations Resear	9,143	3,036
Materials Science	30	26
Mathematics	1,164	2,66
Medicine	441	1,193
Philosophy	462	459
Physics	256	390
Phsychology	3,985	2,122
Social Sciences	8,076	25,776
Sports and Recreation	133	86

Twenty-five different disciplines are identified, but Mendeley splits every discipline in further sub-disciplines. In our dataset most of the readers have a background in Business Administration with a total number of 35,819 users reading BS articles and 10,688 readers saving E articles (see table 3). Articles from E were mostly saved by readers from the Social Sciences.

3.3 Academic status based on Mendeley

Figure 5 shows that for BS and E PhD Students and Master students are the core Mendeley readers. Mohammadi et al. (in press) also observed that the majority of Mendeley-readers are PhD Students from other disciplines, i.e., Clinical Medicine, Engineering and Technology, Social Science, Physics and Chemistry.

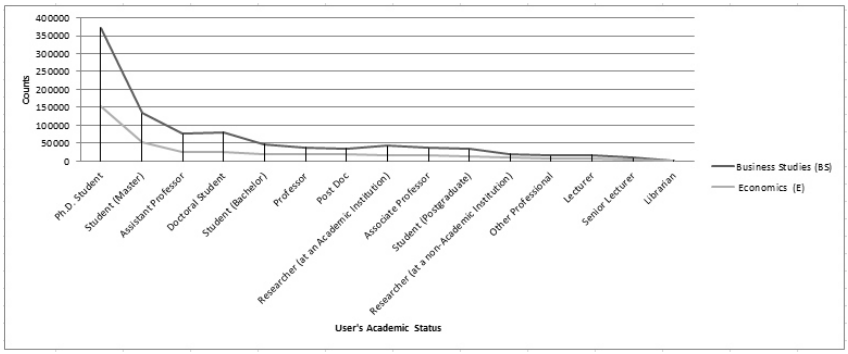

Figure 5. Readers' academic statuses for journals from BS and E

Most PhD students (in this case merged with Doctoral Students) read the *amj* whereas the least amount of PhD students use the *International Economic Review* (*vier*; see fig. 6). The best PhD student and Master student ratio can be found for the *Journal of Marketing* (*jm*).

3.4 User country based on Mendeley

Mendeley users are able to save their geographical location in their profiles leading to 119 different countries found. For economics journals the top 3 countries with the most readers in BS and E are from the USA, Germany and the United Kingdom.

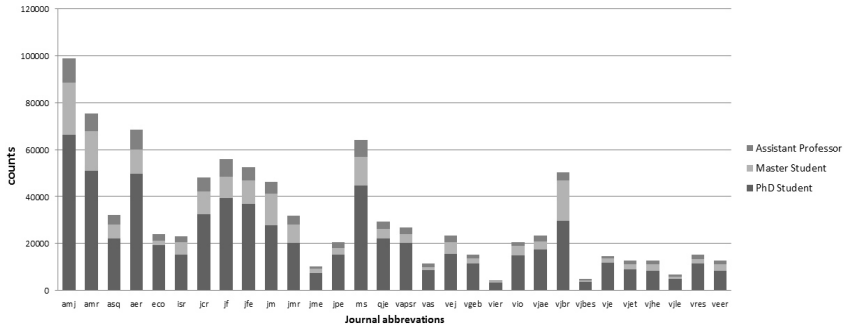

Figure 6. Top 3 academic statuses of readers per journal

4 Conclusion

Previous studies (e.g., Mohammadi, Thelwall, & Kousha, in press; Mohammadi et al., in press) confirm that Mendeley readership counts are important for reflecting scholarly impact. Our analyses also revealed that Mendeley readership information provides helpful information for economic researchers on a short term basis since both coverage and popularity of journals are biased towards current research. The study aimed at characterizing 30 Economics and Business Studies journals in terms of journal output, reader count, user discipline, academic status, and location of readers. Mendeley was used as source for readership information. The *Academy of Management Journal*, which was shared from both disciplines BS and E, was the journal with the largest number of readers coming from Business Administration. Most of the readers of journals from our dataset were PhD students and Master students and often came from USA, Germany, and the United Kingdom. Since we could show that readership information was only moderately correlated with the journals' impact factor we believe that reader counts complement the research assessment toolbox by reflecting another type of journal impact which goes beyond citations although this has to be confirmed by future studies. Moreover, Mendeley can reveal reader characteristics which might be important for economists in order to determine the appropriate publication outlet in terms of target groups.

References

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond Citations: Scholars' Visibility on the Social Web. In: *Proceedings of the 17th International Conference on Science and Technology Indicators*, Montréal, Canada, Vol. 1, pp. 98–109.

Berlemann, M., & Haucap, J. (2012). Which factors drive the decision to boycott and opt out of research rankings? CESifo Working Paper: Economics of Education, No. 3997. http://www.cesifo-group.de/ifoHome/publications/working-papers/CESifoWP/CESifoWPdetails?wp_id=19070339.

