Library and Information Science Papers Discussed on Twitter: A new Network-based Approach for Measuring Public Attention

Robin Haunschild^{1†}, Loet Leydesdorff², Lutz Bornmann³

¹Max Planck Institute for Solid State Research, Heisenbergstr. 1, 70569 Stuttgart, Germany

²Amsterdam School of Communication Research (ASCoR), University of Amsterdam, PB 15793, 1001 NG Amsterdam, The Netherlands

³Administrative Headquarters of the Max Planck Society, Division for Science and Innovation Studies, Hofgartenstr. 8, 80539 Munich, Germany

Citation: Haunschild, Robin, Loet Leydesdorff, and Lutz Bornmann. "Library and information science papers discussed on twitter: A new network-based approach for measuring public attention." *Journal of Data and Information Science*, vol. 5, no. 3, 2020, pp. 5–17. <https://doi.org/10.2478/jdis-2020-0017>

Received: Jan. 21, 2020

Revised: May 4, 2020

Accepted: May 20, 2020

Abstract

Purpose: In recent years, one can witness a trend in research evaluation to measure the impact on society or attention to research by society (beyond science). We address the following question: can Twitter be meaningfully used for the mapping of public and scientific discourses?

Design/methodology/approach: Recently, Haunschild et al. (2019) introduced a new network-oriented approach for using Twitter data in research evaluation. Such a procedure can be used to measure the public discussion around a specific field or topic. In this study, we used all papers published in the Web of Science (WoS, Clarivate Analytics) subject category Information Science & Library Science to explore the publicly discussed topics from the area of library and information science (LIS) in comparison to the topics used by scholars in their publications in this area.

Findings: The results show that LIS papers are represented rather well on Twitter. Similar topics appear in the networks of author keywords of all LIS papers, not tweeted LIS papers, and tweeted LIS papers. The networks of the author keywords of all LIS papers and not tweeted LIS papers are most similar to each other.

Research limitations: Only papers published since 2011 with DOI were analyzed.

Practical implications: Although Twitter data do not seem to be useful for quantitative research evaluation, it seems that Twitter data can be used in a more qualitative way for mapping of public and scientific discourses.

Originality/value: This study explores a rather new methodology for comparing public and scientific discourses.

Keywords Altmetrics; Twitter; News; Hashtags; Author keywords

[†] Corresponding author: Robin Haunschild (E-mail: r.haunschild@fkf.mpg.de).

1 Introduction

In recent years, we witnessed a general trend in research evaluation to measure the impact research has on society (beyond science) or the attention research receives from other parts of society. Whereas in the UK Research Excellence Framework (REF) the case-study approach was used for societal impact measurements, altmetrics has been proposed to measure impact or attention quantitatively (Bornmann, Haunschild, & Adams, 2019). Since the introduction of altmetrics, most quantitative studies focussed on Mendeley or Twitter data (e.g. saves of publications in this online reference manager and short messages with links to publications, respectively). Whereas Mendeley data might be useful in research evaluation to measure the early impact of publications (which can be scarcely measured by citations) (Thelwall, 2018), the usefulness of Twitter counts has frequently been questioned (e.g. Bornmann, 2015; Robinson-Garcia et al., 2017).

Hellsten and Leydesdorff (2020) analyzed Twitter data and mapped the co-occurrences of hashtags (as representation of topics) and usernames (as addressed actors). The resulting networks can show the relationships between three different types of nodes: authors, actors, and topics. The maps demonstrate how actors and topics are co-addressed in science-related communications. Wouters, Zahedi, and Costas (2019) discussed such an approach as a new and valid procedure to use social media data in research evaluation. Recently, Haunschild et al. (2019) explored a network-oriented approach for using Twitter data in research evaluation. Such a methodology can be used to measure the public discussion around a field or topic. For example, Haunschild et al. (2019) based their study on papers about climate change.