Hammerton, J.A., Granitzer, M., Harvey, D., Hristakeva, M., & Jack, K. (2012). On Generating Large-scale Ground truth Datasets for the Deduplication of the Bibliographic Records. In: *Proceedings of 2nd International Conference on Web Intelligence, Mining and Semantics*, Craiova, Romania. doi:10.1145/2254129.2254153

Haustein, S. (2012). *Multidimensional Journal Evaluation*. Berlin: De Gruyter Saur:

Haustein, S., & Larivière, V. (2014). Mendeley as a Source of Readership by Students and Postdocs? Evaluating Article Usage by Academic Status. In: *Proceedings of IAUTL Conference*, Helsinki, Finland. http://docs.lib.purdue.edu/iatul/2014/altmetrics/2/.

Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting Biomedicine: An Analysis of Tweets and Citations in the Biomedical Literature. *Journal of the Association for Information Science and Technology*, 65 (4), 656–669. doi: 10.1002/asi.23101

Mohammadi, E., & Thelwall, M. (2014). Mendeley Readership Altmetrics for the Social Sciences and Humanities: Research Evaluation and Knowledge Flows. *Journal of the Association for Information Science and Technology*, 65 (8), 1627–1638. doi: 10.1002/asi.23071

Mohammadi, E., Thelwall, M., & Kousha, K. (in press). Can Mendeley Bookmarks Reflect Readership? A Survey of User Motivation. *Journal of Association for Information Science and Technology.* http://cba.scit.wlv.ac.uk/~cm1993/papers/CanMendeleyBookmarksReflectReadershipSurvey_preprint.pdf.

Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (in press). Who Reads Research Articles? An Altmetrics Analysis of Mendeley User Categories. *Journal of Association for Information Science and Technology.* http://www.scit.wlv.ac.uk/~cm1993/papers/WhoReadsResearchArticlesPreprint.pdf.

Schlögl, C., Gorraiz, J., Gumpenberger, C., Jack, K., & Kraker, P. (2013). Download vs. Citation vs. Readership Data: The Case of Information Systems Journal. In

Proceedings of the 14th International Society of Scientometrics and Informetrics Conference, Vienna, Austria, pp. 626–634.

Seiler, C., & Wohlrabe, K. (2014). How Robust are Journal Rankings Based on the Impact Factor? Evidence from the Economic Sciences. *Journal of Informetrics,* 8 (4), 904–911.

Siegfried, D., Mazarakis, A., & Peters, I. (2014). Nutzung von Social-Media-Diensten in den Wirtschaftswissenschaften, Kiel: ZBW. http://www.zbw.eu/file-admin/pdf/presse/2014-zbw-studie-nutzung-social-media.pdf.

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. *PloSOne*, 8 (5), e64841. doi: 10.1371/journal.pone.0064841

Zahedi, Z., Costas, R. & Wouters, P. (2014). How Well Developed are Altmetrics? A Cross-Disciplinary Analysis of The Presence of 'Alternative Metrics' in Scientific Publications. *Scientometrics*, 101 (2), 1491–1513.

Exploring Coverage and Distribution of Identifiers on the Scholarly Web

Peter Kraker[1], Asura Enkhbayar[1], Elisabeth Lex[2]

[1] Know-Center
Inffeldgasse 13/VI, 8010 Graz, Austria
{pkraker, aenkhbayar}@know-center.at
[2] Graz University of Technology
Inffeldgasse 13/V, 8010 Graz, Austria
elisabeth.lex@tugraz.at

Abstract

In a scientific publishing environment that is increasingly moving online, identifiers of scholarly work are gaining in importance. In this paper, we analysed identifier distribution and coverage of articles from the discipline of quantitative biology using arXiv, Mendeley and CrossRef as data sources. The results show that when retrieving arXiv articles from Mendeley, we were able to find more papers using the DOI than the arXiv ID. This indicates that DOI may be a better identifier with respect to findability. We also find that coverage of articles on Mendeley decreases in the most recent years, whereas the coverage of DOIs does not decrease in the same order of magnitude. This hints at the fact that there is a certain time lag involved, before articles are covered in crowd-sourced services on the scholarly web.

Keywords: Scholarly identifiers, Pre-prints, arXiv, DOI, Readership

In: F. Pehar/C. Schlögl/C. Wolff (Eds.). Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th–21st May 2015. Glückstadt: Verlag Werner Hülsbusch, pp. 393–403.

1 Introduction

In a scientific publishing environment that is increasingly moving online, identifiers of scholarly work are gaining in importance. With the advent of pre-print archives, there is often more than one version of an article available and these versions may be hosted in various places around the web. Scholarly communication is no longer limited to articles alone, but it also takes place in different forms on various social media platforms. Identifiers are therefore crucial for disambiguation and traceability of scholarly articles and their reception.

The need for persistent identifiers is often mentioned in the literature (see e.g. Davidson & Douglas, 1998; Bourne & Fink 2008) and consequently, a variety of identifier systems have been proposed (see e.g. Van De Sompel et al., 2001; Warner 2010). Prominent examples for identifiers on an article level are the Digital Object Identifier or DOI (DOI Foundation, n.d.) and the arXiv ID. Notable identifiers on the author level are author-based identifiers such as ORCID (Haak et al., 2012) and Researcher ID (Thomson-Reuters, n.d.). Some of the most longstanding identifiers predate the digital age, including the International Standard Book Number (ISBN) and the International Standard Serial Number (ISSN).

Despite their importance, little is empirically known about the coverage and distribution of scholarly identifiers, and how they propagate on the scholarly web. In our work, we are addressing this very gap in the scientometric literature. Specifically, our research was guided by the following research questions:

- How are scholarly identifiers distributed in crowd-sourced systems, e.g. pre-print archives and online reference management systems? Which identifier combinations are the most common? Who are the top providers of identifiers?
- Does the provision of different identifiers have an influence on findability of scientific publications in other bibliographic and bibliometric sources?

2 Data and method

In this study, we analysed arXiv papers from the discipline of quantitative biology (arXiv short code: q-bio). We chose this discipline because it represents one of the largest disciplines on Mendeley (Kraker et al., 2012). Three different data sources were used in this study: (i) arXiv, a preprint archive (ii) CrossRef, a metadata and linking service, and (iii) Mendeley, an online reference management system.

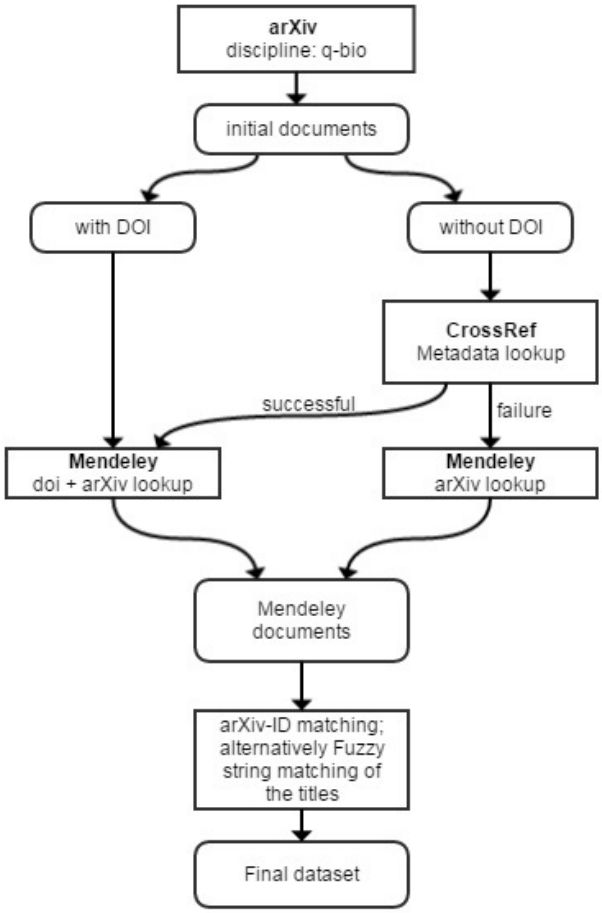

Figure 1. Data collection pipeline

The data collection pipeline is shown in figure 1. At first, we collected metadata on all publicly available articles for quantitative biology. In all cases, the most recent upload to arXiv was used and all older entries were discarded. This resulted in n = 14,195 metadata records. Quantitative biology represents a medium-to-small collection on arXiv. The collected metadata includes: arXiv ID, DOI (optional), title, authors, year, and journal (optional).

This data was sourced on 17.11.2014 and was used as a basis for all following steps. At first, the initial data set was divided into entries with DOI (n = 5,125 entries, 36.7%) and without a DOI (n = 8,980 entries, 63.3%). arXiv is primarily used as a way to disseminate pre-prints, and not all authors add a DOI to the arXiv record after an article has been published. Therefore, we performed a CrossRef meta-data lookup in order to acquire additional DOIs. We used the following metadata to search for an entry: title, author, journal, and year.

With this procedure, we found DOIs for an additional 1,885 entries, bringing the number of entries with a DOI up to 7,100 (50.02%). We then attempted to retrieve the corresponding documents for all entries on Mendeley. We used either the arXiv ID or both the DOI and the arXiv ID to locate the document. If both arXiv ID and DOI yielded a result on Mendeley, the Mendeley IDs were compared. If they didn't match, we used the result, which contained additional identifier fields, e.g. a PubMed ID, if available. If both results contained the same amount of articles, we chose the item found with the DOI.

Finally, we compared the arXiv ID of the obtained Mendeley document with the original arXiv entry. If the obtained Mendeley document did not provide one, the two titles were compared using approximate string matching in order to ascertain matching documents.