This approach can be used to study how the public discusses a certain topic differently from the discussion of the topic in the research community. In this study, we use all papers published during the period 2010–2017 in journals covered by the subject category “Information Science & Library Science” in the Web of Science (WoS, Clarivate Analytics). The objective is to explore the publicly discussed topics in comparison to topics of research as discussed within the journals classified as library and information science (LIS) by Clarivate Analytics.^①

^① This is a substantially extended study based on our ISSI 2019 conference contribution (Haunschild, Leydesdorff, & Bornmann, 2019), entitled “Library and Information Science papers as Topics on Twitter: A network approach to measuring public attention”.

2 Methodology

2.1 Datasets

We used the WoS data of the in-house database of the Max Planck Society (MPG) derived from the Science Citation Index Expanded (SCI-E), Social Sciences Citation Index (SSCI), and Arts and Humanities Citation Index (AHCI) licensed from Clarivate Analytics (Philadelphia, USA). In this database, 86,657 papers were assigned to the WoS subject category “Information Science & Library Science” and published between 2010 and 2017. Of these papers, 31,348 (36.2%) have a DOI in the database. Following previous studies (Bornmann, Haunschild, & Marx, 2016), we used the Perl module Bib::CrossRef[®] to search for additional DOIs. Only 2,478 additional DOIs were obtained by this procedure. The combined set of WoS and CrossRef DOIs was searched for DOIs occurring multiple times. Such DOIs were removed. Finally, a set of 33,312 papers (38.4%) with DOI was obtained.

The company Altmetric.com (see <https://www.altmetric.com>) tracks mentions of scientific papers in various altmetrics sources (e.g. Twitter, Facebook, news outlets, and Wikipedia). Twitter is monitored by the company Altmetric.com for tweets that reference scientific papers. Tweets may refer to the content of papers. Twitter users often use hashtags to index their tweets. News outlets are also monitored by the company Altmetric.com for online news items which reference scientific papers (via direct links and text mining or unique identifiers in, e.g. the *Washington Post*). Altmetric.com provides free access to the resulting datasets for research purposes for free via their API or snapshots.

We received the most recent snapshot from Altmetric.com on October 30, 2019. This snapshot was imported and processed in our locally maintained PostgreSQL database at the Max Planck Institute for Solid State Research. We used the combined set of 33,312 papers to match them via the DOIs with our locally maintained database of altmetrics data. In Haunschild, Leydesdorff, and Bornmann (2019) an earlier snapshot from Altmetric.com from 10th June 2018 was used. Recently, we found data problems regarding this data snapshot: (i) Altmetric.com offered a partial dataset, the limitations of which were not made clear at the time of delivery. (ii) Inadvertently, we did not import all data provided by Altmetric.com at that time into our local database due to an error in our routine. Therefore, we used the newer data snapshot for this study (see also Haunschild et al., 2020).

[®] See <http://search.cpan.org/dist/Bib-CrossRef/lib/Bib/CrossRef.pm>

Research Paper

The following information was appended to the DOIs: (1) links to the tweets which mentioned the respective papers, (2) the numbers of tweets in which the respective paper was mentioned, and (3) the numbers of mentions in news outlets of this same paper. Among the LIS papers with DOI, 13.2% of the (11,421) papers were mentioned in 91,914 tweets; 8.7% (n=7,513) of the papers were mentioned by at least two twitter accounts in 87,529 tweets. Only 0.5% (n=469) were also mentioned in news outlets. The additional consideration of news outlets is intended to identify topics in Twitter discussions which are also reflected in the news sector.

2.2 Data

In the most-recent Altmetric.com data dump no tweet URLs were available but only the IDs of tweets. We used these tweet IDs to download the 87,529 tweets with all additionally available information from the Twitter API using R (R Core Team, 2019) between 5th and 6th November 2019. We are interested in all author keywords and hashtags, including name variants. Since these names start with the # sign, no stop-word list is needed. The most frequently occurring author keywords and hashtags were selected for further analysis (see below). We used a cosine-normalized term co-occurrence matrix generated with a dedicated routine written in Visual Basic (see <https://www.leydesdorff.net/software/twitter>).