After this procedure, we arrived a final set of n = 11,570 articles that could be found on Mendeley (81.5%). For these articles, we retrieved basic readership data and identifier data. Available identifiers on Mendeley are:[1]

- arxiv: arXiv ID
- doi: Digital Object Identifier (DOI)
- isbn: International Standard Book Number (ISBN)
- issn: International Standard Serial Number (ISSN)
- pmid: PubMed ID (assigned to publications indexed in PubMed)

1 See http://dev.mendeley.com/methods/#catalog-documents.

- scopus: Scopus ID (assigned to publications indexed in Scopus)
- ssrn: Social Science Research Network (SSRN) ID

3 Results

3.1 Identifier distribution in arXiv and findability on Mendeley

Table 1 sums up the basic results of the crawling process. Of the 14,195 unique articles, 36.7% had a DOI on arXiv. Using CrossRef, an additional 1,885 DOIs could be found, bringing the share of articles with a DOI up to 50.02%. 11,570 articles (81.5%) could finally be found on Mendeley.

Table 1. Results of the crawling process; n = 14,195 articles

arXiv: total docs	arXiv: docs with DOI	CrossRef: additional DOIs	Mendeley: found
14,195	5,125 (36.7%)	1,885 (13.3%)	11,570 (81.5%)

There was a difference in findability with respect to whether we used a DOI or the arXiv ID to search for the articles on Mendeley (see also table 3). Of the 14,195 articles, 72.6% could be retrieved on Mendeley using the arXiv ID. In contrast to that, 91.4% of the 7,100 articles with a Digital Object Identifier (either on arXiv or via metadata lookup on CrossRef) could be found on Mendeley using the DOI.

One of the reasons for that could be that records with a DOI do represent articles that have eventually been published in a journal. In order to test this assumption, we analysed the registrants for all entries with a DOI (7,100 articles). We used a list of DOI registrants by Alf Eaton[2] with manual extensions to identify registrants. The results confirm our assumption (see table 2). The top registrants are established publishers such as Elsevier and Springer. These publishers usually assign DOIs to articles published in their journals and books, in contrast to archives such as figshare, which assign a DOI to any submitted article regardless of whether it was published in a journal or not.

Table 2. DOI registrants of articles; n = 7,100 articles

2 See https://gist.github.com/hubgit/5974843.

Registrant	# DOIs	Percentage
American Physical Society	1,507	21.2%
Elsevier	1,029	14.5%
Springer-Verlag	668	9.4%
Public Library of Science	502	7.1%
IOP Publishing	439	6.2%
American Institute of Physics	335	4.7%
Proceedings of the National Academy of Sciences	217	3.1%
Oxford University Press	194	2.7%
Springer (Biomed Central Ltd.)	180	2.5%
IOP Publishing – Europhysics Letters	141	2.0%
Other	1,888	26.6%
Sum	7,100	100%

To eliminate effects that relate to the nature of the article that has been posted on arXiv (whether it stayed a pre-print or went on to become a journal article), we also compared findability for articles that have both a DOI and an arXiv ID (see table 3). We also found a difference in these cases: 91.4% of articles with a DOI could be found using the very same identifier, whereas, only 71.4% of articles with a DOI could be found with the arXiv ID. The lowest findability was reported for articles with no DOI: of the 7,095 articles with no DOI, only 69.0% were retrieved using the arXiv ID.

Table 3. Findability of articles on Mendeley, depending on the identifier used; n = 14,195 articles

		found on Mendeley using	
	n	arXiv ID	DOI
arXiv ID & DOI	7,100 (50.02%)	5,414 (76.25%)	6,492 (91.44%)
arXiv ID	7,095 (49.98%)	4,896 (69.01%)	-
Sum	14,195 (100%)	10,310 (72.63%)	-

Another interesting fact found in the top providers is that the American Physical Society, which is, among other things, "working to advance and diffuse the knowledge of physics through its outstanding research journals"[3]

3 See http://www.aps.org/about/index.cfm.

is the top registrant for DOIs in quantitative biology. One of the reasons for that could be that arXiv allows authors to assign more than just one category to each article. The analysis of article categories (see table 4) shows that quantitative biology is the primary discipline for only 61.4% of articles with a DOI (4,358 articles). 30.1% (2,178 articles) are assigned to a primary category that falls into the discipline of physics. This indicates a high number of interdisciplinary articles in the sample.

Table 4. Distribution of disciplines in articles with a DOI (n = 7,100 articles)

Discipline	Number of articles	Percentage
Quantitative Biology	4,358	61.4%
Physics	2,178	30.7%
Computer Science	247	3.5%
Mathematics	211	3.0%
Statistics	105	1.5%
Quantitative Finance	1	0.0%
All	7,100	100.0%

Figure 2 shows the distribution of articles from 1992 to 2013. There is a strong, at times exponential increase in the number of articles. The coverage on Mendeley, however, has declined for the youngest articles as can be seen in figure 3. The percentage of articles with a DOI does not decrease in the same order of magnitude.

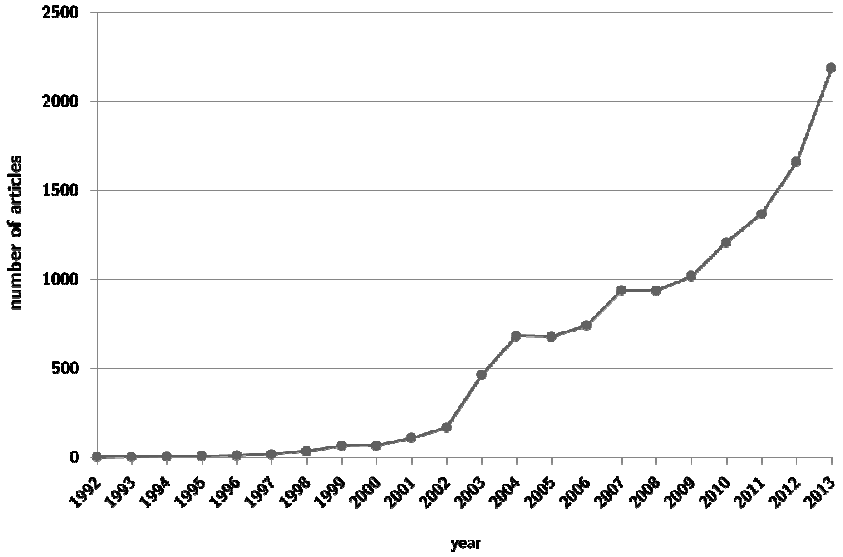

Figure 2. Distribution of articles between 1992 and 2013; n = 12,392 articles

Figure 3. Findability of articles on Mendeley and DOI coverage, 1992–2013; n = 12,392 articles

3.2 Distribution of identifiers on Mendeley

We then investigated the distribution of identifiers of all arXiv articles found on Mendeley in detail. Note that we only took metadata from Mendeley into account, which is why the numbers for arXiv ID and DOI differ to the analyses before. The distribution of identifiers on Mendeley can be seen in table 5. The arXiv ID is the most common identifier, followed by the Scopus ID, DOI and ISSN. In terms of readership, articles with a PubMed ID have the highest average readership.[4]

Table 5.
Identifier frequency and mean readership on Mendeley; n = 11,570 articles

	arxiv	doi	scopus	pmid	issn
frequency	10,351 (89.5%)	8,321 (71.9%)	8,409 (72.7%)	5,477 (47.3%)	8,119 (70.2%)
mean readership	20.4	25.4	25.4	32.4	25.9

Figure 4 shows the most common identifier combinations in the data. Here, a combination of all identifiers on Mendeley included in this analysis (arXiv ID, DOI, ISSN, PubMed ID and Scopus ID) is the most common identifier combination; a single arXiv ID comes second.

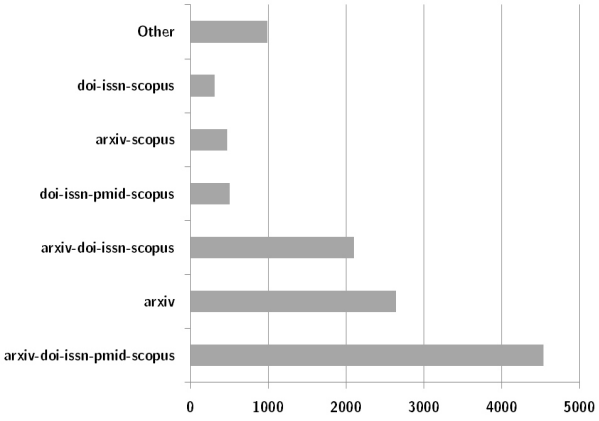

Figure 4.
Identifier combination frequency of articles on Mendeley; n = 11,570 articles

4 Note that we left ISBN out of this analysis, because the metadata quality was very poor with respect to this field on Mendeley.

4 Conclusions and future work

We found that when retrieving arXiv articles in quantitative biology from Mendeley, we were able to obtain more articles using the DOI than the arXiv ID. Even when we only considered articles that were assigned both identifiers, the effect was sizeable (91.4% vs. 72.6%). This indicates that the DOI may be a better identifier with respect to findability. Nevertheless, a single arXiv ID is the second most popular identifier combination on Mendeley. This suggests that pre-prints are being read – if at a lower level – even when they are not yet published in a journal.

We found that coverage of articles on Mendeley decreases in the most recent years, whereas the availability of DOIs does not decrease in the same order of magnitude. This hints at the fact that there is a certain time lag before articles are covered in crowd-sourced services on the scholarly web.

There are certain limitations to this work. We only looked at a single discipline (quantitative biology) and we only used three data sources in our study (arXiv, CrossRef and Mendeley), which may have had a significant influence on the results. Indeed, in a small-scale study using a random sample of 381 articles from Web of Science, Zahedi et al. (2014) report that they were able to retrieve only 47.7% of articles on Mendeley using the DOI or the title.

In the future, we therefore plan to extend this study to more disciplines and fields in order to substantiate the hypotheses emanating from the results in this study. In order to gain a deeper insight into the distribution and the coverage of identifiers on the scientific web, we are looking to include further data sources such as Web of Science, PubMed Central, Altmetric.com, and figshare.