We exported four different sets of author keywords: (1) author keywords of all LIS papers, (2) author keywords of not-tweeted papers, (3) author keywords of papers tweeted at least twice, and (4) author keywords of papers tweeted at least twice and mentioned in news outlets at least once. In total 1,366 different author keywords occurred in LIS papers tweeted by at least two accounts and mentioned in news outlets at least once; 211 of these author keywords occurred at least twice, and 65 of them occurred at least three times. We used the top-65 author keywords of the sets of the different author keywords in order to compare networks of the same and a displayable size.

When we refer below to “tweeted papers”, only papers tweeted at least twice are meant. When we refer to “not-tweeted papers”, indeed not-tweeted papers are meant. Papers tweeted exactly once (n=3,908 papers) are not included in the analysis in order to reduce noise. Many papers are tweeted only a single time by the publisher or the authors themselves for self-promotion. We consider these single occurrences as noise.

2.3 Visualization

The resulting files (containing cosine-normalized distributions of terms in the Pajek format, see <http://mrvar.fdv.uni-lj.si/pajek>) were laid-out using the algorithm of Kamada and Kawai (1989) in Pajek and then exported to VOSviewer v.1.6.12

for visualizations. The community-searching algorithm in VOSviewer was employed with a resolution parameter of 1.0, minimum cluster size of 1, 10 random starts, 10 iterations, a random seed of 0, and the option “merge small clusters” enabled. The size of a node indicates the frequency of co-occurrence of a specific term with all other terms on the map. Lines between two nodes and their thickness indicate the co-occurrence frequency of these specific terms.

3 Results

3.1 Author keywords

Figure 1 shows the semantic map of the top-65 author keywords of LIS publications. This map visualizes the author keywords used within the scholarly communication. Five different clusters are marked by respective colours. These clusters reveal the broad spectrum of LIS research. The green cluster represents the core of scientometrics including bibliometrics and most of altmetrics. The yellow cluster is centred on text mining, data mining, and related topics, such as semantics and machine learning. The red cluster contains author keywords related to social media and social networks. The blue cluster deals mainly with libraries and higher-education issues. The purple cluster contains the author keywords “Social network analysis” and “Network analysis”. These methods are used in many of the other clusters’ papers. Both nodes of the purple cluster have many strong links to the red and green clusters. This also shows their topical relations to scientometrics and social media.

Figure 2 shows the semantic map of the top-64 author keywords of **not-tweeted** LIS publications. The author keywords on ranks 65–67 are tied in this case. Therefore, we decided to display the top-64 author keywords. The author keywords are grouped in six different clusters. Overall, the grouping is like the clustering in Figure 1. The semantic maps in Figure 1 and Figure 2 have an overlap of 55 author keywords (85.9%).

Figure 3 shows the semantic map of the top-63 author keywords of **tweeted** LIS publications. The author keywords on the ranks 64–69 are tied in this case. Therefore, we decided to display the top-63 author keywords. Six different clusters are found: the green, red, yellow, and blue clusters roughly correspond to their counter parts in Figure 1. The purple cluster comprises author keywords about qualitative research and health care while some author keywords related to electronic health records are grouped in the yellow (semantics and text mining) cluster. Overall, the semantic maps in Figure 1 and Figure 2 share 47 (74.6%) and 37 (58.7%), respectively, keywords with the semantic map in Figure 3. Although the quantitative agreement between the semantic maps in Figure 1, Figure 2, and Figure 3 decreases considerably,

Figure 1. Top-65 author keywords of LIS papers published between 2011 and 2017. An interactive version of this network can be viewed at <https://tinyurl.com/qwvtocq>. Note that the colour scheme may be different in the interactive version.

the qualitative agreement is still large for most of the top 63–65 author keywords of LIS papers. The core author keywords of scientometrics, bibliometrics, altmetrics, text mining, data mining, and social networks still appear in all maps and are grouped in the same clusters, independently of the specific variants of the indicator.