Acknowledgments

The Know-Center is funded within the Austrian COMET program – Competence Centers for Excellent Technologies – under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth, and the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

References

Bourne, P. E., & Fink, J. L. (2008). I am not a scientist, I am a number. *PLoS Computational Biology*, 4 (12), e1000247. doi:10.1371/journal.pcbi.1000247

Davidson, L. A., & Douglas, K. (1998). Digital Object Identifiers: Promise and Problems for Scholarly Publishing. *The Journal of Electronic Publishing*, 4 (2). doi:10.3998/3336451.0004.203

DOI Foundation (n.d.). Digital Object Identifier System. http://www.doi.org/ <14.01.2015>.

Haak, L. L., Fenner, M., Paglione, L., Pentz, E., & Ratner, H. (2012). ORCID: a system to uniquely identify researchers. *Learned Publishing*, 25 (4), 259–264. doi:10.1087/20120404

Kraker, P., Körner, C., Jack, K., & Granitzer, M. (2012). Harnessing User Library Statistics for Research Evaluation and Knowledge Domain Visualization. In: *Proceedings of the 21st International Conference Companion on World Wide Web*, pp. 1017–1024. doi:10.1145/2187980.2188236

Sompel, H. Van De, Lagoze, C., Bekaert, J., Liu, X., Payette, S., & Warner, S. (2006). An interoperable fabric for scholarly value chains. *D-Lib Magazine*, 12 (10).

Thomson-Reuters (n.d.). ResearcherID. URL: http://www.researcherid.com <14.01.2015>.

Van de Sompel, H., & Beit-Arie, O. (2001). Open linking in the scholarly information environment using the OpenURL Framework. *New Review of Information Networking*, 7 (1), 59–76. doi:10.1080/13614570109516969

Warner, S. (2010). Author identifiers in scholarly repositories. *Journal of Digital Information*, 11 (1), 1–10.

Zahedi, Z., Haustein, S., & Bowman, T. D. (2014). Exploring data quality and retrieval strategies for Mendeley reader counts. In: *Metrics14 – ASIS&T Workshop on Informetric and Scientometric Research Introduction*. http://www.asis.org/ SIG/SIGMET/data/uploads/sigmet2014/zahedi.pdf <04.03.2014>.

Mapping the Spreading of Cited References over Research Fronts of Bibliographically Coupled Publications

Edgar Schiebel

AIT Austrian Institute of Technology GmbH,
Donau City Straße 1, 1220 Wien, Austria
edgar.schiebel@ait.ac.at

Abstract

This short paper deals with the delineation of research issues based on bibliographic coupling of publications assisted by visualization techniques. Cluster techniques, multidimensional scaling or spring models reveal agglomerations of similar publications but it is always difficult to have a clear picture of the thematic substructure of a research field or even a set of publications with a consistent content. The central research questions of this work are: How can we visualize the occurrence of cited references in an agglomeration of similar publications? Does the visualization of the occurrence of cited reference in bibliographically coupled publications help to understand how to delineate a research topic? Research fronts were defined as a local agglomeration of similar publications in a two dimensional space. This work proposes a visualization method using an overlay technique in 2D heat maps of bibliographically coupled publications. With this approach we could visualize and discuss to what extend research fronts are formed by several highly cited references that are the core of the underlying knowledge base. The approach is demonstrated for research related to foresight.

Keywords: Bibliographic coupling, Science mapping, Visualization, Overlay technique, Delineation of research issues

In: F. Pehar/C. Schlögl/C. Wolff (Eds.). Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th–21st May 2015. Glückstadt: Verlag Werner Hülsbusch, pp. 404–409.

1 Introduction

Bibliographic coupling is an established scientometric approach to delineate thematic structures of a research field, see Kessler (1963). Although cluster techniques, multidimensional scaling or spring models reveal agglomerations of similar publications, it is always difficult to have a clear picture of the thematic substructure of a research field. Bibliographic coupling uses references for the definition of clusters of similar publications that are called research fronts. Sometimes references are spread over the whole research field or just occur in a small number of similar publications. This work proposes a visualization method using an overlay technique in 2D heat maps of bibliographically coupled publications. With this approach we can visualize to what extend research fronts are formed by several highly cited references that form the knowledge base of the agglomeration. Publications about research on foresight were used as a case study.

This study focuses on the following research questions:

1. How can we visualize the occurrence of cited references in an agglomeration of similar publications?
2. Does the visualization of the occurrence of cited reference in bibliographically coupled publications help to understand how to delineate a research topic?

2 Methodology and data

The research fronts were mapped as bibliographically coupled publications using the Jaccard index to calculate the similarity of common references of pairs of publications. Research fronts were defined as a local agglomeration of similar publications in a two dimensional space. The positions of the publications were calculated with a spring model, see Kopcsa and Schiebel (1998). The parameters of the algorithm for the set of publications in this study were: maximum distance: 1.00, minimal distance: 2,35E-02, maximum force: 1; repulsive force: -2,5E-04 all other attracting forces based on the Jaccardindex (0 to 1). The local density of similar publications weighted by the Jaccard index was visualized with a 2D surface "heat" map. Parameters of the surface map are: x, y bt 0 and 1, grid pixel size 1/250, window of the

cosine weighted filter: 20 pixels in x and y direction. The method was introduced by Schiebel (2012).