Figure 4 shows the semantic map of the top-65 author keywords of LIS publications which were **tweeted and mentioned in news outlets**. This network is less dense. Nine different clusters are shown in Figure 4. The rose cluster contains only a single author keyword: “Certification” (rose dot left of “Scientometrics” and “Citation_analysis”). The red cluster represents the core of scientometrics, bibliometrics, altmetrics, and scholarly publishing. The author keywords related to social media

Figure 2. Top-64 author keywords of not-tweeted LIS papers published between 2011 and 2017. An interactive version of this network can be viewed at: <https://tinyurl.com/u3569lc>. Note that the colour scheme may be different in the interactive version.

are split-up into two different clusters: light-blue and orange. The purple cluster contains author keywords related to electronic health issues. The yellow cluster contains various information-related author keywords. The green cluster contains author keywords related to journalism and big data. Health-related author keywords are also mixed in the green and yellow clusters. The blue cluster contains author keywords related to qualitative sociology research. The brown cluster is mainly related to privacy issues in the internet. Rather few author keywords in the semantic map of Figure 4 appeared also in the previous figures: 24 (36.9%) in the case of Figure 1, 19 (29.7%) in the case of Figure 2, and 28 (44.4%) in the case of Figure 3.

Table 1 shows the overlap between the top author keywords of all LIS publications (“All”), not-tweeted LIS publications (“Not tweeted”), LIS publications tweeted at least twice (“Tweeted”), and LIS publications tweeted both at least twice and

Figure 3. Top-63 author keywords of LIS papers tweeted and published between 2011 and 2017. An interactive version of this network can be viewed at: <https://tinyurl.com/rfmy4vz>. Note that the colour scheme may be different in the interactive version.

mentioned in news outlets at least once (“Tweeted and mentioned in the news”). The lower triangle shows the absolute number of overlapping author keywords and the upper triangle shows the proportion of overlapping keywords. The top-65 author keywords of publications which were tweeted and mentioned in news outlets show an overlap of about one third with the sets of top author keywords of all and not tweeted publications. The overlap with the author keywords of tweeted publications is higher. This might be partly due to the fact that the author keywords of publications which were tweeted and mentioned in news outlets are a sub-set of the author keywords of tweeted publications. However, this fact cannot explain all of the differences among the overlaps.

Figure 4. Top-65 author keywords of LIS papers tweeted, mentioned in news outlets at least once, and published between 2011 and 2017. An interactive version of this network can be viewed at: <https://tinyurl.com/twssvt5>. Note that the colour scheme may be different in the interactive version.

The focus of top author keyword selection varies slightly from all and not-tweeted publications to publications tweeted at least twice, but the focus varies significantly to publications tweeted at least twice and also mentioned in news outlets. These results suggest that Twitter activity is rather high in library and information sciences in comparison with other subject categories (Bornmann & Haunschild, 2016). Most of the topics seem to be used both on Twitter and in the scholarly literature. Most of these author keywords of LIS papers which were mentioned also in news outlets have a strong thematic relation to health care.

Research Paper

Table 1. Overlap between top author keywords. The lower triangle shows the absolute number of overlapping keywords and the upper triangle shows the proportion of overlapping keywords.

	All	Not tweeted	Tweeted	Tweeted and mentioned in the news
All	65	85.9%	74.6%	36.9%
Not tweeted	55	64	58.7%	29.7%
Tweeted	47	37	63	44.4%
Tweeted and mentioned in the news	24	19	28	65

3.2 Hashtags

Figure 5 shows the semantic map of the top-65 hashtags of tweets mentioning LIS publications. The hashtags are grouped in eight different clusters. The red cluster mainly contains hashtags related to libraries, scientometrics, bibliometrics, and altmetrics. The hashtags in the green cluster are related to digital and electronic health care. The yellow cluster contains hashtags about big data and open data

Figure 5. Top-65 hashtags from tweets which mentioned a LIS paper published between 2011 and 2017. An interactive version of this network can be viewed at: <https://tinyurl.com/sv8gpax>. Note that the colour scheme may be different in the interactive version.

related to health care issues and financial technology. The blue cluster is mainly related to the World Development Report 2016 entitled “Digital Dividends” and related topics. The purple cluster is focussed on open access and open science. The remaining three clusters are very small: three hashtags are gathered in the light-blue cluster regarding health-related issues which probably also very well could be part of the yellow or green cluster when other parameters in the cluster algorithm would be used. The hashtags (“#PAYWALLED” and “#RICKYPO”) are in the orange cluster. The brown cluster contains only the hashtag “#WIKILEAKS”.

The semantic map in Figure 5 shows many hashtags which are mainly related to the author keywords of the semantic map in Figure 4 but also hashtags which seem to be unrelated to all other semantic maps, e.g. most of the hashtags in the green, light-blue, blue, and orange clusters. Many other hashtags focus stronger on specific events and buzzwords, e.g. “#WDR2016”, “#ICT4D”, “#PAYWALLED”, and “#WIKILEAKS” than the author keywords.

4 Discussion and conclusions

Many scientometric studies used Twitter counts for measuring societal impact, but the meaningfulness of this data for these measurements in research evaluations (or measurements of attention) has been questioned (Haunschild et al., 2019). We followed our recent proposal (Haunschild et al., 2019) to focus on hashtags in tweets and author keywords in scientific papers in separate sets to differentiate public discussions of certain topics from their addressing in research. We analyzed three datasets: (1) author keywords of all LIS papers, (2) author keywords of not tweeted papers, (3) author keywords of papers tweeted at least twice, and (4) author keywords of papers tweeted at least twice and mentioned in news outlets at least once.

Our study is based on the papers in the WoS subject category LIS which have a DOI and were published between 2011 and 2017. Unfortunately, only less than half of the LIS papers have a DOI which is a major limitation of this study. We used Twitter data to reveal topics of public interest and compare them to research-focused topics. Such an analysis can provide insights into a subject category by revealing which topics enter the public discussions and which do not. Furthermore, the connections between the different topics become visible by using a network-oriented approach.

Our results show that topics in LIS papers seem to be represented rather well on Twitter. Similar topics appear in the networks of author keywords of all LIS papers, not tweeted LIS papers, and tweeted LIS papers. The networks of the author keywords of all LIS papers and not tweeted LIS papers are most similar to each

Research Paper

other in terms of author keyword overlap. Larger differences were found between these first three networks of scholarly communications, and the networks of hashtags and the networks author keywords of LIS papers which were tweeted by at least two accounts and mentioned in news outlets at least once as representations of public discourse. Both, the latter scholarly discourses and the tweets, are oriented towards digital and electronic health care more than tweeted LIS papers, not tweeted LIS papers, or all LIS papers. Our results confirm that only specific aspects of research outcomes intersect directly with the attention of the general public. Moving from the author keywords of all LIS papers to those author keywords of tweeted papers and those author keywords of papers additionally mentioned in the news, the focus shifts from theoretical applications and methodologies to health-applications, social media, privacy issues, and sociological studies.

Although we used another dump of data from Altmetric.com in this study than in our ISSI 2019 conference contribution (Haunschild, Leydesdorff, & Bornmann, 2019), the conclusions and interpretations in that conference paper were confirmed. In a similar paper on discussions about climate change, Haunschild et al. (2019) came to the following conclusion: “publications using scientific jargon are less likely to be tweeted than publications using more general keywords” (p. 18). A similar tendency was not visible in the current study using LIS papers and tweets as data. A possible reason for the difference is that the scientific jargon in LIS is less technical than in climate-change research.