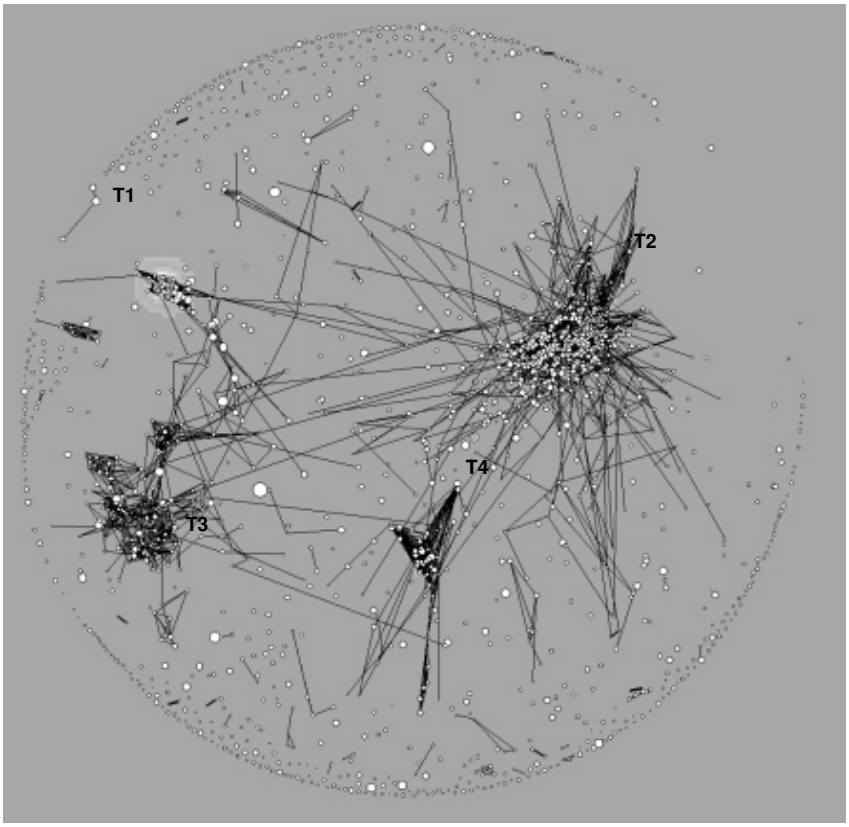

Figure 1. 2D surface map of bibliographically coupled publications, dots are publications, size is the number of references

Research fronts were identified as follows: selection of one publication and all related ones with at least one common reference in a hot zone, graphical exclusion of all publications that are not "near" the core of the hot zone and finally the selection of all publications with a Jaccard index greater than the expected value. In the last step the task was to decide whether the publications from a research front show a consistent common knowledge base or not. An overlay technique was used to visualize the spreading of cited references over the publications of the research front.

In this work we used data from the Web of Science™ Core Collection. The search word "foresight" in the "topic"-field for the time span 2000 to 2014 identified a set of 1980 publications. The selected publications where downloaded and computed with the software BibTechMon™.

3 Results

The "Foresight" map of bibliographically coupled publications reveals four research fronts, see fig. 1: T1 Brain Science: Mental Time Travel, T2 Different Kind of Foresight: Technology, Innovation, Corporation, …; T3 Rational Expectations in Economics: Perfect Foresight, Fiscal Foresight and T4 Hindsight Bias: Afterwards one knows better

The blue flags in fig. 2 show all publications that cited Martin (1995). Grupp et al. (1999) has a distribution in the same region. Additionally other publications about technology foresight suggest the the definition and the name *"F1:Technology Foresight"* for the subtopic in this area.

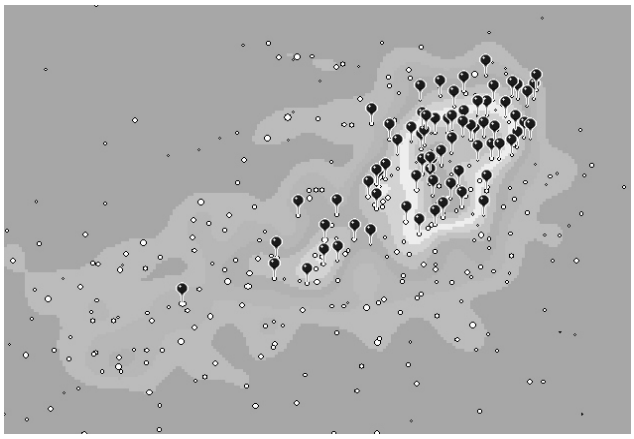

Figure 2. Publications citing Martin (1995)

Cuhls (2003) and Könnölä et al. (2007) as well as other publications about foresight processes show an occurrence as cited references in the middle of the island of fig. 2. This is the research field of "*F2: Foresight Processes*".