Acknowledgments

The bibliometric data used in this paper are from an in-house database developed and maintained in cooperation with the Max Planck Digital Library (MPDL, Munich) and derived from the SCI-E, SSCI, AHCI prepared by Clarivate Analytics, formerly the IP & Science business of Thomson Reuters (Philadelphia, Pennsylvania, USA). The twitter and news data were retrieved from our locally maintained database with data shared with us by the company Altmetric on October 30, 2019. We thank two anonymous reviewers and Stacy Konkiel (Altmetric.com) for their positive and constructive comments.

Author contributions

Proposing the research problems: Robin Haunschild (R.Haunschild@fkf.mpg.de), Loet Leydesdorff (loet@leydesdorff.net), Lutz Bornmann (bornmann@gv.mpg.de); performing the research: RH, LL, LB; designing the research framework: RH, LL, LB; collecting and analyzing the data: RH; software development: RH, LL; writing and revising the manuscript: RH, LL, LB.

References

- Bornmann, L. (2015). Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics. *Scientometrics*, 103(3), 1123–1144. doi: 10.1007/s11192-015-1565-y.
- Bornmann, L., & Haunschild, R. (2016). How to normalize Twitter counts? A first attempt based on journals in the Twitter Index. *Scientometrics*, 107(3), 1405–1422. doi: 10.1007/s11192-016-1893-6.
- Bornmann, L., Haunschild, R., & Adams, J. (2019). Do altmetrics assess societal impact in a comparable way to case studies? An empirical test of the convergent validity of altmetrics based on data from the UK research excellence framework (REF). *Journal of Informetrics*, 13(1), 325–340. doi: 10.1016/j.joi.2019.01.008.
- Bornmann, L., Haunschild, R., & Marx, W. (2016). Policy documents as sources for measuring societal impact: How often is climate change research mentioned in policy-related documents? *Scientometrics*, 109(3), 1477–1495. doi: 10.1007/s11192-016-2115-y.
- Haunschild, R., Leydesdorff, L., & Bornmann, L. (2019). Library and Information Science papers as topics on Twitter: A network approach to measuring public attention. Paper presented at the ISSI 2019—17th International Conference of the International Society for Scientometrics and Informetrics, Rome, Italy.
- Haunschild, R., Leydesdorff, L., Bornmann, L., Hellsten, I., & Marx, W. (2019). Does the public discuss other topics on climate change than researchers? A comparison of explorative networks based on author keywords and hashtags. *Journal of Informetrics*, 13(2), 695–707. doi: 10.1016/j.joi.2019.03.008.
- Haunschild, R., Leydesdorff, L., Bornmann, L., Hellsten, I., & Marx, W. (2020). Corrigendum to “Does the public discuss other topics on climate change than researchers? A comparison of explorative networks based on author keywords and hashtags” [J. Informetrics 13 (2019) 695–707]. *Journal of Informetrics*, 14(1), February 2020, 101020. doi: 10.1016/j.joi.2020.101020
- Hellsten, I., & Leydesdorff, L. (2020). Automated analysis of actor–topic networks on twitter: New approaches to the analysis of socio-semantic networks. *JASIST*, 71(1), 3–15. doi: 10.1002/asi.24207
- R Core Team. (2019). R: A Language and Environment for Statistical Computing (Version 3.6.0). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Robinson-Garcia, N., Costas, R., Isett, K., Melkers, J., & Hicks, D. (2017). The unbearable emptiness of tweeting—About journal articles. *PLOS ONE*, 12(8), e0183551. doi: 10.1371/journal.pone.0183551.
- Thelwall, M. (2018). Early Mendeley readers correlate with later citation counts. *Scientometrics*, 115(3), 1231–1240. doi: 10.1007/s11192-018-2715-9.
- Wouters, P., Zahedi, Z., & Costas, R. (2019) Social media metrics for new research evaluation. In: Glänzel W., Moed H.F., Schmoch U., Thelwall M. (eds) *Springer Handbook of Science and Technology Indicators*. Springer Handbooks. Springer, Cham, pp 687–713.

This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Journal of Data and
Information Science

<http://www.jdis.org>

<https://www.degruyter.com/view/j/jdis>