Finally, the tale of the island of fig. 2 is dominated by publications citing van der Heijden (1996) and Bradfield et al. (2005) i.e. work on *"F3: strategic and corporate foresight"*.

The three sub issues of the research front "T2 Different Kind of Foresight: Technology, Innovation, Corporation, …" are marked in fig. 3

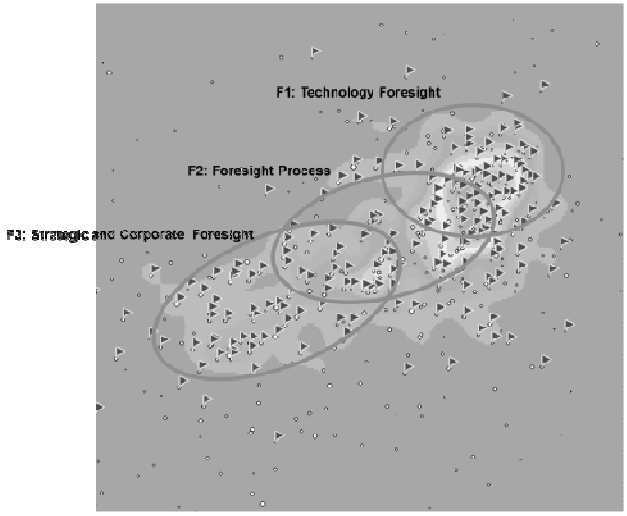

Figure 3. Substructure of the research front T2

4 Conclusions

It could be shown that the visualization of the occurrence of cited publications in an agglomeration of bibliographically coupled publications offer clear information on the underlying knowledge base. The content of the knowledge base that is formed by cited references suggests a good delineation of the research front and its substructure.

Further research is foreseen to develop a quantitative approach for the delineation of research fronts based on the statistic of occurring references.

References

Bradfield, R.; Wright, G.; Burt, G.; Cairns, G. & van der Heijden, K. (2005). The origins and evolution of scenario techniques in long range business planning. *Futures* 37 (8): 795–812.

Cuhls, K. (2003): From Forecasting to Foresight processes – New participative Foresight Activities in Germany. In: *Journal of Forecasting*, 22 (2–3), 93–111.

Grupp, H.; Linstone, H. A. (1999). National Technology Foresight Activities around the Globe: Resurrection and New Paradigms. *Technological Forecasting and Social Change*, 60 (1), 85–94.

Kessler, M. M. (1963). Bibliographic Coupling Between Scientific Papers. *American Documentation* 14, 10–25.

Könnölä, T.; Brummer, V.; Salo, A., (2007). Diversity in Foresight: Insights from the Fostering of Innovation Ideas. *Technological Forecasting and Social Change*, 74, 608–626.

Kopcsa, A.; Schiebel, E. (1998). Science and technology mapping: A new iteration model for representing multidimensional relationships. *Journal of the American Society for Information*, 49 (1), 7–17. doi: 10.1002/(sici)1097-4571(1998)49:1<7::aid-asi3>3.0.co;2-w

Martin, B. R. (1995). Foresight in science and technology. *Technology Analysis & Strategic Management*, 7 (2), 139–168. doi 10.1080/09537329508524202

Noll, M.; Frohlich, D.; Schiebel, E. (2002). Knowledge maps of knowledge management tools – Information visualization with BibTechMon. In: *Practical Aspects of Knowledge Management.* pp. 14–27.

Schiebel, E. (2012): Visualization of research fronts and knowledge bases by three-dimensional areal densities of bibliographically coupled publications and co-citations. *Scientometrics*, 91 (2), 557–566. doi: 10.1007/s11192-012-0626-8

van der Heijden, K. (1996). *Scenarios: The art of strategic conversation*. New York: John Wiley.

Topical Impact Analysis

A New Informetric Indicator for the Assessment of a Scientific Institution

Axel Honka, Lisa Orszullok,
Isabelle Dorsch, Nils Frommelius

Heinrich Heine University
Universitätsstraße 1, Düsseldorf, Germany
{Axel.Honka, Lisa.Orszullok, Isabelle.Dorsch, Nils.Frommelius}@hhu.de

Abstract

This paper presents new specific values regarding the topical impact analysis. These values make comprehensible statements that provide concrete comparative output to describe the differences between an initial topical map and an impact map. The purpose of this study is to evaluate the topical scope any institution has with its output. The Know-Center, an institute for knowledge technology in Graz, acts as a case study. To collect the citations, we used the reference search at Scopus. All publications of the Know-Center from 1st January 2003 until 31th December 2012 have been covered. According to the study, there is a high rate of new topics in the impact whereat durable topics have a higher occurrence comparing to the Know-Center's output.

Keywords: Topical impact analysis, Informetric indicator, Scientific institution, Citation, Topic

In: F. Pehar/C. Schlögl/C. Wolff (Eds.). Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th–21st May 2015. Glückstadt: Verlag Werner Hülsbusch, pp. 410–421.